Generalized Empirical Likelihood Inferences for Nonsmooth Moment Functions With Nonignorable Missing Values

Puying Zhao[1,2], Niansheng Tang[1], and Hongtu Zhu[3]

[1] *Yunnan University,* [2] *University of Waterloo, and*
[3] *The University of Texas MD Anderson Cancer Center*

Supplementary Material

This supplement material contains detailed technical proofs of Propositions 1–2, Theorems 1–2 and Lemmas 1–6. The same assumptions in the Appendix of the main paper are assumed. Throughout the supplement material, $\mathcal{C}$ represents a generic positive constant that may take different values in each case.

S1. Proofs of Propositions 1–2

Proof of Proposition 1. Under nonignorable missing-data assumption (2.2), we have

$$E\left\{\frac{\delta}{\pi(U,Y,\boldsymbol{\alpha}_0)}\Big|X,Y\right\}=1.$$

This, coupled with the iterated conditional expectation formula, shows Proposition 1(i).

We now consider the second statement. Let

$$\pi^*(U,Y)=\frac{\exp\{\varphi(U,\boldsymbol{\alpha}_0)+q(Y)\}}{1+\exp\{\varphi(U,\boldsymbol{\alpha}_0)+q(Y)\}},$$

be the true response model. Recall the following AIPW moment funcitons

$$\begin{aligned}\widetilde{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \frac{\delta_i g(X_i,Y_i,\boldsymbol{\beta})}{\pi(U_i,Y_i,\boldsymbol{\alpha})}-\frac{\delta_i-\pi(U_i,Y_i,\boldsymbol{\alpha})}{\pi(U_i,Y_i,\boldsymbol{\alpha})}m_g^0(U_i,\boldsymbol{\beta},\boldsymbol{\alpha})\\ &= g(X_i,Y_i,\boldsymbol{\beta})+\left\{1-\frac{\delta_i}{\pi(U_i,Y_i,\boldsymbol{\alpha})}\right\}\Big\{m_g^0(U_i,\boldsymbol{\beta},\boldsymbol{\alpha})-g(X_i,Y_i,\boldsymbol{\beta})\Big\},\end{aligned}$$

where

$$m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha})=\frac{E\{\delta g(X,Y,\boldsymbol{\beta})O(U,Y,\boldsymbol{\alpha})\mid U\}}{E\{\delta O(U,Y,\boldsymbol{\alpha})\mid U\}}$$

and $O(U,Y,\boldsymbol{\alpha})=\pi^{-1}(U,Y,\boldsymbol{\alpha})-1$. Suppose that the model $\varphi(U_i,\boldsymbol{\alpha})$ is misspecified, that is

$\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$. Simple algebraic manipulations show that

$$\begin{aligned}
& E\left[\left\{1-\frac{\delta}{\pi(U,Y,\boldsymbol{\alpha})}\right\}\left\{m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha})-g(X,Y,\boldsymbol{\beta})\right\} \mid U\right] \\
= \;& m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) - E\left\{\delta\frac{1-\pi(U,Y,\boldsymbol{\alpha})}{\pi(U,Y,\boldsymbol{\alpha})} \mid U\right\} m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) \\
& -E(\delta \mid U) m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) - E\{g(X,Y,\boldsymbol{\beta}) \mid U\} \\
& +E\left\{\delta\frac{1-\pi(U,Y,\boldsymbol{\alpha})}{\pi(U,Y,\boldsymbol{\alpha})} g(X,Y,\boldsymbol{\beta}) \mid U\right\} + E\{\delta g(X,Y,\boldsymbol{\beta}) \mid U\} \\
= \;& m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) E\{1-\pi^*(U,Y) \mid U\} - E\{(1-\pi^*(U,Y)) g(X,Y,\boldsymbol{\beta}) \mid U\}.
\end{aligned}$$

The last equality is true because

$$\begin{aligned}
& E\left\{\delta\frac{1-\pi(U,Y,\boldsymbol{\alpha})}{\pi(U,Y,\boldsymbol{\alpha})} \mid U\right\} m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) \\
= \;& E\left\{\delta O(U,Y,\boldsymbol{\alpha}) \mid U\right\} \frac{E\{\delta g(X,Y,\boldsymbol{\beta}) O(U,Y,\boldsymbol{\alpha}) \mid U\}}{E\{\delta O(U,Y,\boldsymbol{\alpha}) \mid U\}} \\
= \;& E\left\{\delta g(X,Y,\boldsymbol{\beta}) O(U,Y,\boldsymbol{\alpha}) \mid U\right\} \\
= \;& E\left\{\delta\frac{1-\pi(U,Y,\boldsymbol{\alpha})}{\pi(U,Y,\boldsymbol{\alpha})} g(X,Y,\boldsymbol{\beta}) \mid U\right\}.
\end{aligned}$$

Note that

$$\begin{aligned}
m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) &= \frac{E\{\delta g(X,Y,\boldsymbol{\beta}) O(U,Y,\boldsymbol{\alpha}) \mid U\}}{E\{\delta O(U,Y,\boldsymbol{\alpha}) \mid U\}} \\
&= \frac{E\{\pi^*(U,Y) g(X,Y,\boldsymbol{\beta}) O(U,Y,\boldsymbol{\alpha}) \mid U\}}{E\{\pi^*(U,Y) O(U,Y,\boldsymbol{\alpha}) \mid U\Big\}} \\
&= \frac{E\left\{\dfrac{1}{1+\exp\{\varphi(U,\boldsymbol{\alpha}_0)+q(Y)\}} g(X,Y,\boldsymbol{\beta}) \mid U\right\}}{E\left\{\dfrac{1}{1+\exp\{\varphi(U,\boldsymbol{\alpha}_0)+q(Y)\}} \mid U\right\}} \\
&= \frac{E\{(1-\pi^*(U,Y)) g(X,Y,\boldsymbol{\beta}) \mid U\}}{E\{1-\pi^*(U,Y) \mid U\}}.
\end{aligned}$$

Combining above arguments, we can show that, for any $\boldsymbol{\alpha} \in \mathcal{A}$,

$$E\left[\left\{1-\frac{\delta}{\pi(U,Y,\boldsymbol{\alpha})}\right\}\{m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha})-g(X,Y,\boldsymbol{\beta})\} \mid U\right] = 0.$$

Therefore, $E\{\widetilde{g}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha})\} = E\{g(X_i,Y_i,\boldsymbol{\beta}_0)\} = 0$ even when the model for $\varphi(U_i,\boldsymbol{\alpha})$ is misspecified. This completes the proof of Proposition 1. □

Proof of Proposition 2. Define $\boldsymbol{\gamma} = \lambda_1(1-\boldsymbol{\omega})$, $\boldsymbol{\eta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\omega}, \boldsymbol{\gamma}^\top)^\top$, $\boldsymbol{\eta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\omega}_0, 0)^\top$ and $\tau_n = nn_1^{-1} - \boldsymbol{\omega}_0^{-1}$, where $(\boldsymbol{\alpha}_0, \boldsymbol{\omega}_0)$ denotes the true value of $(\boldsymbol{\alpha}, \boldsymbol{\omega})$. The proof of consistency of estimator $\widehat{\boldsymbol{\eta}} = (\widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\omega}}, \widehat{\boldsymbol{\gamma}}^\top)^\top$ is similar to that of Theorem 1 of Qin et al. (2002), and thus the details are omitted. We now consider the asymptotic normality of $\widehat{\boldsymbol{\eta}}$. We only outline the main steps in proving the asymptotic normality, and the details of the proof can be found in Zhao et al. (2017). Let $\pi_i(\boldsymbol{\alpha}) = \pi(U_i,Y_i,\boldsymbol{\alpha})$, $\phi_i(\boldsymbol{\alpha}) = \phi(X_i,Y_i,\boldsymbol{\alpha})$, and $\mathcal{D}_i(\boldsymbol{\eta},\tau_n) = 1 -$

$\boldsymbol{\omega}/\boldsymbol{\omega}_0 + (1/\boldsymbol{\omega}_0 - 1)\pi_i(\boldsymbol{\alpha}) + \boldsymbol{\gamma}^\top\phi_i(\boldsymbol{\alpha}) + \tau_n(\pi_i(\boldsymbol{\alpha}) - \boldsymbol{\omega})$. Rewrite $1 + \lambda_1^\top\phi_i(\boldsymbol{\alpha}) + \lambda_2(\pi_i(\boldsymbol{\alpha}) - \boldsymbol{\omega}) = \{1-\boldsymbol{\omega}\}^{-1}\{1 - \boldsymbol{\omega}/\boldsymbol{\omega}_0 + (1/\boldsymbol{\omega}_0 - 1)\pi_i(\boldsymbol{\alpha}) + \boldsymbol{\gamma}^\top\phi_i(\boldsymbol{\alpha}) + \tau_n(\pi_i(\boldsymbol{\alpha}) - \boldsymbol{\omega})\}$. Define

$$\mathcal{M}_n(\boldsymbol{\eta}, \tau_n) = \frac{1}{n}\sum_{i=1}^n \delta_i(\mathcal{M}_{i1}^\top(\boldsymbol{\eta},\tau_n), \mathcal{M}_{i2}(\boldsymbol{\eta},\tau_n), \mathcal{M}_{i3}^\top(\boldsymbol{\eta},\tau_n))^\top,$$

where

$$\begin{aligned}
\mathcal{M}_{i1}(\boldsymbol{\eta},\tau_n) &= \partial\log\pi_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^\top - [\boldsymbol{\gamma}^\top\partial\phi_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^\top \\
&\quad + \{\tau_n + (1-\boldsymbol{\omega}_0)/\boldsymbol{\omega}_0\}\partial\pi_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^\top]/\mathcal{D}_i(\boldsymbol{\eta},\tau_n), \\
\mathcal{M}_{i2}(\boldsymbol{\eta},\tau_n) &= (\pi_i(\boldsymbol{\alpha}) - \boldsymbol{\omega})/\mathcal{D}_i(\boldsymbol{\eta},\tau_n), \\
\mathcal{M}_{i3}(\boldsymbol{\eta},\tau_n) &= \phi_i(\boldsymbol{\alpha})/\mathcal{D}_i(\boldsymbol{\eta},\tau_n)
\end{aligned}$$

with $\mathcal{D}_i(\boldsymbol{\eta},\tau_n) = 1 - \boldsymbol{\omega}/\boldsymbol{\omega}_0 + (1/\boldsymbol{\omega}_0 - 1)\pi_i(\boldsymbol{\alpha}) + \boldsymbol{\gamma}^\top\phi_i(\boldsymbol{\alpha}) + \tau_n(\pi_i(\boldsymbol{\alpha}) - \boldsymbol{\omega})$. Simple algebraic manipulation shows that

$$\frac{\partial\mathcal{M}_n(\boldsymbol{\eta},\tau_n)}{\partial\boldsymbol{\eta}}\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0,\tau_n=0} = -\mathbb{A} + o_p(1),$$

where

$$\mathbb{A} = E\begin{pmatrix} 0 & \dfrac{\partial_{\boldsymbol{\alpha}}\pi_i(\boldsymbol{\alpha}_0)}{\tilde{\boldsymbol{\omega}}_0\pi_i(\boldsymbol{\alpha}_0)} & \dfrac{\boldsymbol{\omega}_0\phi_i(\boldsymbol{\alpha}_0)\partial_{\boldsymbol{\alpha}}\pi_i(\boldsymbol{\alpha}_0)}{\tilde{\boldsymbol{\omega}}_0\pi_i(\boldsymbol{\alpha}_0)} \\ -\dfrac{\boldsymbol{\omega}_0^2\partial_{\boldsymbol{\alpha}}\pi_i(\boldsymbol{\alpha}_0)}{\tilde{\boldsymbol{\omega}}_0\pi_i(\boldsymbol{\alpha}_0)} & -\dfrac{\boldsymbol{\omega}_0^2(\pi_i(\boldsymbol{\alpha}_0)+1)}{\tilde{\boldsymbol{\omega}}_0^2\pi_i(\boldsymbol{\alpha}_0)} & \dfrac{\boldsymbol{\omega}_0^2\phi_i(\boldsymbol{\alpha}_0)(\pi_i(\boldsymbol{\alpha}_0)-\boldsymbol{\omega}_0)}{\tilde{\boldsymbol{\omega}}_0^2\pi_i(\boldsymbol{\alpha}_0)} \\ \dfrac{\boldsymbol{\omega}_0\phi_i(\boldsymbol{\alpha}_0)\partial_{\boldsymbol{\alpha}}\pi_i(\boldsymbol{\alpha}_0)}{\tilde{\boldsymbol{\omega}}_0\pi_i(\boldsymbol{\alpha}_0)} & -\dfrac{\boldsymbol{\omega}_0\phi_i(\boldsymbol{\alpha}_0)}{\tilde{\boldsymbol{\omega}}_0^2\pi_i(\boldsymbol{\alpha}_0)} & \dfrac{\boldsymbol{\omega}_0^2\phi_i(\boldsymbol{\alpha}_0)\phi_i^\top(\boldsymbol{\alpha}_0)}{\tilde{\boldsymbol{\omega}}_0\pi_i(\boldsymbol{\alpha}_0)} \end{pmatrix}$$

in which $\tilde{\boldsymbol{\omega}}_0 = 1 - \boldsymbol{\omega}_0$, and $\partial_{\boldsymbol{\alpha}}\pi_i(\boldsymbol{\alpha}_0) = \partial\pi_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^\top|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$. It follows from Rothenberg (1971, p. 581) that $\boldsymbol{\eta}_0$ is identified if and only if $\mathbb{A}$ has rank $l+\kappa+1$. Define $\Phi_i = \delta_i(\mathcal{M}_{i1}^\top(\boldsymbol{\eta}_0,0), \mathcal{M}_{i2}(\boldsymbol{\eta}_0,0), \mathcal{M}_{i3}^\top(\boldsymbol{\eta}_0,0))^\top + \mathbb{H}(1/\boldsymbol{\omega}_0 - \delta_i/\boldsymbol{\omega}_0^2)$ with $\mathbb{H} = (h_1^\top, h_2, h_3^\top)^\top$, where

$$\begin{aligned}
h_1 &= \frac{-\boldsymbol{\omega}_0^2}{1-\boldsymbol{\omega}_0}E\left\{\frac{\partial\log\pi_i(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}}\right\}, \\
h_2 &= \frac{\boldsymbol{\omega}_0^2}{(1-\boldsymbol{\omega}_0)^2}E\left\{\frac{(\pi_i(\boldsymbol{\alpha}_0)-\boldsymbol{\omega}_0)^2}{\pi_i(\boldsymbol{\alpha}_0)}\right\}, \\
h_3 &= \frac{\boldsymbol{\omega}_0^2}{(1-\boldsymbol{\omega}_0)^2}E\left\{\frac{\phi_i(\boldsymbol{\alpha}_0)(\pi_i(\boldsymbol{\alpha}_0)-\boldsymbol{\omega}_0)}{\pi_i(\boldsymbol{\alpha}_0)}\right\}.
\end{aligned}$$

Following Qin et al. (2002) and Zhao et al. (2017), we obtain

$$n^{1/2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = -\mathbb{A}^{-1}n^{-1/2}\sum_{i=1}^n\Phi_i + o_p(1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{A}^{-1}\mathbb{B}(\mathbb{A}^{-1})^\top),$$

where $\mathbb{B} = \mathrm{Var}\{\Lambda_i\}$. Then, the asymptotic expansion for $\widehat{\boldsymbol{\alpha}}$ is given by $n^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^n\Psi_i(\boldsymbol{\eta}_0) + o_p(1)$, where $\Psi_i(\boldsymbol{\eta}_0)$ is the influence function, which is given by the appropriate submatrix of $-\mathbb{A}^{-1}\Lambda_i$. The proof of Proposition 2 is then completed. □

S2. Proofs of Lemmas 1–6

Lemma 1. *Suppose that Assumption C holds. Then, we have*

$$\sup_{\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\alpha}\in\mathcal{A}} \|\widehat{m}_g^0(\boldsymbol{\beta},\boldsymbol{\alpha}) - m_g^0(\boldsymbol{\beta},\boldsymbol{\alpha})\|_\infty = o_p(n^{-1/4}).$$

Proof. Rewrite

$$m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{m_1(U,\boldsymbol{\beta},\boldsymbol{\alpha})}{m_2(U,\boldsymbol{\alpha})}, \quad \text{and} \quad \widehat{m}_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{\widehat{m}_1(U,\boldsymbol{\beta},\boldsymbol{\alpha})}{\widehat{m}_2(U,\boldsymbol{\alpha})},$$

where $m_1(U,\boldsymbol{\beta},\boldsymbol{\alpha}) = E\{\delta g(X,Y,\boldsymbol{\beta})O(U,Y,\boldsymbol{\alpha}) \mid U\}$, $m_2(U,\boldsymbol{\alpha}) = E\{\delta O(U,Y,\boldsymbol{\alpha}) \mid U\}$, and

$$\widehat{m}_1(U,\boldsymbol{\beta},\boldsymbol{\alpha}) = \sum_{i=1}^n \omega_i(U,\boldsymbol{\alpha})g(X_i,Y_i,\boldsymbol{\beta}), \quad \widehat{m}_2(U,\boldsymbol{\alpha}) = \sum_{i=1}^n \omega_i(U,\boldsymbol{\alpha}),$$

in which $\omega_i(U,\boldsymbol{\alpha}) = \delta_i O_i(\boldsymbol{\alpha})\mathcal{K}_h(U-U_i)/\sum_{i=1}^n \mathcal{K}_h(U-U_i)$ with $O_i(\boldsymbol{\alpha}_0) = O(U_i,Y_i,\boldsymbol{\alpha}_0)$. Using the arguments of Lemma B.6 in Kitamura et al. (2004), together with Assumption C, we have

$$\sup_{(u,\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{U}\times\mathcal{B}\times\mathcal{A}} |\widehat{m}_1^0(u,\boldsymbol{\beta},\boldsymbol{\alpha}) - m_1(u,\boldsymbol{\beta},\boldsymbol{\alpha})| = o_p(n^{-1/4}), \quad \text{and}$$
$$\sup_{(u,\boldsymbol{\alpha})\in\mathcal{U}\times\mathcal{A}} |\widehat{m}_2^0(u,\boldsymbol{\alpha}) - m_2(u,\boldsymbol{\alpha})| = o_p(n^{-1/4}).$$

Using Lemma D.1 in Kitamura et al. (2004), we have

$$\sup_{(u,\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{U}\times\mathcal{B}\times\mathcal{A}} \left| \frac{\widehat{m}_1(u,\boldsymbol{\beta},\boldsymbol{\alpha})}{\widehat{m}_2(u,\boldsymbol{\alpha})} - \frac{m_1(u,\boldsymbol{\beta},\boldsymbol{\alpha})}{m_2(u,\boldsymbol{\alpha})} \right| = o_p(n^{-1/4}).$$

The proof of Lemma 1 is then completed. □

Lemma 2. *Suppose that Assumption C holds; that the respondent probability model $\pi(U,Y,\boldsymbol{\alpha}_0)$ is correctly specified; and that $\widehat{\boldsymbol{\alpha}}$ is computed by the SEL approach. Then, we have*

$$\mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^n \widetilde{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0) - \Xi \times (\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o_p(n^{-1/2}),$$

where $\widetilde{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})$ is defined in (2.6), $\Xi = \mathrm{Cov}\{\widetilde{g}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0), \Delta(U,Y,\boldsymbol{\alpha}_0)\}$ with $\Delta(U,Y,\boldsymbol{\alpha}) = \{\delta - \pi(U,Y,\boldsymbol{\alpha})\}\partial\mathrm{logit}\{\pi(U,Y,\boldsymbol{\alpha})\}/\partial\boldsymbol{\alpha}^\top$.

Proof. Taylor expansion of $\mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})$ at $\boldsymbol{\alpha}_0$ yields

$$\mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) = \mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) + \partial\mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}^\top \times (\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o(\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|).$$

We first consider the asymptotic property of $\mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)$. For $\mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)$, we consider the following decomposition

$$\begin{aligned}
&\mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) \\
&\quad = \frac{1}{n}\sum_{i=1}^n \left\{ \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} g(X_i,Y_i,\boldsymbol{\beta}_0) - \frac{\delta_i - \pi_i(\boldsymbol{\alpha}_0)}{\pi_i(\boldsymbol{\alpha}_0)} m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) \right\} \\
&\qquad + \frac{1}{n}\sum_{i=1}^n \left\{ 1 - \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} \right\} \{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) - m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\} \\
&\quad =: I_{n1} + I_{n2}.
\end{aligned}$$

Define $\pi(U) = E(\delta \mid U)$ and $G(U) = f(U)\{1-\pi(U)\}$. Under assumption (2.2) in the main paper, we have $E\{\delta O(U,Y,\boldsymbol{\alpha}_0) \mid U\} = 1-\pi(U)$. Thus, using the kernel regression method yields

$$\widehat{G}(U) = \frac{1}{n}\sum_{i=1}^{n}\delta_i O_i(\boldsymbol{\alpha}_0)\mathcal{K}_h(U_i - U).$$

Define $\mathcal{O}_{ij} = \delta_j O_j(\boldsymbol{\alpha}_0)\mathcal{K}_h(U_j-U_i)$. Then, for I_{n2}, we have $I_{n2} = I_{n21}+I_{n22}+I_{n23}$, where

$$\begin{aligned}
I_{n21} &= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\frac{\frac{1}{n}\sum_{j=1}^{n}\mathcal{O}_{ij}\{g(X_j,Y_j,\boldsymbol{\beta}_0)-m_g^0(U_j,\boldsymbol{\beta}_0)\}}{G(U_i)},\\
I_{n22} &= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0)\}\left\{1-\frac{\widehat{G}(U_i)}{G(U_i)}\right\},\\
I_{n23} &= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\frac{\frac{1}{n}\sum_{j=1}^{n}\mathcal{O}_{ij}\{m_g^0(U_j,\boldsymbol{\beta}_0)-m_g^0(U_i,\boldsymbol{\beta}_0)\}}{G(U_i)}.
\end{aligned}$$

We first derive the asymptotic distribution of I_{n21}. Note that

$$\begin{aligned}
I_{n21} &= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\frac{\frac{1}{n}\sum_{j=1}^{n}\mathcal{O}_{ij}\{g(X_j,Y_j,\boldsymbol{\beta}_0)-m_g^0(U_j,\boldsymbol{\beta}_0)\}}{G(U_i)}\\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathcal{M}_j\frac{\{1-\delta_i/\pi_i(\boldsymbol{\alpha}_0)\}\mathcal{K}_h(U_j-U_i)}{G(U_i)}\\
&=: \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}Q_{ij},
\end{aligned}$$

where $\mathcal{M}_j = \delta_j O_j(\boldsymbol{\alpha}_0)\{g(X_j,Y_j,\boldsymbol{\beta}_0)-m_g^0(U_j,\boldsymbol{\beta}_0)\}$. Define

$$\check{I}_{n21} = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}E\{Q_{ij}|(X_j,Y_j,\delta_j)\},$$

and write $I_{n21} = \check{I}_{n21} + I_{n21} - \check{I}_{n21}$. In what follows, we will establish the fact that I_{n21} is dominated by $\check{I}_{n21}$, whilst $I_{n21}-\check{I}_{n21}$ is of smaller order. Under nonignorable missing data mechanism, together with conditional independence assumption $\delta \perp\!\!\!\perp Z \mid (U,Y)$, we can show

$$\begin{aligned}
\check{I}_{n21} &= \frac{1}{n}\sum_{j=1}^{n}\mathcal{M}_j E\left\{\frac{\{1-\delta_i/\pi_i(\boldsymbol{\alpha}_0)\}\mathcal{K}_h(U_j-U_i)}{G(U_i)}\Big|X_j,Y_j\right\}\\
&= \frac{1}{n}\sum_{j=1}^{n}\mathcal{M}_j E\left[E\left\{\frac{\{1-\delta_i/\pi_i(\boldsymbol{\alpha}_0)\}\mathcal{K}_h(U_j-U_i)}{G(U_i)}\Big|X_j,Y_j,X_i,Y_i\right\}\Big|X_j,Y_j\right]\\
&= \frac{1}{n}\sum_{j=1}^{n}\mathcal{M}_j E\left[\frac{\mathcal{K}_h(U_j-U_i)}{G(U_i)}E\left\{1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\Big|X_j,Y_j,X_i,Y_i\right\}\Big|X_j,Y_j\right]\\
&= 0.
\end{aligned}$$

Let $I_{n21i} = n^{-1}\sum_{j=1}^{n}Q_{ij}$ and $\check{I}_{n21i} = n^{-1}\sum_{j=1}^{n}E\{Q_{ij}|(X_j,Y_j,\delta_j)\}$. Simple algebraic manipulations lead to

$$\begin{aligned}
nE(I_{n21}-\check{I}_{n21})^2 &= \frac{1}{n}\sum_{i=1}^{n}E(I_{n21i}-\check{I}_{n21i})^2 + \frac{2}{n}\sum_{i\neq j}E\{(I_{n21i}-\check{I}_{n21i})(I_{n21j}-\check{I}_{n21j})\}\\
&= E(I_{n21i}-\check{I}_{n21i})^2.
\end{aligned}$$

The last equality is true, since $E_{i\neq j}\{(I_{n21i}-\check{I}_{n21i})(I_{n21j}-\check{I}_{n21j})\}=0$. It follows from Lemma 1 that

$$\frac{\frac{1}{n}\sum_{j=1}^{n}\mathcal{M}_j\mathcal{K}_h(U_j-U_i)}{\widehat{G}(U_i)} \quad = \quad o_p(1).$$

In addition, using Assumption (C2), we have $\max_{1\leq i\leq n}\{1-\delta_i/\pi_i(\boldsymbol{\alpha}_0)\}\leq \max(1,|1-\mathcal{C}^{-1}|)=:\mathcal{C}_0$, which is bounded. Therefore, we obtain

$$\begin{aligned} nE(I_{n21}-\check{I}_{n21})^2 &= EI_{n21i}^2-E\check{I}_{n21i}^2\leq EI_{n21i}^2\\ &\leq \mathcal{C}_0E\left\{\frac{n^{-1}\sum_{j=1}^{n}\mathcal{M}_j\mathcal{K}_h(U_j-U_i)}{G(U_i)}\right\}^2\\ &\to 0. \end{aligned}$$

This yields $I_{n21}-\check{I}_{n21}=o_p(n^{-1/2})$. Therefore, we have $I_{n21}=o_p(n^{-1/2})$.

The standard arguments can be used to prove $\sup_{u\in\mathcal{U}}|\widehat{G}(u)-G(u)|=o_p(n^{-1/4})$. This together with the result from Lemma 1 leads to

$$|I_{n22}| \quad \leq \quad o_p(n^{-1/2})\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1-\delta_i}{G(U_i)}\right|.$$

By the law of large number, we can show

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1-\delta_i}{G(U_i)}=E\{f^{-1}(U)\}+o_p(1).$$

Then, we obtain $I_{n22}=o_p(n^{-1/2})$ due to $E\{f^{-1}(U)\}<\infty$. For I_{n23}, a similar derivation to that for I_{n21} shows $I_{n23}=o_p(n^{-1/2})$.

We now consider the asymptotic property of $\partial\mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}^\top$. By calculation, we obtain

$$\begin{aligned} \partial\mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}^\top &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\delta}{\pi_i^2(\boldsymbol{\alpha}_0)}\{g(X_i,Y_i,\boldsymbol{\beta}_0)-m_g^0(U_i,\boldsymbol{\beta},\boldsymbol{\alpha}_0)\}\frac{\partial\pi_i(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}^\top}\\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\frac{\delta}{\pi_i^2(\boldsymbol{\alpha}_0)}\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\}\frac{\partial\pi_i(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}^\top}\\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\frac{\partial\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}^\top}\\ &= T_{n1}+T_{n2}+T_{n3}. \end{aligned}$$

Let $\xi(U,Y,\boldsymbol{\alpha})=\partial\text{logit}\{\pi(U,Y,\boldsymbol{\alpha})\}/\partial\boldsymbol{\alpha}$. Note that

$$\frac{\partial\pi(U,Y,\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}^\top}=\pi(U,Y,\boldsymbol{\alpha}_0)(1-\pi(U,Y,\boldsymbol{\alpha}_0))\xi(U,Y,\boldsymbol{\alpha}_0)^\top, \quad \text{and}$$

$$\begin{aligned} &E\left\{\frac{\delta}{\pi(U,Y,\boldsymbol{\alpha}_0)}\{1-\pi(U,Y,\boldsymbol{\alpha}_0)\}\{g(X,Y,\boldsymbol{\beta})-m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}_0)\}\xi(U,Y,\boldsymbol{\alpha}_0)^\top\right\}\\ &=E\left\{\frac{\delta}{\pi(U,Y,\boldsymbol{\alpha}_0)}\{g(X,Y,\boldsymbol{\beta})-m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}_0)\}\{\delta-\pi(U,Y,\boldsymbol{\alpha}_0)\}\xi(U,Y,\boldsymbol{\alpha}_0)^\top\right\}\\ &=E\left\{\frac{\delta}{\pi(U,Y,\boldsymbol{\alpha}_0)}\{g(X,Y,\boldsymbol{\beta})-m_g^0(U,\boldsymbol{\beta},\boldsymbol{\alpha}_0)\}\Delta(U,Y,\boldsymbol{\alpha}_0)^\top\right\}\\ &=E\left\{\widetilde{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\Delta(U,Y,\boldsymbol{\alpha}_0)^\top\right\}=\text{Cov}\{\widetilde{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0),\Delta(U,Y,\boldsymbol{\alpha}_0)\}=:\Xi. \end{aligned}$$

The third equality is true because $E\{\Delta(U,Y,\boldsymbol{\alpha}_0)|X,Y\}=0$ from the assumption $\delta \perp\!\!\!\perp Z \mid (U,Y)$. Then, for T_{n1}, we have $T_{n1}=-\Xi+o_p(1)$.

It follows from Lemma 1 and Assumption (A2) that $T_{n2}=o_p(1)$. We now consider asymptotic property of T_{n3}. Define $m_\xi^0(U,\boldsymbol{\alpha})=E\{\xi(U,Y,\boldsymbol{\alpha})^\top \mid U,\delta=0\}$ and $m_{g\xi}^0(U,\boldsymbol{\beta},\boldsymbol{\alpha})=E\{g(X,Y,\boldsymbol{\beta})\xi(U,Y,\boldsymbol{\alpha})^\top \mid U,\delta=0\}$. By calculation, we obtain $\partial O_i(\boldsymbol{\alpha}_0)/\partial\alpha=-O_i(\boldsymbol{\alpha}_0)\xi(U_i,Y_i,\boldsymbol{\alpha}_0)$. Then, we have

$$\frac{\partial \widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^\top}=\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha})\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha})-\widehat{m}_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}),$$

where

$$\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha})=\frac{\sum_{j=1}^n \delta_j O_j(\boldsymbol{\alpha})\mathcal{K}_h(U_j-U_i)\xi(U_j,Y_j,\boldsymbol{\alpha})^\top}{\sum_{j=1}^n \delta_j O_j(\boldsymbol{\alpha})\mathcal{K}_h(U_j-U_i)},$$

$$\widehat{m}_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha})=\frac{\sum_{j=1}^n \delta_j O_j(\boldsymbol{\alpha})\mathcal{K}_h(U_j-U_i)g(X_j,Y_j,\boldsymbol{\beta}_0)\xi(U_j,Y_j,\boldsymbol{\alpha})^\top}{\sum_{j=1}^n \delta_j O_j(\boldsymbol{\alpha})\mathcal{K}_h(U_j-U_i)}.$$

Here $\widehat{m}_\xi^0(U,\boldsymbol{\alpha})$ and $\widehat{m}_{g\xi}^0(U_i,\boldsymbol{\beta},\boldsymbol{\alpha})$ are nonparametric regression estimators of $m_\xi^0(U,\boldsymbol{\alpha})$ and $m_{g\xi}^0(U,\boldsymbol{\beta},\boldsymbol{\alpha})$, respectively. Let $\Delta_n(U_i)=\widehat{G}(U_i)-G(U_i)$. Taking further decomposition for T_{n3}, we have

$$\begin{aligned}
T_{n3} &= \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha}_0)-\widehat{m}_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\}\\
&= \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)m_\xi^0(U_i,\boldsymbol{\alpha}_0)\}\\
&\quad -\frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)m_\xi^0(U_i,\boldsymbol{\alpha}_0)\}\\
&=: T_{n31}-T_{n32}
\end{aligned}$$

For T_{n31}, we have

$$\begin{aligned}
T_{n31} &= \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)m_\xi^0(U_i,\boldsymbol{\alpha}_0)\}\\
&= \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\}m_\xi^0(U_i,\boldsymbol{\alpha}_0)\\
&+ \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\}\{\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha}_0)-m_\xi^0(U_i,\boldsymbol{\alpha}_0)\}\\
&+ \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\{\widehat{m}_\xi^0(U_i,\boldsymbol{\alpha}_0)-m_\xi^0(U_i,\boldsymbol{\alpha}_0)\}\\
&=: T_{n311}+T_{n312}+T_{n313}.
\end{aligned}$$

For T_{n32}, we have

$$\begin{aligned}
T_{n32} &= \frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{\widehat{m}_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-m_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\beta}_0,\boldsymbol{\alpha})\}\\
&\quad +\frac{1}{n}\sum_{i=1}^n\left(1-\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\right)\{m_{g\xi}^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha})-m_g^0(U_i,\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)m_\xi^0(U_i,\boldsymbol{\alpha}_0)\}\\
&=: T_{n321}+T_{n322}.
\end{aligned}$$

A similar derivation to that for I_{n21}, together with standard arguments, leads to $T_{n31j} = o_p(n^{-1/2})$ for $j = 1,2,3$, and $T_{n321} = o_p(1)$. By the law of large number, we obtain $T_{n322} = o_p(1)$. Combining the above arguments leads to $T_{n3} = o_p(1)$. The proof of Lemma 2 is thus completed. □

Lemma 3. *Suppose that Assumption C holds; that the respondent probability model* $\pi(U, Y, \boldsymbol{\alpha}_0)$ *is correctly specified; and that* $\widehat{\boldsymbol{\alpha}}$ *is computed by the SEL approach. Then, we have*

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{g}_i(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}})\widehat{g}_i(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}})^\top = V_1 + o_p(1),$$

where $V_1 = E\{\widetilde{g}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\widetilde{g}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)^\top\}$.

Proof. Define

$$\widehat{V}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{g}_i(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}})\widehat{g}_i(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}})^\top, \quad \text{and} \quad \widetilde{V}_n = \frac{1}{n}\sum_{i=1}^{n} \widetilde{g}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\widetilde{g}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)^\top.$$

As $n \to \infty$, we have $\widetilde{V}_n = V_1 + o_p(1)$. Next, using Holder's Inequality, we have

$$\left|\widehat{V}_n - V_1\right| \le \left|\widehat{V}_n - \widetilde{V}_n\right| + \left|\widetilde{V}_n - V_1\right| \le R_{1n} + R_{2n} + R_{3n} + R_{4n} + R_{5n} + o_p(1),$$

where

$$R_{1n} = \frac{1}{n}\sum_{i=1}^{n}\Big|\frac{\delta_i}{\pi_i(\widehat{\boldsymbol{\alpha}})} g(X_i, Y_i, \boldsymbol{\beta}_0) - \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} g(X_i, Y_i, \boldsymbol{\beta}_0)\Big|^2,$$

$$R_{2n} = \frac{1}{n}\sum_{i=1}^{n}\Big|\frac{\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})} \widehat{m}_g^0(U_i, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}}) - \frac{\delta_i - \pi_i(\boldsymbol{\alpha}_0)}{\pi_i(\boldsymbol{\alpha}_0)} m_g^0(U_i, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\Big|^2,$$

$$R_{3n} = 2R_{1n}^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{g}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\right|^2\right)^{1/2}, \quad R_{4n} = 2R_{2n}^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{g}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\right|^2\right)^{1/2},$$

and $R_{5n} = 2R_{1n}^{1/2}R_{2n}^{1/2}$. For R_{1n}, as $n \to \infty$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Big|\frac{\delta_i}{\pi_i(\widehat{\boldsymbol{\alpha}})} g(X_i, Y_i, \boldsymbol{\beta}_0) - \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} g(X_i, Y_i, \boldsymbol{\beta}_0)\Big|^2 \\
&\quad \le \mathcal{C}\|\pi_i(\widehat{\boldsymbol{\alpha}}) - \pi_i(\boldsymbol{\alpha}_0)\|_\infty^2 \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}|g(X_i, Y_i, \boldsymbol{\beta}_0)|^2 \\
&\quad = o_p(1).
\end{aligned}$$

For R_{2n}, as $n \to \infty$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Big|\frac{\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})} \widehat{m}_g^0(U_i, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}}) - \frac{\delta_i - \pi_i(\boldsymbol{\alpha}_0)}{\pi_i(\boldsymbol{\alpha}_0)} m_g^0(U_i, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\Big|^2 \\
&\quad \le \frac{\mathcal{C}}{n}\sum_{i=1}^{n}\left|m_g^0(U_i, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}}) - m_g^0(U_i, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\right|^2 + o_p(1) \\
&\quad \le \frac{\mathcal{C}}{n}\sum_{i=1}^{n}\sup_{|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0| < \varrho}\left|\frac{\partial}{\partial \boldsymbol{\alpha}^\top} m_g^0(U_i, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\right|^2 |\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0|^2 + o_p(1) \\
&\quad = o_p(1).
\end{aligned}$$

Therefore, we have $R_{jn} = o_p(1)$ for $j = 3,4,5$. Then, the proof of Lemma 3 is completed. □

Lemma 4. *Suppose that Assumptions A, B and C hold; that the respondent probability model $\pi(U, Y, \boldsymbol{\alpha}_0)$ is correctly specified; and that $\widehat{\boldsymbol{\alpha}}$ is computed by the SEL approach. Then, for all positive $\varrho_n = o_p(1)$, we have*

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n} \frac{|\mathcal{G}_n(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) + \mathcal{G}(\boldsymbol{\beta}_0)|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|} = o_p(n^{-1/2}).$$

Proof. Let $\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \{\delta_i \pi_i^{-1}(\boldsymbol{\alpha}) g(X_i, Y_i, \boldsymbol{\beta})\}$. Consider the following decomposition

$$\begin{aligned}
&|\mathcal{G}_n(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) + \mathcal{G}(\boldsymbol{\beta}_0)| \\
&\quad \le |\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) + \mathcal{G}(\boldsymbol{\beta}_0)| \\
&\qquad + \left|\frac{1}{n}\sum_{i=1}^n \{\widehat{m}_g^0(U_i,\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})\}(\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right| \\
&\quad =: J_1 + J_2.
\end{aligned}$$

We now define

$$\Delta_n(\boldsymbol{\beta}, \pi_i(\boldsymbol{\alpha}) - \pi_i(\boldsymbol{\alpha}_0)) = -n^{-1}\sum_{i=1}^n \delta_i \pi_i^{-2}(\boldsymbol{\alpha}_0) g(X_i, Y_i, \boldsymbol{\beta})\{\pi_i(\boldsymbol{\alpha}) - \pi_i(\boldsymbol{\alpha}_0)\},$$

and consider the following decomposition:

$$\begin{aligned}
&\left|\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) + \mathcal{G}(\boldsymbol{\beta}_0)\right| \\
&\quad \le \left|\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) + \mathcal{G}(\boldsymbol{\beta}_0)\right| \\
&\qquad + \left|\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0) - \Delta_n(\boldsymbol{\beta}, \pi_i(\widehat{\boldsymbol{\alpha}}) - \pi_i(\boldsymbol{\alpha}_0))\right| \\
&\qquad + \left|\mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) - \Delta_n(\boldsymbol{\beta}_0, \pi_i(\widehat{\boldsymbol{\alpha}}) - \pi_i(\boldsymbol{\alpha}_0))\right| \\
&\qquad + \left|\Delta_n(\boldsymbol{\beta}, \pi_i(\widehat{\boldsymbol{\alpha}}) - \pi_i(\boldsymbol{\alpha}_0)) - \Delta_n(\boldsymbol{\beta}_0, \pi_i(\widehat{\boldsymbol{\alpha}}) - \pi_i(\boldsymbol{\alpha}_0))\right| \\
&\quad =: J_{11} + J_{12} + J_{13} + J_{14}.
\end{aligned}$$

Define $\mathcal{F} = \{\delta_i \pi_i^{-1}(\boldsymbol{\alpha}_0) g(X_i, Y_i, \boldsymbol{\beta}) : |\boldsymbol{\beta}-\boldsymbol{\beta}_0| \le \varrho\}$. Using the arguments of Van der Vaart and Wellner (1996), we can show that the class of functions $\mathcal{F}$ is Donsker because $\{g(X, Y, \boldsymbol{\beta}) : |\boldsymbol{\beta}-\boldsymbol{\beta}_0| \le \varrho\}$ is Donsker under assumption (A3), and $\pi_i^{-1}(\boldsymbol{\alpha}_0)$ and δ_i is uniformly bounded under Assumption (C2). Let $M_{ij}(\boldsymbol{\beta})$ denote the jth coordinates of $\delta_i \pi_i^{-1}(\boldsymbol{\alpha}_0) g(X_i, Y_i, \boldsymbol{\beta})$. Using Assumptions (A4) and (C2), we obtain

$$\begin{aligned}
&E\left\{\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho} |M_{ij}(\boldsymbol{\beta}) - M_{ij}(\boldsymbol{\beta}_0)|^2\right\} \\
&\quad = E\left\{\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho} \left|\frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}[g_j(X_i, Y_i, \boldsymbol{\beta}) - g_j(X_i, Y_i, \boldsymbol{\beta}_0)]\right|^2\right\} \le \mathcal{C}\varrho^{2s}
\end{aligned}$$

for some constants $s \in (0, 1]$. Therefore, we can show that the class of functions $\mathcal{F}$ is $\mathscr{L}_2(P)$ continuous. By applying Lemma 2.17 in Pakes and Pollard (1989), we have

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0) + \mathcal{G}(\boldsymbol{\beta}_0)| = o_p(n^{-1/2}),$$

as $n \to \infty$. Note that

$$\begin{aligned}
&\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \frac{|\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0)-\mathcal{G}(\boldsymbol{\beta})-\mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)+\mathcal{G}(\boldsymbol{\beta}_0)|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}\\
&\quad\le \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0)-\mathcal{G}(\boldsymbol{\beta})-\mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)+\mathcal{G}(\boldsymbol{\beta}_0)|\\
&\quad= o_p(n^{-1/2}).
\end{aligned}$$

By assumption, we have $|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0| = o_p(n^{-1/2+r})$ for every $0<r<1/2$. Moreover, we have $|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0| = o_p(n^{-1/4})$ with $r=1/4$. This, together with Assumptions (A1) and (C2), and nonignorable missing data mechanism, implies that

$$\begin{aligned}
&\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \frac{\Big|\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})-\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0)-\Delta_n(\boldsymbol{\beta},\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0))\Big|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}\\
&\quad\le \mathcal{C}\|\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0)\|_\infty^2 \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} |g(X_i,Y_i,\boldsymbol{\beta})|\\
&\quad= o_p(n^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \frac{\Big|\mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})-\mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)-\Delta_n(\boldsymbol{\beta}_0,\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0))\Big|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}\\
&\quad\le \mathcal{C}\|\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0)\|_\infty^2 \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} |g(X_i,Y_i,\boldsymbol{\beta}_0)|\\
&\quad= o_p(n^{-1/2}).
\end{aligned}$$

Next, for J_{14}, using Assumption (C2), we have

$$\begin{aligned}
&\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} |\Delta_n(\boldsymbol{\beta},\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0))-\Delta_n(\boldsymbol{\beta}_0,\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0))|\\
&\quad\le \mathcal{C}\|\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0)\|_\infty \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \left|\frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\{g_i(\boldsymbol{\beta})-g_i(\boldsymbol{\beta}_0)\}\right|\\
&\quad\le \mathcal{C}\|\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0)\|_\infty (J_{141}+J_{142}),
\end{aligned}$$

where

$$\begin{aligned}
J_{141} &= \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \left|\frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)}\{g_i(\boldsymbol{\beta})-g_i(\boldsymbol{\beta}_0)\} - E\{g_i(\boldsymbol{\beta})-g_i(\boldsymbol{\beta}_0)\}\right|,\\
J_{142} &= \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} |E\{g(X_i,Y_i,\boldsymbol{\beta})-g(X_i,Y_i,\boldsymbol{\beta}_0)\}|.
\end{aligned}$$

Since the class of functions $\mathcal{F}=\{\delta_i\pi_i^{-1}(\boldsymbol{\alpha}_0)g(X_i,Y_i,\boldsymbol{\beta}), |\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho\}$ is Donsker and $\mathscr{L}^2(P)$ is continuous, we obtain $J_{141}=o_p(n^{-1/2})$. For J_{142}, we have

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \frac{n^{1/2}|E\{g(X_i,Y_i,\boldsymbol{\beta})-g(X_i,Y_i,\boldsymbol{\beta}_0)\}|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|} = O_p(1).$$

Note that $|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0| = o_p(n^{-1/2+r})$ for every $0<r<1/2$. This, coupled with Assumption (C2), implies that

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le \varrho_n} \frac{\Big|\Delta_n(\boldsymbol{\beta},\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0))-\Delta_n(\boldsymbol{\beta}_0,\pi_i(\widehat{\boldsymbol{\alpha}})-\pi_i(\boldsymbol{\alpha}_0))\Big|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|} = o_p(n^{-1/2}).$$

Combining the above arguments leads to

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n}\frac{|\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})-\mathcal{G}(\boldsymbol{\beta})-\mathcal{G}_n^{IPW}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})+\mathcal{G}(\boldsymbol{\beta}_0)|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}=o_p(n^{-1/2}).$$

For J_2, we have

$$\begin{aligned}
&\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n}\frac{\left|\frac{1}{n}\sum_{i=1}^n\{\widehat{m}_g^0(U_i,\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})-\widehat{m}_g^0(U_i,\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})\}(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}\\
&\quad\le\sup{}^{**}\frac{\left|\frac{1}{n}\sum_{i=1}^n\{\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\widetilde{m}_g^0(U_i,\boldsymbol{\beta}_0)\}(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}\\
&\quad\le J_{21}+J_{22},
\end{aligned}$$

where

$$\begin{aligned}
J_{21}&=\sup{}^{**}\frac{\left|\frac{1}{n}\sum_{i=1}^n\left\{\partial_{\boldsymbol{\beta}}\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\partial_{\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta}_0)\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|},\\
J_{22}&=\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n}\frac{\left|\frac{1}{n}\sum_{i=1}^n\partial_{\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}.
\end{aligned}$$

Here $\sup^{**}$ is the supremum over all $|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n$ and $\|\widetilde{m}_g^0-m_g^0\|_\infty\le\varrho_n$ with $\varrho_n=o(1)$. For J_{21}, as $n\to\infty$, we obtain

$$\begin{aligned}
&\sup{}^{**}\frac{\left|\frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial}{\partial\boldsymbol{\beta}}\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta}_0)\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right|}{1+\mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|}\\
&\quad\le\mathcal{C}n^{-1/2}\sup{}^{**}\left|\frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial}{\partial\boldsymbol{\beta}}\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta})\right\}\frac{(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))}{\pi_i(\widehat{\boldsymbol{\alpha}})}\right|\\
&\qquad+\mathcal{C}n^{-1/2}\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n}\left|\frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta})-\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta}_0)\right\}\frac{(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))}{\pi_i(\widehat{\boldsymbol{\alpha}})}\right|\\
&\quad\le\mathcal{C}n^{-1/2}J_{211}+\mathcal{C}n^{-1/2}J_{212}.
\end{aligned}$$

where

$$\begin{aligned}
J_{211}&=\sup{}^{**}\left|\frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial}{\partial\boldsymbol{\beta}}\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta})\right\}\frac{(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))}{\pi_i(\widehat{\boldsymbol{\alpha}})}\right|,\\
J_{212}&=\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n}\left|\frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta})-\frac{\partial}{\partial\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta}_0)\right\}\frac{(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))}{\pi_i(\widehat{\boldsymbol{\alpha}})}\right|.
\end{aligned}$$

Define $\mathbb{I}_n^\pi=\mathbf{1}(\inf_{(u,y)\in\mathcal{U}\times\mathcal{Y}}\pi(u,y,\widehat{\boldsymbol{\alpha}})\ge\mathcal{C})$ and $\mathbb{I}_n^m=\mathbf{1}(\sup_{\boldsymbol{\beta}\in\mathcal{N}_\varrho}\|\widehat{m}_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})-m_g^0(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\|_\infty\le\varrho_n)$. It is easily verified that $\mathbb{I}_n^m\xrightarrow{p}1$ as $\varrho_n=o_p(1)$ and $\mathbb{I}_n^\pi\xrightarrow{p}1$. Since $\mathbb{I}_n^m\mathbb{I}_n^\pi|1-\delta_i\pi_i^{-1}(\widehat{\boldsymbol{\alpha}})|\le\max(1,|1-\mathcal{C}^{-1}|)$ is bounded, it follows from Assumptions (B3) and (C2) that

$$\begin{aligned}
&\sup{}^{**}\left|\frac{1}{n}\sum_{i=1}^n\mathbb{I}_n^m\mathbb{I}_n^\pi\left\{\partial_{\boldsymbol{\beta}}\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\partial_{\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta})\right\}\frac{(\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}}))}{\pi_i(\widehat{\boldsymbol{\alpha}})}\right|\\
&\quad\le\sup{}^{**}\frac{1}{n}\sum_{i=1}^n\mathbb{I}_n^m\mathbb{I}_n^\pi\left|\frac{\delta_i-\pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})}\right|\left|\partial_{\boldsymbol{\beta}}\widetilde{m}_g^0(U_i,\boldsymbol{\beta})-\partial_{\boldsymbol{\beta}}m_g^0(U_i,\boldsymbol{\beta})\right|\\
&\quad\le\mathcal{C}\varrho_n^\epsilon\frac{1}{n}\sum_{i=1}^n b(U_i).
\end{aligned}$$

Recalling $\varrho_n \xrightarrow{p} 0$, $\mathbb{I}_n^m \xrightarrow{p} 1$ and $\mathbb{I}_n^\pi \xrightarrow{p} 1$, together with $E[|b(U)|] < \infty$, we have $J_{211} = o_p(1)$.

For J_{212}, we have decomposition $J_{212} = J_{212}^1 + J_{212}^2$, where

$$\begin{aligned} J_{212}^1 &= \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n} \left| \frac{1}{n}\sum_{i=1}^n \left\{ \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}) - \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}_0) \right\} \frac{(\delta_i - \pi_i(\boldsymbol{\alpha}_0))}{\pi_i(\boldsymbol{\alpha}_0)} \right|, \\ J_{212}^2 &= \sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n} \left| \frac{1}{n}\sum_{i=1}^n \left\{ \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}) - \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}_0) \right\} \left\{ \frac{\delta_i}{\pi_i(\widehat{\boldsymbol{\alpha}})} - \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} \right\} \right|. \end{aligned}$$

Since the function class $\mathcal{F} = \{(\partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}) - \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}_0))(\delta_i - \pi_i(\boldsymbol{\alpha}_0))/\pi_i(\boldsymbol{\alpha}_0) : |\boldsymbol{\beta}-\boldsymbol{\beta}_0| \le \varrho_n\}$ is Glienko-Cantelli for some $\varrho_n > 0$ by Assumption B, we have $J_{212}^1 = o_p(1)$. It is easy to show from Assumption B that $J_{212}^2 = o_p(1)$. Therefore, we obtain $J_{21} = o_p(n^{-1/2})$. For J_{22}, by Assumption B, we have

$$\begin{aligned} &\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n} \frac{n^{1/2}\left|\frac{1}{n}\sum_{i=1}^n \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}}))/\pi_i(\widehat{\boldsymbol{\alpha}})\right|}{1 + \mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|} \\ &\quad \le \left|\frac{1}{n}\sum_{i=1}^n \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}_0)\right| \left| \frac{\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})} - \frac{\delta_i - \pi_i(\boldsymbol{\alpha})}{\pi_i(\boldsymbol{\alpha})} \right| \\ &\qquad + \left|\frac{1}{n}\sum_{i=1}^n \partial_{\boldsymbol{\beta}} m_g^0(U_i,\boldsymbol{\beta}_0) \frac{\delta_i - \pi_i(\boldsymbol{\alpha})}{\pi_i(\boldsymbol{\alpha})}\right| = o_p(1). \end{aligned}$$

Combining the above arguments yields

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\varrho_n} \frac{|\mathcal{G}_n(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta}) - \mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) + \mathcal{G}(\boldsymbol{\beta}_0)|}{1 + \mathcal{C}n^{1/2}|\boldsymbol{\beta}-\boldsymbol{\beta}_0|} = o_p(n^{-1/2}).$$

The proof of Lemma is completed. □

Lemma 5. *Suppose that Assumptions (A1) and C hold; that the respondent probability model* $\pi(U,Y,\boldsymbol{\alpha}_0)$ *is correctly specified; and that* $\widehat{\boldsymbol{\alpha}}$ *is computed by the SEL approach. Then, for* $\Lambda_n = \{\lambda : |\lambda| \le \mathcal{C}n^{-1/2}\}$, *we have* $\sup_{\boldsymbol{\beta}\in\mathcal{B},\lambda\in\Lambda_n,1\le i\le n} |\lambda^\top \widehat{g}_i(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})| \xrightarrow{p} 0$ *and* w.p.1, $\Lambda_n \subseteq \widehat{\Lambda}_n(\boldsymbol{\beta},\boldsymbol{\alpha})$ *for all* $\boldsymbol{\beta}\in\mathcal{B}$ *and* $\boldsymbol{\alpha}\in\mathcal{A}$.

Proof. Let $b_i = \sup_{\boldsymbol{\beta}\in\mathcal{B}} |g(X_i,Y_i,\boldsymbol{\beta})|$. Using the arguments of Owen (1990, Lemma 3), it follows from Assumption (A1) that $\max_{1\le i\le n} b_i = o_p(n^{1/2})$. On the other hand, we have

$$\frac{\delta_i}{\pi_i(\widehat{\boldsymbol{\alpha}})} = \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} \left\{ 1 - \frac{\partial\pi_i(\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}^\top}{\pi_i(\boldsymbol{\alpha}_0)} (\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o_p(n^{-1/2}) \right\},$$

and $\pi(U_i,Y_i,\boldsymbol{\alpha}) \ge \mathcal{C} > 0$ uniformly in i. Hence, we obtain

$$\begin{aligned} \sup_{\boldsymbol{\beta}\in\mathcal{B}} |\widehat{g}_i(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})| &\le \sup_{\boldsymbol{\beta}\in\mathcal{B}} |\delta_i\pi_i^{-1}(\widehat{\boldsymbol{\alpha}})g(X_i,Y_i,\boldsymbol{\beta})| + \sup_{\boldsymbol{\beta}\in\mathcal{B}} |(1 - \delta_i\pi_i^{-1}(\widehat{\boldsymbol{\alpha}}))\widehat{m}_g^0(U_i,\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})| \\ &\le O_p(1)b_i. \end{aligned}$$

This leads to $\sup_{\boldsymbol{\beta}\in\mathcal{B},\lambda\in\Lambda_n,1\le i\le n} |\lambda^\top \widehat{g}_i(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})| \le \mathcal{C}n^{-1/2}o_p(n^{1/2}) = o_p(1)$. □

Lemma 6. *Suppose that Assumptions (A1) and C hold; that the respondent probability model* $\pi(U,Y,\boldsymbol{\alpha}_0)$ *is correctly specified; and that* $\widehat{\boldsymbol{\alpha}}$ *is computed by the SEL approach. Then, we have* $|\mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\boldsymbol{\alpha}})| = O_p(n^{-1/2})$.

Proof. Let $\widehat{g}_i = \widehat{g}_i(\widehat{\boldsymbol{\beta}}_S, \widehat{\boldsymbol{\alpha}})$, $\widehat{\mathcal{G}}_n = \mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S, \widehat{\boldsymbol{\alpha}})$, and $\widetilde{\lambda} = -n^{-1/2}\widehat{\mathcal{G}}_n/|\widehat{\mathcal{G}}_n|$. By Lemma 5, we get $\max_{1\le i\le n} |\widetilde{\lambda}^\top \widehat{g}_i| \xrightarrow{p} 0$ and $\widetilde{\lambda} \in \widehat{\Lambda}_n(\boldsymbol{\beta}_S, \widehat{\boldsymbol{\alpha}})$ with probability one. Thus, we have $\rho_2(\dot{\lambda}^\top \widehat{g}_i) \ge -\mathcal{C}$ $(i = 1, \cdots, n)$ for any $\dot{\lambda}$ on the line joining $\widetilde{\lambda}$ and 0. By Lemma 3, we obtain $n^{-1}\sum_{i=1}^n \widehat{g}_i\widehat{g}_i^\top = O_p(1)$. By the second-order Taylor expansion, we have

$$\begin{aligned} \widehat{P}_n(\widehat{\boldsymbol{\beta}}_S, \widetilde{\lambda}, \widehat{\boldsymbol{\alpha}}) &= -\widetilde{\lambda}^\top \widehat{\mathcal{G}}_n + \frac{1}{2}\widetilde{\lambda}^\top \left[\frac{1}{n}\sum_{i=1}^n \rho_2(\dot{\lambda}^\top \widehat{g}_i)\widehat{g}_i\widehat{g}_i^\top\right]\widetilde{\lambda} \\ &\ge n^{-1/2}|\widehat{\mathcal{G}}_n| - \frac{\mathcal{C}}{2}\widetilde{\lambda}^\top\left[\frac{1}{n}\sum_{i=1}^n \widehat{g}_i\widehat{g}_i^\top\right]\widetilde{\lambda} \\ &\ge n^{-1/2}|\widehat{\mathcal{G}}_n| - \mathcal{C}n^{-1}. \end{aligned}$$

This, together with the constructions of $\widehat{\boldsymbol{\beta}}_S$ and $\widehat{\lambda}_S$, implies

$$\begin{aligned} n^{-1/2}|\widehat{\mathcal{G}}_n| - \mathcal{C}n^{-1} &\le \widehat{P}_n(\widehat{\boldsymbol{\beta}}_S, \widetilde{\lambda}, \widehat{\boldsymbol{\alpha}}) \\ &\le \widehat{P}_n(\widehat{\boldsymbol{\beta}}_S, \widehat{\lambda}_S, \widehat{\boldsymbol{\alpha}}) \le \widehat{P}_n(\boldsymbol{\beta}_0, \lambda(\boldsymbol{\beta}_0), \widehat{\boldsymbol{\alpha}}) \end{aligned} \tag{S2.1}$$

Also, by the second-order Taylor expansion and the proof of Theorem 2, we have

$$2n\widehat{P}_n(\boldsymbol{\beta}_0, \lambda(\boldsymbol{\beta}_0), \widehat{\boldsymbol{\alpha}}) = O_p(1).$$

Solving equation (S2.1) for $|\widehat{\mathcal{G}}_n|$ then gives $|\widehat{\mathcal{G}}_n| \le \mathcal{C}n^{-1/2} + O_p(n^{-1/2}) = O_p(n^{-1/2})$. By arguments similar to those used in Newey and Smith (2004), we have $\varepsilon_n|\widehat{\mathcal{G}}_n|^2 = O_p(n^{-1})$. The proof of Lemma is completed. □

S3. Proofs of Theorems 1–2

Proof of Theorem 1. We first prove that $\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0 = o_p(1)$. According to Theorem 3.1 in Newey and Smith (2004), to prove the consistency of the proposed semiparametric AIPW-GEL estimator, it suffices to prove $\sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta})| = o_p(1)$. Let $\mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n\{\delta_i\pi_i^{-1}(\boldsymbol{\alpha})g(X_i, Y_i, \boldsymbol{\beta})\}$ with $\pi_i(\boldsymbol{\alpha}) = \pi(U_i, Y_i, \boldsymbol{\alpha})$. Note that

$$\begin{aligned} \sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta})| &\le \sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta})| + \sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\sum_{i=1}^n \frac{\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})}\widehat{m}_g^0(U_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}})\right| \\ &\le \sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)| + \sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0) - \mathcal{G}(\boldsymbol{\beta})| \\ &\quad + \sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\sum_{i=1}^n \frac{\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})}\widehat{m}_g^0(U_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}})\right|. \end{aligned}$$

By the consistency of $\widehat{\boldsymbol{\alpha}}$ and Assumptions (A1) and (C2), it can be shown that

$$\begin{aligned} &\sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n^{IPW}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)| \\ \le\ &\mathcal{C}\|\pi_i(\widehat{\boldsymbol{\alpha}}) - \pi_i(\boldsymbol{\alpha}_0)\|_\infty \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i(\boldsymbol{\alpha}_0)} \sup_{\boldsymbol{\beta}\in\mathcal{B}} |g(X_i, Y_i, \boldsymbol{\beta})| = o_p(1), \end{aligned}$$

as $n \to \infty$. On the other hand, we obtain

$$\begin{aligned}
& \sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - \pi_i(\widehat{\boldsymbol{\alpha}})}{\pi_i(\widehat{\boldsymbol{\alpha}})} \widehat{m}_g^0(U_i,\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) \right| \\
\leq\ & \mathcal{C} \sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\widehat{m}_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - m_g^0(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\|_\infty + \sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - \pi_i(\boldsymbol{\alpha}_0)}{\pi_i(\boldsymbol{\alpha}_0)} m_g^0(U_i,\boldsymbol{\beta},\boldsymbol{\alpha}_0) \right| + o_p(1) \\
\leq\ & \mathcal{C} \sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\widehat{m}_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - m_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})\|_\infty + \mathcal{C} \sup_{\boldsymbol{\beta}\in\mathcal{B}} \|m_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - m_g^0(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\|_\infty \\
& + \sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - \pi_i(\boldsymbol{\alpha}_0)}{\pi_i(\boldsymbol{\alpha}_0)} m_g^0(U_i,\boldsymbol{\beta},\boldsymbol{\alpha}_0) \right| + o_p(1) \\
=\ & o_p(1).
\end{aligned}$$

The last equality holds because $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\widehat{m}_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - m_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})\|_\infty = o_p(1)$ by Lemma 1, and $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|m_g^0(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - m_g^0(\boldsymbol{\beta},\boldsymbol{\alpha}_0)\|_\infty \leq |\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0| \sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\partial m_g^0(\boldsymbol{\beta},\boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}^\top\|_\infty = o_p(1)$. Define $\mathcal{F} = \{\delta_i \pi_i^{-1}(\boldsymbol{\alpha}_0) g(X_i,Y_i,\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}\}$. The function class $\mathcal{F}$ is Glienko-Cantelli because $\{g(X,Y,\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}\}$ is Glivenko-Cantelli under Assumption (A2), and $\pi_i^{-1}(\boldsymbol{\alpha}_0)$ and δ_i are uniformly bounded. It follows from the Glienko-Cantelli theorem that $\sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n^{IPW}(\boldsymbol{\beta},\boldsymbol{\alpha}_0) - \mathcal{G}(\boldsymbol{\beta})| = o_p(1)$. Combining the above arguments leads to $\sup_{\boldsymbol{\beta}\in\mathcal{B}} |\mathcal{G}_n(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\boldsymbol{\beta})| = o_p(1)$. Using the arguments of Newey and Smith (2004), we have $\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0 = o_p(1)$.

Next we prove the asymptotic normality of $\widehat{\boldsymbol{\beta}}_S$. The proof is very similar to that of Theorem 2.2 in Parente and Smith (2011); hence we just sketch the main steps here. We first establish $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_S$ to $\boldsymbol{\beta}_0$. By triangle inequality, we have that $|\mathcal{G}(\widehat{\boldsymbol{\beta}}_S)| \leq |\mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\widehat{\boldsymbol{\beta}}_S)| + |\mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\boldsymbol{\alpha}})| + |\mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})|$. From Lemma 6, $|\mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\boldsymbol{\alpha}})| = O_p(n^{-1/2})$. By Lemma 2 and the central limit theorem, $n^{1/2}|\mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}})| = O_p(1)$. Using Lemma 4, together with $\mathcal{G}(\boldsymbol{\beta}_0) = 0$, shows that $n^{1/2}|\mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) - \mathcal{G}(\widehat{\boldsymbol{\beta}}_S)| \leq (1 + n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0|)o_p(1)$. This leads to $n^{1/2}|\mathcal{G}(\widehat{\boldsymbol{\beta}}_S)| \leq (1 + n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0|)o_p(1) + O_p(1)$. Additionally, by the smoothness of $\mathcal{G}(\boldsymbol{\beta})$, we have that $|\mathcal{G}(\widehat{\boldsymbol{\beta}}_S)| \geq \mathcal{C}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0|$. Therefore, $n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0| \leq (1 + n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0|)o_p(1) + O_p(1)$. Solving the above equation for $n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0|$ then gives $(1 - o_p(1))n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0| \leq O_p(1)$, which leads to the desired result $n^{1/2}|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0| = O_p(1)$. Define the linearization $\widehat{\mathcal{L}}_n(\boldsymbol{\beta},\lambda,\boldsymbol{\alpha}) = [-\Gamma(\boldsymbol{\beta} - \boldsymbol{\beta}_0)]^\top\lambda - \mathcal{G}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha})^\top\lambda - 0.5\lambda^\top V_1\lambda$. Applying Taylor's expansion of $\widehat{P}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S,\widehat{\boldsymbol{\alpha}})$ and triangle inequality leads to

$$\begin{aligned}
& \left|\widehat{P}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S,\widehat{\boldsymbol{\alpha}}) - \widehat{\mathcal{L}}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S,\widehat{\boldsymbol{\alpha}})\right| \\
\leq\ & |-[\mathcal{G}_n(\widehat{\boldsymbol{\beta}}_S,\widehat{\boldsymbol{\alpha}}) - \mathcal{G}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}}) - \Gamma(\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0)]^\top\widehat{\lambda}_S| \\
& + \left|\frac{1}{2}\widehat{\lambda}_S^\top\left(\sum_{i=1}^{n}\rho_2(\dot{\lambda}^\top\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S))\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S)\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S)^\top/n + V_1\right)\widehat{\lambda}_S\right|,
\end{aligned}$$

where $\dot{\lambda}$ lies between $\widehat{\lambda}_S$ and 0. Using arguments of Lemma A2 in Newey and Smith (2004), we can show that $\widehat{\lambda}_S = O_p(n^{-1/2})$. By Cauchy-Schwarz inequality, Lemmas 3 and 5, we obtain

$$\begin{aligned}
& \left|\widehat{\lambda}_S^\top\left(\sum_{i=1}^{n}\rho_2(\dot{\lambda}^\top\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S))\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S)\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S)^\top/n + V_1\right)\widehat{\lambda}_S\right| \\
\leq\ & \left|\widehat{\lambda}_S\right|^2\left|\sum_{i=1}^{n}\rho_2(\dot{\lambda}^\top\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S))\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S)\widehat{g}_i(\widehat{\boldsymbol{\beta}}_S,\widehat{\lambda}_S)^\top/n + V_1\right| = O_p(n^{-1})o_p(1) = o_p(n^{-1}).
\end{aligned}$$

Additionally, we can show

$$\begin{aligned} & |-[\mathcal{G}_n(\widehat{\boldsymbol\beta}_S,\widehat{\boldsymbol\alpha})-\mathcal{G}_n(\boldsymbol\beta_0,\widehat{\boldsymbol\alpha})-\Gamma(\widehat{\boldsymbol\beta}_S-\boldsymbol\beta_0)]| \\ \leq\ & |-[\mathcal{G}_n(\widehat{\boldsymbol\beta}_S,\widehat{\boldsymbol\alpha})-\mathcal{G}_n(\boldsymbol\beta_0,\widehat{\boldsymbol\alpha})-\mathcal{G}(\widehat{\boldsymbol\beta}_S)]|+|\Gamma(\widehat{\boldsymbol\beta}_S-\boldsymbol\beta_0)-\mathcal{G}(\widehat{\boldsymbol\beta}_S)| \\ \leq\ & (1+n^{1/2}|\widehat{\boldsymbol\beta}_S-\boldsymbol\beta_0|)o_p(n^{-1/2})+o_p(|\widehat{\boldsymbol\beta}_S-\boldsymbol\beta_0|)=o_p(n^{-1/2}). \end{aligned}$$

This leads to $|-[\mathcal{G}_n(\widehat{\boldsymbol\beta}_S,\widehat{\boldsymbol\alpha})-\mathcal{G}_n(\boldsymbol\beta_0,\widehat{\boldsymbol\alpha})-\Gamma(\widehat{\boldsymbol\beta}_S-\boldsymbol\beta_0)]^\top\widehat\lambda_S|=o_p(n^{-1})$. Combining above arguments, we have $|\widehat{P}_n(\widehat{\boldsymbol\beta}_S,\widehat\lambda_S,\widehat{\boldsymbol\alpha})-\widehat{\mathcal{L}}_n(\widehat{\boldsymbol\beta}_S,\widehat\lambda_S,\widehat{\boldsymbol\alpha})|=o_p(n^{-1})$. Now consider the problem $\min_{\boldsymbol\beta\in\mathcal{B}}\sup_{\lambda\in\mathcal{R}^r}\widehat{\mathcal{L}}_n(\boldsymbol\beta,\lambda,\widehat{\boldsymbol\alpha})$. Therefore, the following first order conditions are satisfied: $-\Gamma^\top\widetilde\lambda=0$ and $-\Gamma(\widetilde{\boldsymbol\beta}-\boldsymbol\beta_0)-\mathcal{G}_n(\boldsymbol\beta_0,\widehat{\boldsymbol\alpha})-V_1\widetilde\lambda=0$. Next, simple algebraic manipulations show $n^{1/2}(\widetilde{\boldsymbol\beta}-\boldsymbol\beta_0)=\{\Gamma^\top V_1^{-1}\Gamma\}^{-1}\Gamma^\top V_1^{-1}n^{1/2}\mathcal{G}_n(\boldsymbol\beta_0,\widehat{\boldsymbol\alpha})$. Consequently, by Lemma 1 and the asymptotic linear expansion of $\widehat{\boldsymbol\alpha}$, we have $n^{1/2}(\widetilde{\boldsymbol\beta}_S-\boldsymbol\beta_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma_S)$, where

$$\Sigma_S=(\Gamma^\top V_1^{-1}\Gamma)^{-1}\Gamma^\top V_1^{-1}V_2V_1^{-1}\Gamma(\Gamma^\top V_1^{-1}\Gamma)^{-1}.$$

A little more work similar to those of Parente and Smith (2011) shows $n^{1/2}(\widehat{\boldsymbol\beta}-\widetilde{\boldsymbol\beta})=o_p(1)$. The asymptotic normality of $\widehat{\boldsymbol\beta}_S$ is then established. This completes the proof of Theorem 1. □

Proof of Theorem 2. Let $\widehat\Omega(\boldsymbol\beta,\widehat{\boldsymbol\alpha})=n^{-1}\sum_{i=1}^n\widehat{g}_i(\boldsymbol\beta,\widehat{\boldsymbol\alpha})\widehat{g}_i(\boldsymbol\beta,\widehat{\boldsymbol\alpha})^\top$. It follows from arguments similar to those of Newey and Smith (2004) that any solution $\lambda(\boldsymbol\beta)$ to the first order conditions $\mathcal{G}_n(\boldsymbol\beta,\widehat{\boldsymbol\alpha})+\widehat\Omega(\boldsymbol\beta,\widehat{\boldsymbol\alpha})\lambda=0$ will maximize $\widehat{P}_n(\boldsymbol\beta,\lambda,\widehat{\boldsymbol\alpha})$ with respect to λ holding $\boldsymbol\beta$ fixed. Then $\lambda(\boldsymbol\beta)=-\widehat\Omega(\boldsymbol\beta,\widehat{\boldsymbol\alpha})^{-1}\mathcal{G}_n(\boldsymbol\beta,\widehat{\boldsymbol\alpha})$ solves the first order conditions. Since $\rho(\cdot)$ is twice differentiable, a second order Taylor expansion is exact, giving $\widehat{P}_n(\boldsymbol\beta,\lambda(\boldsymbol\beta),\widehat{\boldsymbol\alpha})=-\mathcal{G}_n(\boldsymbol\beta,\widehat{\boldsymbol\alpha})^\top\lambda(\boldsymbol\beta)-\frac{1}{2}\lambda(\boldsymbol\beta)^\top\widehat\Omega(\boldsymbol\beta,\widehat{\boldsymbol\alpha})\lambda(\boldsymbol\beta)=\frac{1}{2}\mathcal{G}_n(\boldsymbol\beta,\widehat{\boldsymbol\alpha})^\top\widehat\Omega(\boldsymbol\beta,\widehat{\boldsymbol\alpha})^{-1}\mathcal{G}_n(\boldsymbol\beta,\widehat{\boldsymbol\alpha})$. This, together with Lemmas 1 and 2, implies $2n\widehat{P}_n(\boldsymbol\beta_0,\lambda(\boldsymbol\beta_0),\widehat{\boldsymbol\alpha})\xrightarrow{\mathcal{L}}Q^\top\Omega Q$, where $Q\sim\mathcal{N}(0,I_r)$ and $\Omega=V_2^{1/2}V_1^{-1}V_2^{1/2}$. □

References

Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72**, 1667–1714.

Newey, W. and Smith, R. J. (2004). Higher-order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219–255.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90–120.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

Parente, P. M., and Smith, R. J. (2011). GEL methods for nonsmooth moment indicators. *Econometric Theory* **27**, 74–113.

Qin, J., Leung, D. and Shao, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling. *Journal of the American Statistical Association* **97**, 193–200.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica* **39**, 577–591.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Zhao, P. Y., Zhao, H., Tang, N. S. and Li, Z. H. (2017). Weighted composite quantile regression analysis for nonignorable missing data using nonresponse instrument. *Journal of Nonparametric Statistics* **29**, 189–212.