Empirical Likelihood for Estimating Equations with Nonignorably Missing Data

Niansheng Tang, Puying Zhao, and Hongtu Zhu

Yunnan University and University of North Carolina at Chapel Hill.

Supplementary Material

S1 Assumptions and Proofs of Theorems 1 – 7

Let $f(\cdot)$ be the probability density function of X and $G(X) = f(X)\exp\{g(X)\}\{1-\pi(X)\}$, where $g(X)$ is defined in (2.1). Take $\pi(x, y) = P(\delta = 1|X = x, Y = y), \pi(x) = P(\delta = 1|X = x)$, $m_\psi^0(x;\theta) = E\{\psi(Y, Z;\theta)|X = x, \delta = 0\}$, $m_0(X) = E(Y|X, \delta = 0)$, and $m_{Y\psi}(X) = E(Y\psi(Y, Z;\theta)|X, \delta = 0)$. The symbol ∂ denotes partial differentiation with respect to parameter θ.

Some regularity conditions are required for the proofs of Theorems 1 – 7.

(C1) The probability density function $f(x)$ is bounded away from ∞ on the support of X, and the second derivative of $f(x)$ is continuous and bounded.

(C2) The probability function $\pi(X, Y)$ satisfies $\min_i \pi(X_i, Y_i) \geq c_0 > 0$ a.s. for some positive constant c_0, and $\pi(X) = E(\pi(X, Y)|X) \neq 1$ a.s.

(C3) $E(Y^2)$ and $E\{\exp(2\gamma Y)\}$ are finite.

(C4) $\psi(\cdot;\theta)$ is twice continuously differentiable in the neighborhood of the true value θ_0, and $m_\psi(x;\theta)$ is twice continuously differentiable in the neighborhood of x.

(C5) $0 < E|\psi(Y, Z;\theta)|^2 < \infty$ and $0 < E|\alpha^T\partial_\theta\psi(Y, Z;\theta)|^2 < \infty$ for any constant vector α; $\partial_\theta\psi(\cdot;\theta)$ and $\psi^3(\cdot;\theta)$ are bounded by some integrable function $M(z)$ in the neighborhood of θ.

(C6) Matrices V_1, V_2, $\tilde{V}_1$, $\tilde{V}_{\rm v}$, and D_1 are positive definite, and $E\{\partial_\theta\psi(Y, Z;\theta)\}$ has full column rank p.

(C7) The kernel function $K(\cdot)$ is a probability density function such that

(i) is bounded and has compact support;

(ii) is symmetric with $\sigma^2 = \int \omega^2 K(\omega)d\omega < \infty$;

(iii) $K(\omega) \geq d_1$ for some $d_1 > 0$ in some closed interval centered at zero.

(C8) $nh \to \infty$ and $nh^4 \to 0$ as $n \to \infty$.

These assumptions are common in the missing data and nonparametric literatures. Conditions (C2) is similar to that used in Kim and Yu (2011); (C3) – (C6) are standard assumptions for empirical likelihood based inference with estimating equations; (C7) and (C8) are common in the nonparametric literature.

Lemma 1 *Suppose (C1)–(C8) hold. Then*

$$n^{-1/2}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i,\theta) \xrightarrow{\mathcal{L}} N(0,V_1),\ \ n^{-1}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i,\theta_0)^{\otimes 2} \xrightarrow{\mathcal{P}} V_2,\ \ n^{-1}\sum_{i=1}^n \partial_\theta \hat{\psi}_M(Y_i,Z_i,\theta_0) \xrightarrow{\mathcal{P}} \Gamma.$$

Proof We first proof $n^{-1/2}\sum\limits_{i=1}^n \hat{\psi}_M(Y_i,Z_i,\theta) \xrightarrow{\mathcal{L}} N(0,V_1)$. By the definition of $\hat{\psi}_M(Y_i,Z_i,\theta_0)$, we obtain the following decomposition

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\{\psi(Y_i,Z_i;\theta_0)-m_\psi^0(X_i;\theta_0)\}+\frac{1}{\sqrt{n}}\sum_{i=1}^n m_\psi^0(X_i;\theta_0)\\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_\psi(X_i;\theta_0)-m_\psi^0(X_i;\theta_0)\}\\
&:= I_1+I_2+I_3.
\end{aligned}$$

It is easily shown that $E\{\delta\exp(\gamma Y)|X\}=\exp\{g(X)\}(1-\pi(X))$ with $\pi(X)=E(\delta|X)$. Define $G(X)=f(X)\exp\{g(X)\}(1-\pi(X))$. Thus, using kernel regression method, we have $\hat{G}(X)=\sum_{j=1}^n \delta_i \exp(\gamma Y_j)K_h(X_j-X)$. Then, for I_3, we have

$$\begin{aligned}
I_3 &= \frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^n\delta_j\exp(\gamma Y_j)K_h(X_j-X_i)\{\psi(Y_j,Z_j;\theta)-m_\psi^0(X_j;\theta_0)\}}{G(X_i)}\\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_\psi(X_i;\theta_0)-m_\psi^0(X_i;\theta_0)\}\left\{1-\frac{\hat{G}(X_i)}{G(X_i)}\right\}\\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^n\delta_j\exp(\gamma Y_j)K_h(X_j-X_i)\{m_\psi^0(X_j;\theta_0)-m_\psi^0(X_i;\theta_0)\}}{G(X_i)}\\
&:= I_{31}+I_{32}+I_{33}.
\end{aligned}$$

We first derive asymptotic distribution of I_{31}.

Let $\mathcal{B}_\psi(X_j,X_i)=\delta_j\exp(\gamma Y_j)\{m_\psi^0(X_j;\theta_0)-m_\psi^0(X_i;\theta_0)\}$, $\mathcal{W}_j=\delta_j\exp(\gamma Y_j)\{\psi(Y_j,Z_j;\theta)-m_\psi^0(X_j;\theta_0)\}$, $\varphi_n(X_i)=\frac{1}{n}\sum_{j=1}^n K_h(X_j-X_i)\mathcal{W}_j$, $\mathcal{B}_n(X_i)=\frac{1}{n}\sum_{j=1}^n K_h(X_j-X_i)\mathcal{B}_\psi(X_j,X_i)$ and $S_j=(X_j,Y_j,\delta_j)$. Define a kernel function of the U statistic for all pair (i,j), $H(S_i,S_j)=\frac{1}{2}K_h(X_i-X_j)\{(1-\delta_j)\mathcal{W}_i/G(X_j)+(1-\delta_i)\mathcal{W}_j/G(X_i)\}$. By some tedious calculations, we obtain $E\{\mathcal{W}_j\}=0$, from which it can be shown that $E\{K_h(X_i-X_j)(1-\delta_i)\mathcal{W}_j/G(X_i)\}=0$. By the symmetry of U statistic $H(S_i,S_j)$, we have $E\{H(S_i,S_j)\}=0$.

On the other hand, we have

$$\begin{aligned}
E\{H(S_i,S_j)|S_j\} &= \frac{1}{2h}E\{K(X_i-X_j)\{\frac{(1-\delta_j)\mathcal{W}_i}{G(X_j)}+\frac{(1-\delta_i)\mathcal{W}_j}{G(X_i)}\}|S_j\}\\
&= \frac{1}{2h}\{\frac{1-\delta_j}{G(X_j)}E\{K(X_i-X_j)\mathcal{W}_i|S_j\}+\mathcal{W}_jE\{K(X_i-X_j)\frac{1-\delta_i}{G(X_i)}|S_j\}\}\\
&:= J_1+J_2.
\end{aligned}$$

It follows from $E\{\mathcal{W}_j|X_j\}=0$ that $J_1=0$. It is easily shown from the expression of J_2 that $J_2=\frac{\mathcal{W}_j}{2}\int K(u)\exp\{-g(hu+X_j)\}du$. Taking Taylor expansion of $\exp\{-g(hu+X_j)\}$ yields $J_2=\frac{1-\pi(X_j,Y_j)}{2\pi(X_j,Y_j)}\delta_j\{\psi(Y_j,Z_j;\theta)-m_\psi^0(X_j;\theta_0)\}+O_p(h^2)$. Then, we have

$$H_1(S_j):=E\{H(S_i,S_j|S_j)\}=\frac{\{1-\pi(X_j,Y_j)\}\delta_j\{\psi(Y_j,Z_j;\theta)-m_\psi^0(X_j;\theta_0)\}}{2\pi(X_j,Y_j)}\{1+O(h^2)\}.$$

Now, we prove that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\varphi_n(X_i)/G(X_i) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}H_1(S_i), \tag{S1.1}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{B}_n(X_i)/G(X_i) = O(h^2). \tag{S1.2}$$

To save space, we only prove (S1.1), (S1.2) can be similarly shown.

It can be shown that $n^{-1}\sum_{i=1}^{n}(1-\delta_i)\varphi_n(X_i)/G(X_i) = n^{-2}\sum_{i=1}^{n}H(S_i,S_i)+U_n$, where $U_n = \frac{2}{n^2}\sum_{i=1}^{n}\sum_{i<j}^{n}H(S_i,S_j)$. By the definition of $H(S_i,S_i)$, it is easy to obtain that $E\{H(S_i,S_i)\} = 0$. By the Law of Large Numbers, we obtain that $\frac{1}{n^2}\sum_{i=1}^{n}H(S_i,S_i)$ approximates $o(n^{-1})$ in probability, which indicates that it suffices to only consider statistic U_n in statistic $n^{-1}\sum_{i=1}^{n}(1-\delta_i)\varphi_n(X_i)/G(X_i)$. By direct calculation, we obtain $\zeta_1 := \mathrm{Var}\{H(S_j)\} = E\{(4\pi(X_i,Y_i))^{-1}(1-\pi(X_i,Y_i))\ (\psi(Y_i,Z_i;\theta_0)-m_\psi^0(X_i;\theta_0))^{\otimes 2}\}(1+O(h^2))$. Similarly,

$$\begin{aligned}
\zeta_2 &\triangleq \mathrm{Var}\{H(S_i,S_j)\}\\
&= \tfrac{1}{4}E\left\{K_h^2(X_i-X_j)\tfrac{(1-\delta_j)\mathcal{W}_i^{\otimes 2}}{G^2(X_j)}\right\} + \tfrac{1}{4}E\left\{K_h^2(X_i-X_j)\tfrac{(1-\delta_i)\mathcal{W}_j^{\otimes 2}}{G^2(X_i)}\right\}\\
&:= K_1 + K_2.
\end{aligned}$$

For K_1, we have

$$\begin{aligned}
K_1 &= \tfrac{1}{4}E\left\{E\{K_h^2(X_i-X_j)\tfrac{(1-\delta_j)\mathcal{W}_i^{\otimes 2}}{G^2(X_j)}|X_i,Y_i,\delta_i\}\right\}\\
&= \tfrac{1}{4}E\left\{\mathcal{W}_i^{\otimes 2}E\{E(K_h^2(X_i-X_j)\tfrac{(1-\delta_j)}{G^2(X_j)}|X_i,X_j,Y_i)|X_i,Y_i\}\right\}\\
&= \tfrac{1}{4}E\left\{\mathcal{W}_i^{\otimes 2}E\{\tfrac{K_h^2(X_i-X_j)}{G^2(X_j)}(1-\pi(X_j))|X_i\}\right\}\\
&= \tfrac{1}{4h}E\left\{\mathcal{W}_i^{\otimes 2}\tfrac{[1-\pi(X_i)]f(X_i)}{G^2(X_i)}\int K^2(u)du\right\} + o_p(1)\\
&= \tfrac{1}{4h}E\left\{\tfrac{(1-\pi(X,Y))^2(\psi(Y,Z;\theta_0)-m_\psi^0(X;\theta_0))^{\otimes 2}}{\pi(X,Y)(1-\pi(X))f(X)}\int K^2(u)du\right\} + o_p(1).
\end{aligned}$$

Similarly, for K_2, we obtain

$$K_2 = \frac{1}{4h}E\left\{\frac{(1-\pi(X,Y))^2(\psi(Y,Z;\theta_0)-m_\psi^0(X;\theta_0))^{\otimes 2}}{\pi(X,Y)(1-\pi(X))f(X)}\int K^2(u)du\right\} + o_p(1).$$

Combing the above two equations yields

$$\zeta_2 = \frac{1}{2h}E\left\{\frac{[1-\pi(X,Y)]^2[\psi(Y,Z;\theta_0)-m_\psi^0(X;\theta_0)]^{\otimes 2}}{\pi(X,Y)[1-\pi(X)]f(X)}\int K^2(u)du\right\} + o_p(1).$$

Define $\hat{U}_n = \frac{2}{n}\sum_{i=1}^{n}H_1(S_i)$, from which it can be shown that $E\hat{U}_n^2 = \frac{4}{n}\zeta_1$. Also, by the definition of U_n, we have $E(U_n^2) = \frac{4(n-2)\zeta_1}{n(n-1)} + \frac{2\zeta_2}{n(n-1)}$. Combining the above equations yields $E(U_n-\hat{U}_n)^2 = \frac{2\zeta_2}{n(n-1)} + O(n^{-2})$, from which it can be shown that

$$\begin{aligned}
U_n &= \hat{U}_n + \left\{\tfrac{2\zeta_2}{n(n-1)} + O(n^{-2})\right\}^{\frac{1}{2}}\\
&= \tfrac{1}{n}\sum_{i=1}^{n}\tfrac{\delta_i(\psi(Y_i,Z_i;\theta_0)-m_\psi^0(X_i;\theta_0))(1-\pi(X_i,Y_i))}{\pi(X_i,Y_i)}\{1+O(h^2)\} + O((n^2h)^{-\frac{1}{2}}).
\end{aligned}$$

This completes the proof of (S1.1). Using same arguments, we can prove (S1.2).

It follows from the results given in (S1.1) and (S1.2) that

$$I_{33} = o_p(1), \quad I_{31} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(1-\pi(X_i, Y_i))\delta_i(\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0))}{\pi(X_i, Y_i)} + o_p(1).$$

Using similar arguments given in Wang and Chen (2009), we obtain that $I_{32} = o_p(1)$. Combing the above results leads to

$$\begin{aligned}
\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta_0) &= \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \{\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0)\} \\
&\quad + \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{(1-\pi(X_i,Y_i))\delta_i(\psi(Y_i,Z_i;\theta_0) - m_\psi^0(X_i;\theta_0))}{\pi(X_i,Y_i)} \\
&\quad + \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} m_\psi^0(X_i; \theta_0) + o_p(1) \\
&= \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{\delta_i}{\pi(X_i,Y_i)} \{\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0)\} \\
&\quad + \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} m_\psi^0(X_i; \theta_0) + o_p(1).
\end{aligned} \tag{S1.3}$$

It is easily shown that $E\{\frac{\delta_i}{\pi(X_i,Y_i)}(\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0))\} = 0$. Again, since $m_\psi(X_i; \theta_0)$ is independent of $\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0)$, thus we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta_0) \xrightarrow{\mathcal{L}} N(0, V_1(\theta_0)),$$

where

$$\begin{aligned}
V_1(\theta_0) &= \mathrm{Var}\left\{\tfrac{\delta_i}{\pi(X_i,Y_i)}[\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0)]\right\} + E\left\{m_\psi^0(X_i; \theta_0) m_\psi^{0T}(X_i; \theta_0)\right\} \\
&= E\left\{\tfrac{\delta_i}{\pi^2(X_i,Y_i)}[\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0)]^{\otimes 2}\right\} + E\left\{m_\psi^0(X_i; \theta_0) m_\psi^{0T}(X_i; \theta_0)\right\} \\
&= E\left\{\tfrac{1}{\pi(X_i,Y_i)}[\psi(Y_i, Z_i; \theta_0) - m_\psi^0(X_i; \theta_0)]^{\otimes 2}\right\} + E\left\{m_\psi^0(X_i; \theta_0) m_\psi^{0T}(X_i; \theta_0)\right\}.
\end{aligned}$$

We now proof the results

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i, \theta_0) \hat{\psi}_M^T(Y_i, Z_i, \theta_0) \xrightarrow{\mathcal{P}} V_2, \quad \frac{1}{n} \sum_{i=1}^{n} \partial_\theta \hat{\psi}_M(Y_i, Z_i, \theta_0) \xrightarrow{\mathcal{P}} \Gamma.$$

Firstly, we prove that $\hat{m}_\psi(X; \theta)$ defined as (2.6) is a consistent estimator of $m_\psi^0(X; \theta) = E\{\psi(Y, Z; \theta)|X; \delta = 0\}$. In fact, it follows from the standard kernel regression theory that

$$\begin{aligned}
p \lim_{n\to\infty} \sum_{i=1}^{n} \omega_{i0}(X; \gamma) \psi(Y_i, Z_i; \theta) &= \tfrac{E\{\delta\psi(Y,Z;\theta)\exp(\gamma Y)|X\}}{E\{\delta \exp(\gamma y)|X\}} \\
&= \tfrac{E\{\pi(X,Y)\psi(Y,Z;\theta)\exp(\gamma Y)|X\}}{E\{\pi(X,Y)\exp(\gamma Y)|X\}} \\
&= \tfrac{E\{\frac{\psi(Y,Z;\theta)}{1+\exp(g(X)-\gamma Y)}|X\}}{E\{\frac{1}{1+\exp(g(X)-\gamma Y)}|X\}} \\
&= \tfrac{E\{[1-\delta]\psi(Y,Z;\theta)|X\}}{E\{[1-\delta]|X\}} \\
&= E(\psi(Y, Z; \theta)|X, \delta = 0).
\end{aligned} \tag{S1.4}$$

By (S1.4) and Law of Large Numbers, we have

$$\begin{aligned}
\tfrac{1}{n} \sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i, \theta_0) \hat{\psi}_M^T(Y_i, Z_i, \theta_0) &= \tfrac{1}{n} \sum_{i=1}^{n} \{\delta_i \psi(Y_i, Z_i; \theta) + (1-\delta_i) m_\psi^0(X_i; \theta)\}^{\otimes 2} + o_p(1) \\
&= \tfrac{1}{n} \sum_{i=1}^{n} [\delta_i \{\psi(Y_i, Z_i; \theta) - m_\psi^0(X_i; \theta)\} + m_\psi^0(X_i; \theta)]^{\otimes 2} + o_p(1) \\
&= V_2(\theta_0) + o_p(1).
\end{aligned}$$

Let $\hat{m}_{\partial\psi}(X;\theta,\gamma) = \sum_{i=1}^n \omega_{i0}(X;\gamma)\partial_\theta\psi(Y_i,Z_i;\theta)$ and $m^0_{\partial\psi}(X;\theta) = E\{\partial_\theta\psi(Y,Z;\theta)|X;\delta=0\}$, Then $p\lim_{n\to\infty}\sum_{i=1}^n \omega_{i0}(X;\gamma)\partial_\theta\psi(Y_i,Z_i;\theta) = E\{\partial_\theta\psi(Y,Z;\theta)|X;\delta=0\}$. Thus

$$\begin{aligned}
\tfrac{1}{n}\sum_{i=1}^n \partial_\theta\hat{\psi}_M(Y_i,Z_i,\theta_0) &= \tfrac{1}{n}\sum_{i=1}^n \partial_\theta\{\delta_i\psi(Y_i,Z_i;\theta)+(1-\delta_i)\hat{m}_\psi(X_i;\theta)\}\\
&= \tfrac{1}{n}\sum_{i=1}^n \{\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)+(1-\delta_i)m^0_{\partial\psi}(X_i;\theta)\}+o_p(1)\\
&= E\{\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)+(1-\delta_i)m^0_{\partial\psi}(X_i;\theta)\}+o_p(1).
\end{aligned}$$

We note that

$$\begin{aligned}
& E\{\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)+(1-\delta_i)m^0_{\partial\psi}(X_i;\theta)\}\\
= \;& E\Big\{E(\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)|X_i)+\mathrm{pr}(\delta_i=0|X_i)E\{\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)\exp(\gamma Y_i)|X_i\}/E\{\delta_i\exp(\gamma Y_i)|X_i\}\Big\}\\
= \;& E\Big\{E(\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)|X_i)+\mathrm{pr}(\delta_i=0|X_i)E(\partial_\theta\psi(Y_i,Z_i;\theta)|\delta_i=0,X_i)\Big\}\\
= \;& E\Big\{E(\delta_i\partial_\theta\psi(Y_i,Z_i;\theta)|X_i)+\mathrm{pr}(\delta_i=0|X_i)E\{(1-\delta_i)\partial_\theta\psi(Y_i,Z_i;\theta)|X_i\}/E\{(1-\delta_i)|X_i\}\Big\}\\
= \;& E\{\partial_\theta\psi(Y_i,Z_i;\theta)\}=\Gamma.
\end{aligned}$$

Therefore $n^{-1}\sum_{i=1}^n \partial_\theta\hat{\psi}_M(Y_i,Z_i,\theta_0) \xrightarrow{\mathcal{P}} \Gamma$.

Lemma 2 *Suppose (C1)–(C8) hold. Then, as $n\to\infty$, with probability tending to 1, $\hat{\ell}_M(\theta)$ attains its minimum at some point $\hat{\theta}_e$ in the interior of the ball $\parallel \theta-\theta_0 \parallel \le n^{-\frac{1}{3}}$, and the solutions $\hat{\theta}_e$ and $\hat{\lambda}_{n1}=\lambda_{n1}(\hat{\theta}_e)$ satisfy*

$$Q_{n1}(\hat{\theta}_e,\hat{\lambda}_{n1})=0 \quad \textit{and} \quad Q_{n2}(\hat{\theta}_e,\hat{\lambda}_{n1})=0.$$

Proof. Let $\theta=\theta_0+un^{-\frac{1}{3}}$, for $\theta\in\{\theta|\parallel\theta-\theta_0\parallel\le n^{-\frac{1}{3}}\}$, where $\parallel u\parallel=1$. Following the arguments of Owen (1990), it can be shown that

$$\lambda_{n1}(\theta)=\left\{\frac{1}{n}\sum_{i=1}^n\hat{\psi}_M(Y_i,Z_i;\theta)\hat{\psi}_M^T(Y_i,Z_i;\theta)\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^n\hat{\psi}_M(Y_i,Z_i;\theta)\right\}+o(n^{-\frac{1}{3}}) \quad (a.s.)$$

uniformly for $\theta\in\{\theta|\parallel\theta-\theta_0\parallel\le n^{-\frac{1}{3}}\}$. By this and Taylor expansion, we have (uniformly for

u)

$$\begin{aligned}
\hat{\ell}_M(\theta) &= \sum_{i=1}^n t_n^T \hat{\psi}_M(Y_i, Z_i;\theta) - \tfrac{1}{2}\sum_{i=1}^n (t_n^T\hat{\psi}_M(Y_i,Z_i;\theta))^2 + o(n^{1/3}) \quad (a.s.)\\
&= \tfrac{n}{2}\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M^T(Y_i,Z_i;\theta)\right)\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta)\hat{\psi}_M^T(Y_i,Z_i;\theta)\right)^{-1}\\
&\quad\times\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_{NP}(Y_i,Z_i;\theta)\right)^T + o(n^{1/3}) \quad (a.s.)\\
&= \tfrac{n}{2}\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta_0) + \tfrac{1}{n}\sum_{i=1}^n \partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta_0)un^{-1/3}\right)^T\\
&\quad\times\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta)\hat{\psi}_M^T(Y_i,Z_i;\theta)\right)^{-1}\\
&\quad\times\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta_0) + \tfrac{1}{n}\sum_{i=1}^n \partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta_0)un^{-1/3}\right) + o(n^{1/3}) \quad (a.s.)\\
&= \tfrac{n}{2}\left(O(n^{-1/2}(\log\log n)^{1/2}) + E(\partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta_0))un^{-1/3}\right)^T\\
&\quad\times\left(E(\hat{\psi}_M(Y_i,Z_i;\theta_0)\hat{\psi}_M^T(Y_i,Z_i;\theta_0))\right)^{-1}\\
&\quad\times\left(O(n^{-1/2}(\log\log n)^{1/2}) + E(\partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta))un^{-1/3}\right) + o(n^{1/3}) \quad (a.s.)\\
&\geq (c-\varepsilon)n^{1/3}, \quad (a.s.)
\end{aligned}$$

where $c-\varepsilon>0$ and c is the smallest eigenvalue of

$$\left[E\{\partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta_0)\}\right]^T\left[E(\hat{\psi}_M(Y_i,Z_i;\theta_0)\hat{\psi}_M^T(Y_i,Z_i;\theta_0))\right]^{-1}\left[E\{\partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta_0)\}\right].$$

Similarly,

$$\begin{aligned}
\hat{\ell}(\theta_0) &= \tfrac{n}{2}\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta)\right)^T\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta)\hat{\psi}_M^T(Y_i,Z_i;\theta)\right)^{-1}\\
&\quad\times\left(\tfrac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta)\right) + o(1) \quad a.s.\\
&= O(\log\log n). \quad a.s.
\end{aligned}$$

Since $\hat{\ell}_M(\theta)$ is a continuous function about θ as $\theta\in\{\theta|\parallel\theta-\theta_0\parallel\leq n^{-\frac{1}{3}}\}$, with probability tending to 1, $\hat{\ell}_M(\theta)$ has a minimum $\hat{\theta}_e$ in the interior of the ball and $\hat{\theta}_e$ and λ_{n1} satisfy

$$\frac{\partial\hat{\ell}_M(\theta)}{\partial\theta} = \sum_{i=1}^n \frac{\{\partial_\theta\hat{\psi}_M(Y_i,Z_i;\theta)\}^T\lambda_{n1}(\theta)}{1+\lambda_{n1}^T\hat{\psi}_M(Y_i,Z_i;\theta)} = 0.$$

$$\frac{\partial\hat{\ell}_M(\theta)}{\partial\lambda_{n1}} = \sum_{i=1}^n \frac{\hat{\psi}_M(Y_i,Z_i;\theta)}{1+\lambda_{n1}^T\hat{\psi}_M(Y_i,Z_i;\theta)} = 0.$$

Lemma 3. *Suppose (C1)–(C8) hold. Then*

$$n^{-1/2}\sum_{i=1}^n \Lambda_i(\theta_0)\xrightarrow{\mathcal{L}} N(0,V_{1,\mathrm{AU}}), \quad n^{-1}\sum_{i=1}^n \Lambda_i(\theta_0)\Lambda_i^T(\theta_0)\xrightarrow{\mathcal{P}} V_{2,\mathrm{AU}},$$

where

$$V_{2,\mathrm{AU}} = \begin{pmatrix} V_2 & D_1\\ D_1^T & D_2\end{pmatrix}$$

Proof. Let $\bar{A}_n = \frac{1}{n}\sum_{i=1}^n A(X_i)$, from the proof of Lemma 1, we obtain

$$\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta_0), \frac{1}{\sqrt{n}}\sum_{i=1}^n A(X_i)\right) = \mathrm{Cov}(I_1,\sqrt{n}\bar{A}) + \mathrm{Cov}(I_2,\sqrt{n}\bar{A}) + \mathrm{Cov}(I_3,\sqrt{n}\bar{A}).$$

By direct calculation, it can be obtained that

$$\begin{aligned}\mathrm{Cov}(I_1, \sqrt{n}\bar{A}) &= E\{\pi(X,Y)[\psi(Y,Z;\theta_0) - m_\psi^0(X;\theta_0)]A^T(X)\},\\ \mathrm{Cov}(I_2, \sqrt{n}\bar{A}) &= E\left\{m_\psi^0(X;\theta_0)A^T(X)\right\}.\end{aligned}$$

Using some of arguments employed in the proof of the Lemma 1, we can prove

$$\begin{aligned}\mathrm{Cov}(I_3, \sqrt{n}\bar{A}) &= \mathrm{Cov}(I_{31}, \sqrt{n}\bar{A}) + o(1)\\ &= -E\{\pi(X,Y)[\psi(Y,Z;\theta_0) - m_\psi^0(X;\theta_0)]A^T(X)\} + o(1).\end{aligned}$$

Then

$$\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta_0), \frac{1}{\sqrt{n}}\sum_{i=1}^n A(X_i)\right) = E\left\{m_\psi^0(X;\theta_0)A^T(X)\right\} + o(1).$$

Thus, by the Central Limit Theorem, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \Lambda_i(\theta_0) \xrightarrow{\mathcal{L}} N(0, V_{1,\mathrm{AU}}(\theta_0)),$$

Similarly , we can show that $\frac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i,Z_i;\theta_0)A^T(X_i) \xrightarrow{\mathcal{P}} E\left\{m_\psi^0(X;\theta_0)A^T(X)\right\}$. The Law of Large Numbers implies that $\frac{1}{n}\sum_{i=1}^n A(X_i)A^T(X_i) \xrightarrow{\mathcal{P}} E\left\{A(X)A^T(X)\right\}$. Then, it can be shown that

$$\frac{1}{n}\sum_{i=1}^n \Lambda_i(\theta_0)\Lambda_i^T(\theta_0) \xrightarrow{\mathcal{P}} V_{2,\mathrm{AU}}(\theta_0),$$

Lemma 4. *Let U be r-vector of random variables that satisfies $U \xrightarrow{\mathcal{L}} N(0, I_r)$, where I_r is the $r \times r$ identity matrix. Let P be a $r \times r$ nonnegative definite matrix with eigenvalues $l_1, \cdots, l_r$. Then, $U^TPU \xrightarrow{\mathcal{L}} l_1\chi_1^2 + \cdots + l_r\chi_r^2$, where χ_i^2's ($i = 1, \ldots, r$) are χ^2 random variables each with one degree of freedom.*

Proof. Since P is a $r \times r$ nonnegative definite matrix, thus there exists an orthogonal matrix Q such that $P = QDQ^T$, where $D = \mathrm{diag}(l_1, \ldots, l_r)$ is a diagonal matrix with diagonal elements $l_1, \cdots, l_r$. Let $\tilde{U} = Q^TU = (\tilde{U}_1, \ldots, \tilde{U}_r)^T$. Then, it follows from $U \xrightarrow{\mathcal{L}} N(0, I_r)$ that $\tilde{U} \xrightarrow{\mathcal{L}} N(0, I_r)$. Therefore,

$$U^TPU = U^TQDQ^TU = (Q^TU)^TD(Q^TU) = \tilde{U}^TD\tilde{U} = l_1\tilde{U}_1^2 + \cdots + l_r\tilde{U}_r^2,$$

where $\tilde{U}_i^2 \xrightarrow{\mathcal{L}} \chi_1^2$.

Lemma 5. *Suppose (C1)-(C8) hold. Then*

(i) when the parameter estimate for γ is compute from an independent survey,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_T(Y_i,Z_i,\theta) \xrightarrow{\mathcal{L}} N(0, \tilde{V}_1),$$

where $\tilde{V}_1 = V_1 + H^2V_r$.

(ii) when the parameter estimate for γ is obtained from a validation sample,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_T(Y_i,Z_i,\theta) \xrightarrow{\mathcal{L}} N(0, \tilde{V}_{\mathrm{v}}),$$

where $\tilde{V}_{\mathrm{v}} = \mathrm{Var}(\eta_{1i})$,

$$\eta_{1i} = m^0_\psi(X_i;\theta,\gamma_0) + \{\frac{r_i}{\nu}(1-\delta_i)+\delta_i\}\{\psi(Y_i,Z_i;\theta) - m^0_\psi(X_i;\theta,\gamma_0)\},$$

$m^0_\psi(X_i;\theta,\gamma) = \mathrm{pr}\lim_{n\to\infty} \hat{m}_\psi(X;\theta,\gamma)$, $\nu = E(r|\delta=0)$, *and* γ_0 *is the probability limit of* $\hat{\gamma}$.

Proof. **(i)** By the definition of $\hat{\psi}_T(Y_i,Z_i,\theta_0)$, we have the following decomposition

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_T(Y_i,Z_i;\theta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\{\psi(Y_i,Z_i;\theta) - m^0_\psi(X_i;\theta)\} + \frac{1}{\sqrt{n}}\sum_{i=1}^n m^0_\psi(X_i;\theta)\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n (1-\delta_i)\{\hat{m}_\psi(X_i;\theta) - m^0_\psi(X_i;\theta)\}\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n (1-\delta_i)\{\hat{m}_\psi(X_i;\theta,\hat{\gamma}) - \hat{m}_\psi(X_i;\theta)\}\\
&:= I_1+I_2+I_3+I_4,
\end{aligned}$$

where I_1, I_2, I_3 were defined as in Lemma 1. For I_4, taking a Taylor expansion of $\hat{m}_\psi(X_i;\theta,\hat{\gamma})$ at γ yields

$$I_4 = (\hat{\gamma}-\gamma)^T \frac{1}{\sqrt{n}}\sum_{i=1}^n (1-\delta_i)\frac{\partial \hat{m}_\psi(X_i;\theta,\gamma)}{\partial\gamma} \triangleq \sqrt{n}(\hat{\gamma}-\gamma)W,$$

where

$$\begin{aligned}
\frac{\partial \hat{m}_\psi(X_i;\theta,\gamma)}{\partial\gamma} &= \frac{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)Y_j\psi(Y_j,Z_j;\theta_0)}{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)}\\
&\quad - \frac{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)\psi(Y_j,Z_j;\theta_0)\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)Y_j}{\{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)\}^2}.
\end{aligned}$$

Let $m_0(X) = E(Y|X,\delta=0)$ and $m_{Y\psi}(X;\theta) = E(Y\psi(Y,Z;\theta)|X,\delta=0)$. Then, using kernel regression method, we obtain

$$\hat{m}_0(X_i) = \frac{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)Y_j}{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)}, \quad \hat{m}_{Y\psi}(X_i;\theta) = \frac{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)Y_j\psi(Y_j,Z_j;\theta)}{\sum_{j=1}^n \delta_j \exp(\gamma Y_j)K_h(X_j-X_i)}.$$

Thus, we have $\partial\hat{m}_\psi(X;\gamma)/\partial\gamma = \hat{m}_{Y\psi}(X;\theta) - \hat{m}_\psi(X;\theta,\gamma)\hat{m}_0(X)$. Let $\Delta_n(X) = \hat{G}(X) - G(X)$. Decomposition of W is given by

$$\begin{aligned}
W &= \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^n \delta_j\exp(\gamma Y_j)K_h(X_j-X_i)\{Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m^0_\psi(X_j;\theta)\}}{G(X_i)}\\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^n \delta_j\exp(\gamma Y_j)K_h(X_j-X_i)\{m_0(X_j)m^0_\psi(X_j;\theta) - m_0(X_i)m^0_\psi(X_i;\theta)\}}{G(X_i)}\\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\frac{m_0(X_i)m^0_\psi(X_i;\theta)\Delta_n(X_i)}{G(X_i)} - \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\frac{m_{Y\psi}(X_i,\theta)G(X_i)}{G^2(X_i)}\Delta_n(X_i)\\
&\quad - \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\frac{\hat{m}_{Y\psi}(X_i;\theta)\hat{G}(X_i) - m_{Y\psi}(X_i;\theta)G(X_i)}{G^2(X_i)}\Delta_n(X_i)\\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\frac{\hat{m}_{Y\psi}(X_i;\theta)\hat{G}(X_i)}{G^2(X_i)\hat{G}(X_i)}\Delta_n^2(X_i)\\
&\quad - \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\{\hat{m}_\psi(X_i;\theta)\hat{m}_0(X_i) - m_0(X_i)m^0_\psi(X_i;\theta)\}\\
&:= W_1+\cdots+W_7.
\end{aligned}$$

According to the same arguments as given in (A.4) of Wang and Rao (2002), we have $W_3 + W_4 + W_5 + W_6 = o_p(n^{-1/2})$. For W_7, we have

$$\begin{aligned} W_7 &= -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_\psi(X_i;\theta)\hat{m}_0(X_i) - m_0(X_i)m_\psi^0(X_i;\theta)\} \\ &= -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m_0(X_i)\{\hat{m}_\psi(X_i;\theta) - m_\psi^0(X_i;\theta)\} \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(\hat{m}_0(X_i) - m_0(X_i))\{\hat{m}_\psi(X_i;\theta) - m_\psi^0(X_i;\theta)\} \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m_\psi(X_i;\theta)(\hat{m}_0(X_i) - m_0(X_i)) \\ &:= W_{71} + W_{72} + W_{73}. \end{aligned}$$

Standard argument yields $W_{72} = o_p(n^{-1/2})$. According to the same arguments as given in I_3 of proof of Lemma A1, we have

$$\begin{aligned} W_{71} &= -\frac{1}{n}\sum_{j=1}^{n}\{\tfrac{\delta_j(1-\pi(X_j,Y_j))}{\pi(X_j,Y_j)}(\psi(Y_j,Z_j;\theta) - m_\psi^0(X_j;\theta))m_0(X_j)\} + o_p(n^{-1/2}), \\ W_{73} &= -\frac{1}{n}\sum_{j=1}^{n}\{\tfrac{\delta_j(1-\pi(X_j,Y_j))}{\pi(X_j,Y_j)}(Y_j - m_0(X_j))m_\psi^0(X_j;\theta)\} + o_p(n^{-1/2}). \end{aligned}$$

By the Law of Large Numbers, we have

$$\begin{aligned} W_{71} &\xrightarrow{P} -E\left\{\tfrac{\delta(1-\pi(X,Y))}{\pi(X,Y)}(\psi(Y,Z;\theta) - m_\psi^0(X;\theta))m_0(X)\right\} \\ &\quad = -E\{(1-\delta)(m_0(X)\psi(Y,Z;\theta) - m_0(X)m_\psi^0(X;\theta))\}, \\ W_{73} &\xrightarrow{P} -E\left\{\tfrac{\delta(1-\pi(X,Y))}{\pi(X,Y)}(Y - m_0(X))m_\psi^0(X;\theta)\right\} \\ &\quad = -E\{(1-\delta)(Ym_\psi(X;\theta) - m_0(X)m_\psi^0(X;\theta))\}. \end{aligned}$$

We now derive the asymptotic distribution of W_1. Note that

$$\begin{aligned} W_1 &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^{n}\delta_j\exp(\gamma Y_j)K_h(X_j-X_i)\{Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta)\}}{G(X_i)} \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\delta_j\left\{\exp(\gamma Y_j)(Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta))\right\}\frac{(1-\delta_i)K_h(X_j-X_i)}{G(X_i)} \\ &:= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}Q_{ij}. \end{aligned}$$

Denote $\check{W}_1 = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}E\{Q_{ij}|(X_j,Y_j,\delta_j)\}$, and write $W_1 = \check{W}_1 + W_1 - \check{W}_1$. From the arguments given in proof of Lemma 1 of Wang and Chen (2009), W_1 is dominated by $\check{W}_1$, whilst $W_1 - \check{W}_1$ is of smaller order. Thus, we just only consider the asymptotic properities of $\check{W}_1$. It can be shown that

$$\begin{aligned} &\check{W}_1 = \frac{1}{n}\sum_{j=1}^{n}\delta_j E\left\{\exp(\gamma Y_j)\{Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta)\}\frac{(1-\delta_i)K_h(X_j-X_i)}{G(X_i)}|X_j,Y_j\right\} \\ &= \frac{1}{n}\sum_{j=1}^{n}\delta_j E\left\{\exp(\gamma Y_j)(Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta))\frac{(1-\pi(X_i))K_h(X_j-X_i)}{G(X_i)}|X_j,Y_j\right\} \\ &= \frac{1}{n}\sum_{j=1}^{n}\delta_j\{\exp(\gamma Y_j)(Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta))\}\int\frac{(1-\pi(x))K_h(X_j-x)}{G(x)}f(x)dx \\ &= \frac{1}{n}\sum_{j=1}^{n}\delta_j\{\exp(\gamma Y_j)(Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta))\}\int\exp(-g(x))K_h(X_j-x)dx \\ &= \frac{1}{n}\sum_{j=1}^{n}\delta_j\{\exp(\gamma Y_j)(Y_j\psi(Y_j,Z_j;\theta) - m_0(X_j)m_\psi^0(X_j;\theta))\}\int\exp\{-g(hu+X_j)K(u)\}du. \end{aligned}$$

Taking the Taylor expansion of $\exp\{-g(hu+X_j)\}$ yields

$$\begin{aligned}\check{W}_1 &= \tfrac{1}{n}\sum_{j=1}^{n}\delta_j\{\exp(\gamma Y_j)(Y_j\psi(Y_j,Z_j;\theta)-m_0(X_j)m_\psi^0(X_j;\theta))\}\exp(-g(X_j))+O_p(h^2)\\ &= \tfrac{1}{n}\sum_{j=1}^{n}\delta_j\{\exp(\gamma Y_j-g(X_j)(Y_j\psi(Y_j,Z_j;\theta)-m_0(X_j)m_\psi^0(X_j;\theta))\}+O_p(h^2)\\ &= \tfrac{1}{n}\sum_{j=1}^{n}\tfrac{\delta_j(1-\pi(X_j,Y_j))}{\pi(X_j,Y_j)}(Y_j\psi(Y_j,Z_j;\theta)-m_0(X_j)m_\psi^0(X_j;\theta))+O_p(h^2).\end{aligned}$$

By the Law of Large Numbers, we obtain

$$\begin{aligned}\check{W}_1 \ &\xrightarrow{\mathcal{P}}\ E\left\{\tfrac{\delta(1-\pi(X,Y))}{\pi(X,Y)}(Y\psi(Y,Z;\theta)-m_0(X)m_\psi^0(X;\theta))\right\}\\ &= E\{(1-\delta)(Y\psi(Y,Z;\theta)-m_0(X)m_\psi^0(X;\theta))\}.\end{aligned}$$

According to the arguments given in the proof of Lemma 1, we have $W_2=o_p(n^{-1/2})$. Combing the above equations leads to

$$\begin{aligned}W \ \xrightarrow{\mathcal{P}}\ & E\{(1-\delta)(Y\psi(Y,Z;\theta)-m_0(X)m_\psi^0(X;\theta))\}\\ &-E\{(1-\delta)(m_0(X)\psi(Y,Z;\theta)-m_0(X)m_\psi^0(X;\theta))\}\\ &-E\{(1-\delta)(Ym_\psi^0(X;\theta)-m_0(X)m_\psi^0(X;\theta))\}\\ &=E\left\{(1-\delta)(Y-m_0(X))(\psi(Y,Z;\theta)-m_\psi^0(X;\theta))\right\}.\end{aligned}$$

Hence, I_4 is asymptotically equivalent to $\sqrt{n}(\hat{\gamma}-\gamma)E\{(1-\delta)(Y-m_0(X))(\psi(Y,Z;\theta)-m_\psi^0(X;\theta))\}$ and $I_4\xrightarrow{\mathcal{L}}N(0,H^2V_r)$, where $H=E\{(1-\delta)(Y-m_0(X))(\psi(Y,Z;\theta)-m_\psi^0(X;\theta))\}$.

From the decomposition of $\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{\psi}_T(Y_i,Z_i;\theta)$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\psi}_T(Y_i,Z_i;\theta)=I_1+I_2+I_3+I_4=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\psi}_M(Y_i,Z_i;\theta)+I_4.$$

Under the condition that $\hat{\gamma}$ is independent of $\hat{\psi}_M(Y_i,Z_i;\theta)$ and the Lemma 1, it can be shown that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\psi}_T(Y_i,Z_i;\theta)\xrightarrow{\mathcal{L}}N(0,\tilde{V}_1),$$

where $\tilde{V}_1=V_1+H^{\otimes 2}V_r$.

(ii) Denote

$$\hat{\psi}_T(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\{\delta_i\psi(Y_i,Z_i;\theta)+(1-\delta_i)\hat{m}_\psi(X_i;\theta,\gamma)\}+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{r_i}{\nu}\{\psi(Y_i,Z_i;\theta)-\hat{m}_\psi(X;\theta,\gamma)\}.$$

One may note that

$$\hat{\psi}_T(\hat{\gamma})=\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_T(Y_i,Z_i;\theta),\quad E\left\{\frac{\partial}{\partial\gamma}\hat{\psi}_T(\gamma)|\gamma=\gamma_0\right\}=0.$$

where γ_0 is the probability limit of $\hat{\gamma}$. According to Kim and Yu (2011) and Randles (1982), using $\sqrt{n}(\hat{\gamma}-\gamma)=O_p(1)$, by Taylor expansion, we have

$$\hat{\psi}_T(\hat{\gamma})=\hat{\psi}_T(\gamma_0)+o_p(n^{-1/2}).$$

Writing

$$m_\psi^0(X;\theta,\gamma)=\mathrm{pr}\lim_{n\to\infty}\hat{m}_\psi(X;\theta,\gamma)=E\{\delta\psi(Y,Z;\theta)\exp(\gamma Y)|X\}/E\{\delta\exp(\gamma Y)|X\},$$

we have

$$\begin{aligned}\hat{\psi}_T(\gamma_0) &= \tfrac{1}{n}\sum_{i=1}^{n}\left\{\delta_i\psi(Y_i,Z_i;\theta)+(1-\delta_i)\,m_\psi^0(X_i;\theta,\gamma_0)\right\}\\ &\quad+\tfrac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\tfrac{r_i}{\nu}\{\psi(Y_i,Z_i;\theta)-m_\psi^0(X_i;\theta,\gamma_0)\}\\ &\quad+\tfrac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(1-\tfrac{r_i}{\nu})\{\hat{m}_\psi(X_i;\theta,\gamma)-m_\psi^0(X_i;\theta,\gamma_0)\}\\ &:= V_{n1}(y)+V_{n2}(y)+V_{n3}(y).\end{aligned}$$

Since $nV_{n1}(y)+nV_{n2}(y)$ is sum of i.i.d random variables, it follows from the Central Limit Theorem that

$$\sqrt{n}\{V_{n1}(y)+V_{n2}(y)\}\xrightarrow{\mathcal{L}}N(0,\tilde{V}_{\mathrm{v}}),$$

where $\tilde{V}_{\mathrm{v}}=\mathrm{Var}(\eta_{1i})$,

$$\eta_{1i}=m_\psi^0(X_i;\theta,\gamma_0)+\{\frac{r_i}{\nu}(1-\delta_i)+\delta_i\}\{\psi(Y_i,Z_i;\theta)-m_\psi^0(X_i;\theta,\gamma_0)\},$$

$m_\psi^0(X_i;\theta,\gamma)=\mathrm{pr}\lim_{n\to\infty}\hat{m}_\psi(X;\theta,\gamma)$, $\nu=E(r|\delta=0)$ and γ_0 is the probability limit of $\hat{\gamma}$.

According to the Lemma 1 and the proof of Theorem 3 in Kim and Yu (2011), we can prove that $\sqrt{n}V_{n3}(y)=o_p(1)$ similarly. Then, Lemma 5 is completed.

Lemma 6. *Suppose (C1)-(C8) hold. Then*

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_T(Y_i,Z_i;\theta_0)\hat{\psi}_T^T(Y_i,Z_i;\theta_0)\xrightarrow{\mathcal{P}}V_2,\quad \frac{1}{n}\sum_{i=1}^{n}\partial_\theta\hat{\psi}_T(Y_i,Z_i;\theta_0)\xrightarrow{\mathcal{P}}\Gamma,$$

where $V_2=E\left\{[\delta_i\{\psi(Y_i,Z_i;\theta)-m_\psi^0(X_i;\theta)\}+m_\psi^0(X_i;\theta)]^{\otimes 2}\right\}$ *and* $\Gamma=E\{\partial_\theta\psi(Y,Z;\theta)\}$.

Proof. Since $\hat{\gamma}$ is a consistent estimator of γ, using the same methods given in the proof of Lemma 1 and Lemma 5, we can show that the result of Lemma 6 hold.

Lemma 7. *Suppose (C1)-(C8) hold. Then*

(i) when the parameter estimate for γ is compute from an independent survey,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\Lambda}_i(\theta_0)\xrightarrow{\mathcal{L}}N(0,\tilde{V}_{1,\mathrm{AU}}),\quad \frac{1}{n}\sum_{i=1}^{n}\tilde{\Lambda}_i(\theta_0)\tilde{\Lambda}_i^T(\theta_0)\xrightarrow{\mathcal{P}}V_{2,\mathrm{AU}},$$

where

$$\tilde{V}_{1,\mathrm{AU}}=\begin{pmatrix}\tilde{V}_1 & D_1\\ D_1^T & D_2\end{pmatrix},\quad V_{2,AU}=\begin{pmatrix}V_2 & D_1\\ D_1^T & D_2\end{pmatrix}$$

with $D_1=E\left\{m_\psi^0(X;\theta_0)A^T(X)\right\}$, $D_2=E\left\{A(X)A^T(X)\right\}$, $\tilde{V}_1$ *is defined in Theorem 4 and* V_2 *is defined in Theorem 2.*

(ii) when the parameter estimate for γ is obtained from a validation sample,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\Lambda}_i(\theta)\xrightarrow{\mathcal{L}}N(0,\tilde{V}_{\mathrm{v,AU}}),\quad ,\quad \frac{1}{n}\sum_{i=1}^{n}\tilde{\Lambda}_i(\theta_0)\tilde{\Lambda}_i^T(\theta_0)\xrightarrow{\mathcal{P}}V_{2,\mathrm{AU}},$$

where

$$\tilde{V}_{\mathrm{v,AU}}=\begin{pmatrix}\tilde{V}_{\mathrm{v}} & D_1\\ D_1^T & D_2\end{pmatrix},\quad V_{2,\mathrm{AU}}=\begin{pmatrix}V_2 & D_1\\ D_1^T & D_2\end{pmatrix}$$

$\tilde{V}_{\rm v}$ *is defined in Theorem 6.*

Proof. According to the proof of Lemma 3 and Lemma 5, and by the definition of $\hat{\psi}_T(Y_i, Z_i; \theta)$ and the consistency of $\hat{\gamma}$, we can show the Lemma hold.

Proof of Theorem 1 Let $\hat{\theta}_e$ and $\hat{\lambda}_{n1}$ be the solutions of the following equations

$$Q_{n1}(\theta, \lambda_{n1}) = 0, \quad Q_{n2}(\theta, \lambda_{n1}) = 0.$$

Taking the Taylor expansion of $Q_{n1}(\hat{\theta}_e, \hat{\lambda}_{n1})$ and $Q_{n2}(\hat{\theta}_e, \hat{\lambda}_{n1})$ at $(\theta_0, 0)$ yields

$$\begin{array}{rcl} 0 &=& Q_{n1}(\hat{\theta}_e, \hat{\lambda}_{n1}) = Q_{n1}(\theta_0, 0) + \frac{\partial Q_{n1}(\theta_0,0)}{\partial \theta}(\hat{\theta}_e - \theta_0) + \frac{\partial Q_{n1}(\theta_0,0)}{\partial \lambda_{n1}^T}(\hat{\lambda}_{n1} - 0) + o_p(\sigma_n), \\ 0 &=& Q_{n2}(\hat{\theta}_e, \hat{\lambda}_{n1}) = Q_{n2}(\theta_0, 0) + \frac{\partial Q_{n2}(\theta_0,0)}{\partial \theta}(\hat{\theta}_e - \theta_0) + \frac{\partial Q_{n2}(\theta_0,0)}{\partial \lambda_{n1}^T}(\hat{\lambda}_{n1} - 0) + o_p(\sigma_n), \end{array} \tag{S1.5}$$

where $\sigma_n =\| \hat{\theta}_e - \theta_0 \| + \| \hat{\lambda}_{n1} \|$. By direct calculation, we obtain

$$\begin{array}{rclrcl} \frac{\partial Q_{n1}(\theta_0,0)}{\partial \theta} &=& \frac{1}{n}\sum\limits_{i=1}^{n} \partial_\theta \hat{\psi}_M(Y_i, Z_i; \theta), & \frac{\partial Q_{n1}(\theta_0,0)}{\partial \lambda_{n1}^T} &=& -\frac{1}{n}\sum\limits_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta)\hat{\psi}_M^T(Y_i, Z_i; \theta) \\ \frac{\partial Q_{n2}(\theta_0,0)}{\partial \theta} &=& 0, & \frac{\partial Q_{n2}(\theta_0,0)}{\partial \lambda_{n1}^T} &=& \frac{1}{n}\sum\limits_{i=1}^{n} \partial_\theta \hat{\psi}_M(Y_i, Z_i; \theta). \end{array}$$

Then, from (S1.5), we have

$$\left(\begin{array}{c} \hat{\lambda}_{n1} \\ \hat{\theta}_e - \theta_0 \end{array} \right) = S_n^{-1} \left(\begin{array}{c} -Q_{n1}(\theta_0, 0) + o_p(\sigma_n) \\ o_p(\sigma_n) \end{array} \right), \tag{S1.6}$$

where

$$S_n = \left(\begin{array}{cc} -\frac{1}{n}\sum\limits_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta)\hat{\psi}_M^T(Y_i, Z_i; \theta) & \frac{1}{n}\sum\limits_{i=1}^{n} \partial_\theta \hat{\psi}_M(Y_i, Z_i; \theta) \\ \frac{1}{n}\sum\limits_{i=1}^{n} \partial_\theta \hat{\psi}_M^T(Y_i, Z_i; \theta) & 0 \end{array} \right).$$

From Lemma 1, we obtain

$$S_n \xrightarrow{\mathcal{P}} S = \left(\begin{array}{cc} -V_2(\theta_0) & \Gamma \\ \Gamma^T & 0 \end{array} \right).$$

From Lemma 1, we have $Q_{n1}(\theta_0, 0) = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta) = O_p(n^{-1/2})$, which yields $\sigma_n = O_p(n^{-1/2})$. Thus, it can be shown from (S1.6) and the expression of S^{-1} that

$$\sqrt{n}(\hat{\theta}_e - \theta_0) = -\Sigma_1 \Gamma^T V_2^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta) + o_p(1).$$

Since $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta) \xrightarrow{\mathcal{L}} N(0, V_1)$, we have $\sqrt{n}(\hat{\theta}_e - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_e)$, where $\Sigma_e = \Sigma_1 \Gamma^T V_2^{-1} V_1 V_2^{-1} \Gamma \Sigma_1$ and $\Sigma_1 = (\Gamma^T V_2^{-1} \Gamma)^{-1}$.

Proof of Theorem 2 Using the same arguments as done in Zhou, Wan and Wang (2008), we have

$$\begin{array}{rcl} \hat{\ell}_M(\theta_0) &=& n^{-1/2}\sum\limits_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta) \left\{ n^{-1}\sum\limits_{i=1}^{n} \hat{\psi}_M(Y_i, Z_i; \theta)\hat{\psi}_M^T(Y_i, Z_i; \theta) \right\}^{-1} \\ && \times n^{-1/2}\sum\limits_{i=1}^{n} \hat{\psi}_M^T(Y_i, Z_i; \theta) + o_p(1). \end{array}$$

From Lemma 1, we have $\xi_n = V_1^{-1/2}\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\psi}_M(Y_i, Z_i;\theta)\} \xrightarrow{\mathcal{L}} N(0, I)$. Therefore, it follows from Lemma A5 that

$$\hat{\ell}_M(\theta_0) \quad = \quad \xi_n^T V_1^{1/2} V_2^{-1} V_1^{1/2} \xi_n + o_p(1) \xrightarrow{\mathcal{L}} \varrho_1 \chi_1^2 + \cdots + \varrho_q \chi_q^2,$$

where χ_i^2's $(i = 1, \ldots, q)$ are the χ^2 random variables each with one degree of freedom, and the weights ϱ_i $(i = 1, \cdots, q)$ are eigenvalues of matrix $V_2^{-1}V_1$.

Proof of Theorem 3 **(i)** Let $\hat{\theta}_{ae}$ and $\hat{\lambda}_{n2}$ be the solutions of the following equations

$$M_{n1}(\theta, \lambda_{n2}) = 0, \quad M_{n2}(\theta, \lambda_{n2}) = 0 \;\text{ and }\; M_{n3}(\theta, \lambda_{n2}) = 0,$$

where $\lambda_{n2} = (\lambda_{n21}^T, \lambda_{n22}^T)^T$ and $\hat{\lambda}_{n2} = (\hat{\lambda}_{n21}^T, \hat{\lambda}_{n22}^T)^T$. Taking the Taylor expansion of $M_{n1}(\hat{\theta}_{ae}, \hat{\lambda}_{n2})$ at $(\theta_0, 0)$ yields

$$0 = M_{n1}(\theta_0, 0) + \frac{\partial M_{n1}(\theta_0, 0)}{\partial \theta}(\hat{\theta}_{ae} - \theta_0) + \frac{\partial M_{n1}(\theta_0, 0)}{\partial \lambda_{n21}^T}(\hat{\lambda}_{n21} - 0) + \frac{\partial M_{n1}(\theta_0, 0)}{\partial \lambda_{n22}^T}(\hat{\lambda}_{n22} - 0) + o_p(\sigma_n).$$

Similarly, taking the Taylor expansion of $M_{n2}(\hat{\theta}_{ae}, \hat{\lambda}_{n2})$ and $M_{n3}(\hat{\theta}_{ae}, \hat{\lambda}_{n2})$ at $(\theta_0, 0)$ yields

$$\begin{aligned} 0 &= M_{n2}(\theta_0, 0) + \tfrac{\partial M_{n2}(\theta_0, 0)}{\partial \theta}(\hat{\theta}_{ae} - \theta_0) + \tfrac{\partial M_{n2}(\theta_0, 0)}{\partial \lambda_{n21}^T}(\hat{\lambda}_{n21} - 0) + \tfrac{\partial M_{n2}(\theta_0, 0)}{\partial \lambda_{n22}^T}(\hat{\lambda}_{n22} - 0) + o_p(\sigma_n), \\ 0 &= M_{n3}(\theta_0, 0) + \tfrac{\partial M_{n3}(\theta_0, 0)}{\partial \theta}(\hat{\theta}_{ae} - \theta_0) + \tfrac{\partial M_{n3}(\theta_0, 0)}{\partial \lambda_{n21}^T}(\hat{\lambda}_{n21} - 0) + \tfrac{\partial M_{n3}(\theta_0, 0)}{\partial \lambda_{n22}^T}(\hat{\lambda}_{n22} - 0) + o_p(\sigma_n), \end{aligned}$$

where $\sigma_n = \| \hat{\theta} - \theta_0 \| + \| \hat{\lambda}_{n2} \|$. By direct calculation, we obtain

$$\frac{\partial M_{n1}(\theta_0, 0)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^n \partial_\theta \hat{\psi}_M(Y_i, Z_i;\theta), \quad \frac{\partial M_{n1}(\theta_0, 0)}{\partial \lambda_{n21}^T} = -\frac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i, Z_i;\theta)\hat{\psi}_M^T(Y_i, Z_i;\theta),$$

$$\frac{\partial M_{n1}(\theta_0, 0)}{\partial \lambda_{n22}^T} = -\frac{1}{n}\sum_{i=1}^n A(X_i)\hat{\psi}_M^T(Y_i, Z_i;\theta), \quad \frac{\partial M_{n2}(\theta_0, 0)}{\partial \lambda_{n21}^T} = -\frac{1}{n}\sum_{i=1}^n A(X_i)\hat{\psi}_M^T(Y_i, Z_i;\theta),$$

$$\frac{\partial M_{n2}(\theta_0, 0)}{\partial \lambda_{n22}^T} = -\frac{1}{n}\sum_{i=1}^n A(X_i)A(X_i)^T, \quad \frac{\partial M_{n2}(\theta_0, 0)}{\partial \lambda_{n21}^T} = \frac{1}{n}\sum_{i=1}^n \{\partial_\theta \hat{\psi}_M(Y_i, Z_i;\theta)\}^T,$$

$$\frac{\partial M_{n2}(\theta_0, 0)}{\partial \lambda_{n22}^T} = 0, \quad \frac{\partial M_{n2}(\theta_0, 0)}{\partial \theta} = 0, \quad \frac{\partial M_{n3}(\theta_0, 0)}{\partial \theta} = 0.$$

Then, we have

$$\begin{pmatrix} \hat{\lambda}_{n21} \\ \hat{\lambda}_{n22} \\ \hat{\theta} - \theta \end{pmatrix} = S_n^{-1} \begin{pmatrix} -M_{n1}(\theta, 0) + o_p(\sigma_n) \\ -M_{n2}(\theta, 0) + o_p(\sigma_n) \\ o_p(\sigma_n) \end{pmatrix},$$

where

$$S_n = \begin{pmatrix} -\frac{1}{n}\sum\limits_{i=1}^n \hat{\psi}_M(Y_i, Z_i;\theta)\hat{\psi}_M^T(Y_i, Z_i;\theta) & -\frac{1}{n}\sum\limits_{i=1}^n \hat{\psi}_M(Y_i, Z_i;\theta)A^T(X_i) & \frac{1}{n}\sum\limits_{i=1}^n \partial_\theta \hat{\psi}_M(Y_i, Z_i;\theta) \\ -\frac{1}{n}\sum\limits_{i=1}^n A(X_i)\hat{\psi}_M^T(Y_i, Z_i;\theta) & -\frac{1}{n}\sum\limits_{i=1}^n A(X_i)A^T(X_i) & 0 \\ \frac{1}{n}\sum\limits_{i=1}^n \partial_\theta \hat{\psi}_M^T(Y_i, Z_i;\theta) & 0 & 0 \end{pmatrix}.$$

From Lemma 1 and Lemma 3, it follows that

$$S_n \xrightarrow{\mathcal{P}} \begin{pmatrix} -V_2 & -D_1 & \Gamma \\ -D_1^T & -D_2 & 0 \\ \Gamma^T & 0 & 0 \end{pmatrix} := S = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B}^T & 0 \end{pmatrix},$$

where $\mathcal{B} = (\Gamma^T, 0^T)^T$, $\mathcal{C} = -\mathcal{B}^T \mathcal{A}^{-1} \mathcal{B}$ and

$$\mathcal{A} = \begin{pmatrix} -V_2 & -D_1 \\ -D_1^T & -D_2 \end{pmatrix}.$$

From Lemma 1, we have $M_{n1}(\theta_0, 0) = \frac{1}{n}\sum_{i=1}^n \hat{\psi}_M(Y_i, Z_i; \theta) = O_p(n^{-1/2})$, $M_{n2}(\theta_0, 0) = \frac{1}{n}\sum_{i=1}^n A(X_i) = O_p(n^{-1/2})$, which yields $\sigma_n = O_p(n^{-1/2})$. Thus, it can be shown that

$$\sqrt{n}(\hat{\theta}_{ae} - \theta_0) = -\mathcal{C}^{-1}\mathcal{B}^T \mathcal{A}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \Lambda_i(\theta) + o_p(1).$$

Since $\frac{1}{\sqrt{n}}\sum_{i=1}^n \Lambda_i(\theta_0) \xrightarrow{\mathcal{L}} N(0, V_{1,\mathrm{AU}})$, we have $\sqrt{n}(\hat{\theta}_{ae} - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_{ae})$.

(ii) Similar to the proof of Theorem 2, we have

$$\hat{\ell}_{AU}(\theta_0) = \{\frac{1}{\sqrt{n}}\sum_{i=1}^n \Lambda_i(\theta_0)\} V_{2,\mathrm{AU}}^{-1} \{\frac{1}{\sqrt{n}}\sum_{i=1}^n \Lambda_i(\theta_0)\}^T + o_p(1).$$

From Lemma 1, we have $\eta_n = V_{1,\mathrm{AU}}^{-1/2} \frac{1}{\sqrt{n}}\sum_{i=1}^n \Lambda_i(\theta) \xrightarrow{\mathcal{L}} N(0, I)$. From Lemma 4, Combing the above equations yields

$$\hat{\ell}_{AU}(\theta_0) = \eta_n^T V_{1,\mathrm{AU}}^{1/2} V_{2,\mathrm{AU}}^{-1} V_{1,\mathrm{AU}}^{1/2} \eta_n + o_p(1) \xrightarrow{\mathcal{L}} \varrho_1 \chi_1^2 + \cdots + \varrho_{r+q}\chi_{r+q}^2,$$

where χ_i^2's $(i = 1, \ldots, r+q)$ are the χ^2 random variables each with one degree of freedom, and the weights ϱ_i $(i = 1, \ldots, r+q)$ are eigenvalues of matrix $V_{2,\mathrm{AU}}^{-1} V_{1,\mathrm{AU}}$.

Proof of Theorem 4 – Theorem 7 Using the same methods given in the proofs of Theorems 1 – 3 and the above Lemma 1 – Lemma 7, we can prove that Theorems 4 – 7 hold.

S2 The weighted estimating equations with nonignorably missing response data

In practice, it is common that initial nonresponders are followed up and are given a second opportunity to respond to the survey. Let $R_1 = (R_{11}, \cdots, R_{n1})$ and $R_2 = (R_{12}, \cdots, R_{n2})$ be vectors of indicator functions for the observation (Y_i, Z_i), where $R_{ik} = I(Y_i, Z_i$ is observed at time $k)$ for $k = 1, 2$. Note that if (Y_i, Z_i) is already observed at the first time, that is $(R_{i1} = 1)$, no effort will be made to observe it again, that is $(R_{i2} = 0)$. The two sets of missingness probabilities are defined as $\pi_{i1} = P(R_{i1} = 1|X_i, Y_i)$ and $\pi_{i2} = P(R_{i2} = 1|X_i, Y_i, R_{i1} = 0)$, where X_i is Z_i or a subset of Z_i. Based on this twice survey, Troxel, Lipsitz, and Brennan (1997) (denoted as TLB below) proposed the following weighted estimating function

$$G(\beta) = \sum_{i=1}^n (R_{i1} + R_{i2})\pi_i^{-1} \dot{\mu}_i^{\mathrm{T}} W_i^{-1}(Y_i - \mu_i), \tag{S2.7}$$

where $\pi_i = \pi_{i1} + (1 - \pi_{i1})\pi_{i2}$, $\mu_i = E(Y_i|X_i) = g(\beta^T X_i)$, $\dot{\mu}_i = \partial\mu_i/\partial\beta$, and $W_i = Var(Y_i)$. It can be shown that $E\{G(\beta)\} = 0$ under the specified nonignorable missing data mechanism assumption. Let $\hat{\beta}^{\mathrm{TLB}}$ be an estimator of (S2.7).

S3 Algorithms for computing MELE and EL-basd CI

The algorithm for computing MELEs under unknown γ includes the following two steps:

1. Use the Levenberg-Marquardt algorithm to solve Equation (3.7) to get an estimated tilting parameter $\hat{\gamma}$.
2. For the estimated tilting parameter $\hat{\gamma}$, optimize the proposed LELRFs to get our proposed MELEs.

The grid search algorithm for constructing the EL-based CIs of θ is given as follows:

1. Arbitrarily choose an interval, which contains the true values θ_0;
2. Use the Levenberg-Marquardt algorithm to solve (3.7) to calculate an estimated tilting parameter $\hat{\gamma}$;
3. Given a search step length, we evaluate LELRFs (i.e., $\hat{\ell}_T(\theta_0)$) at each search point in the given interval, and find the grid-point $\hat{\theta}_0$ such that $\hat{\ell}_T(\hat{\theta}_0) \leq \hat{c}_\alpha$, which indicates that $\hat{\theta}_0$ is just the upper (or lower) bound of the EL-based CI. Moreover, $\hat{c}_\alpha$ is the $1-\alpha$ quantile of the distribution of LELRFs $\hat{\ell}_T(\theta_0)$. In practice, $\hat{c}_\alpha$ can be calculated by using Monte Carlo simulations.

References

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

Kim, J. K. and Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association* **106**, 157–165.

Wang, D. and Chen, S. X. (2009). Empirical likelihood for estimating equations with missing values. *The Annals of Statistics* **37**, 490–517.

Wang, Q. H. and Rao, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statistics* **30**, 896–924.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics* **10**, 462–474.

Zhou, Y., Wan, A. T. K. and Wang, X. J. (2008). Estimating equations inference with missing data. *Journal of the American Statistical Association* **103**, 1187–1199.

Niansheng Tang
Department of Statistics, Yunnan University,
Kunming 650091, China

E-mail: nstang@ynu.edu.cn

Puying Zhao
Department of Statistics, Yunnan University,

Kunming 650091, China
E-mail: pyzhao@live.cn

Hongtu Zhu
Department of Biostatistics,
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA
E-mail: hzhu@bios.unc.edu