A Weighted Composite Likelihood Approach for Analysis of Survey Data under Two-Level Models

Grace Y. Yi[1], JNK Rao[2] and Haocheng Li[3]

1. Department of Statistics and Actuarial Science, University of Waterloo

2. School of Mathematics and Statistics, Carleton University

3. Departments of Oncology and Community Health Sciences, University of Calgary

Supplementary Material

S1 Appendix A: Regularity Conditions

Let $\boldsymbol{\psi}_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik}) = B_{jk}\mathbf{s}_{ijk}$, then $\mathbf{U}_w(\boldsymbol{\theta}) = \sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i}\boldsymbol{\psi}_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})$, $\boldsymbol{\theta} \in \Theta \subset R^p$, where p is the dimension of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_0$ be the value such that

$$E_\xi E_d\{\mathbf{U}_w(\boldsymbol{\theta}_0)\} = \mathbf{0},$$

and

$$h_{ijk}(y_{ij}, y_{ik}) = \sup_{\boldsymbol{\theta}\in\Theta} ||\boldsymbol{\psi}_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})|| \text{ for the triples } (i, j, k),$$

where $||\cdot||$ is the L_1 norm. We assume the following regularity conditions. Some of these conditions are somewhat parallel to those in Carrillo, Chen and Wu (2010) and Shao (2003, Lemma 5.3) for one-level models, but additional conditions and more complex derivations are required here due to the accommodation of the two-level models with survey weights.

(1). Θ is a compact subset of the Euclidean space R^p.

(2). $\sup_{(i,j,k)} E_\xi\{h^2_{ijk}(Y_{ij}, Y_{ik})\} < \infty$ and $\sup_{1\le i\le N} E_\xi\{||\mathbf{Y}_i||\} < \infty$, where $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{iM_i})^{\mathrm{T}}$.

(3). For any given $c > 0$ and a given sequence $\{\mathbf{y}_i\}$ satisfying $||\mathbf{y}_i|| \le c$, the sequence of functions in $\boldsymbol{\theta}$, $\{\boldsymbol{\psi}_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})\}$, is equicontinuous on Θ.

(4). Define

$$\Delta_T(\boldsymbol{\theta}) = E_\xi E_d\{T^{-1}\mathbf{U}_w(\boldsymbol{\theta})\},$$

where $T = \sum_{i=1}^N M_i(M_i - 1)/2$. For any $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that

$$\inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon} ||\Delta_T(\boldsymbol{\theta})|| > \delta_\epsilon.$$

(5). There exists a $\widehat{\boldsymbol{\theta}}_w \in \Theta$ such that $\mathbf{U}_w(\widehat{\boldsymbol{\theta}}_w) = \mathbf{0}$.

(6). For any variable V_{ijk}, write

$$\overline{V} = \frac{1}{T}\sum_{i=1}^{N}\sum_{1\le j<k\le M_i} V_{ijk}.$$

If variables V_{ijk} satisfy $\sum_{i=1}^{N}\sum_{1\le j<k\le M_i} V_{ijk}^2/T = O_\xi(1)$, then

$$\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k, j,k\in s(i)} w_{jk|i}V_{ijk} - \overline{V}$$

converges to 0 in design probability as $n \to \infty$.

(7). When the number of clusters in a sample approaches infinity, the number of clusters in the corresponding population will tend to infinity as well. That is, if $n \to \infty$, then $N \to \infty$.

(8).

$$\frac{N\ \sup_{i\le N} M_i(M_i-1)}{\sum_{i\le N} M_i(M_i-1)} < \infty \quad \text{as } N \to \infty.$$

S2 Appendix B: Establishment of Consistency of the Estimator $\widehat{\boldsymbol{\theta}}_w$

First, we present the following results which are useful for proving Lemma 1. Suppose A and B are nonempty sets of real numbers. Define

$$A + B = \{x + y : x \in A, y \in B\}$$

and

$$A - B = \{x - y : x \in A, y \in B\}.$$

Then

$$\sup(A + B) = \sup A + \sup B \tag{S2.1}$$

and

$$\sup(A - B) = \sup A - \inf B. \tag{S2.2}$$

Proof: First, we note that by definition, $A + B$ is bounded from above if and only if A and B are bounded from above. Thus, $\sup(A + B)$ exists if and only if $\sup A$ and $\sup B$ exist. In this case, for any $x \in A$ and $y \in B$, we have

$$x + y \le \sup A + \sup B,$$

suggesting that $\sup A + \sup B$ is an upper bound of $A + B$. Therefore,

$$\sup(A + B) \le \sup A + \sup B.$$

To show that $\sup(A+B) \geq \sup A + \sup B$, we consider any $\epsilon > 0$. By definition of supermum, there exist $x_0 \in A$ and $y_0 \in B$ such that

$$x_0 > \sup A - \epsilon/2 \text{ and } y_0 > \sup B - \epsilon/2.$$

Thus,

$$x_0 + y_0 > \sup A + \sup B - \epsilon.$$

By the arbitrariness of ϵ, this inequality implies that

$$\sup(A+B) \geq \sup A + \sup B.$$

Therefore, (S2.1) follows. Result (S2.2) follows from (S2.1) and the fact that $\sup(-B) = -\inf B$.

Lemma 1: Under the regularity conditions in Appendix A, we have

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right|\right| \xrightarrow{p} 0 \quad \text{as } n \to \infty, \tag{S2.3}$$

where "p" denotes convergence in probability with respect to joint model ξ and sampling design d.

Proof: For ease of exposition, we consider the case where the ψ_{ijk} are scalar; the derivation for the vector case follows in the same manner. The proof consists of six parts.

Part 1:

By Assumption 3 of equicontinuity and Assumption 1 of compactness, we know that for any given $c > 0$ and any components $\{y_{ij}, y_{ik}\}$ of any fixed $\mathbf{y}_i = (y_{i1}, \cdots, y_{im_i})^{\mathrm{T}}$ with $||\mathbf{y}_i|| \leq c$, $\psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})$ is uniformaly equicontinuous in Θ. Therefore, for a given $\epsilon > 0$, there exists $\delta = \delta(c, \epsilon) > 0$ such that whenever $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{O}$ with $||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| < \delta$, we have

$$|\psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})| \leq \frac{\epsilon}{K}$$

for some $K \geq 2$ and components $\{y_{ij}, y_{ik}\}$ of a given $\mathbf{y}_i = (y_{i1}, \cdots, y_{im_i})^{\mathrm{T}}$ with $||\mathbf{y}_i|| \leq c$. Consequently,

$$\sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta, ||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||<\delta}\left|\{\psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})\}I(||\mathbf{y}_i|| \leq c)\right| \leq \frac{\epsilon}{K}. \tag{S2.4}$$

For any $\boldsymbol{\theta} \in \Theta$, let

$$\mathcal{O}(\boldsymbol{\theta}) = \{\boldsymbol{\theta}^* \in \Theta : ||\boldsymbol{\theta}^* - \boldsymbol{\theta}|| < \delta\}$$

be the δ-neighborhood of $\boldsymbol{\theta}$. Then $\{\mathcal{O}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ is an open covering of Θ with

$$\cup_{\boldsymbol{\theta}\in\Theta}\mathcal{O}(\boldsymbol{\theta}) \supset \Theta.$$

By the compactness of Θ, there exists a finite subcovering of Θ. That is, there exist a finite number of points $\boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_F^* \in \Theta$ such that

$$\cup_{r=1}^F \mathcal{O}(\boldsymbol{\theta}_r^*) \supset \Theta. \tag{S2.5}$$

For any $r = 1, \cdots, F$, we have that by (S2.4),

$$\sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2 \in \mathcal{O}(\boldsymbol{\theta}_r^*)} \left| \{\psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})\} I(||\mathbf{y}_i|| \le c) \right| \le \frac{\epsilon}{K}. \tag{S2.6}$$

As a result, we obtain

$$\begin{aligned}
& \frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik}) \cdot I(||\mathbf{y}_i|| \le c) \\
& \quad -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik}) \cdot I(||\mathbf{y}_i|| \le c) \\
= & \frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \left\{ \sup_{\boldsymbol{\theta}_1\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \inf_{\boldsymbol{\theta}_2\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik}) \right\} \cdot I(||\mathbf{y}_i|| \le c) \\
= & \frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \left\{ \sup_{\boldsymbol{\theta}_1\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) + \sup_{\boldsymbol{\theta}_2\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \{-\psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})\} \right\} \cdot I(||\mathbf{y}_i|| \le c) \\
= & \frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \left[\sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \{\psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})\} \right] \cdot I(||\mathbf{y}_i|| \le c) \\
\le & \frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \left\{ \sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})| \right\} \cdot I(||\mathbf{y}_i|| \le c) \\
\le & \frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \left\{ \sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}_1; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_r^*; y_{ij}, y_{ik})| \right. \\
& \left. + \sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}_r^*; y_{ij}, y_{ik}) - \psi_{ijk}(\boldsymbol{\theta}_2; y_{ij}, y_{ik})| \right\} \cdot I(||\mathbf{y}_i|| \le c) \\
\le & \left(\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \right) \left(\frac{2\epsilon}{K} \right),
\end{aligned} \tag{S2.7}$$

where the last inequality is due to (S2.6).

Part 2: By Assumption 2, we denote $c_0 = \sup_{(i,j,k)} E_\xi\{h_{ijk}^2(Y_{ij}, Y_{ik})\}$ and $c_1 = \sup_i E_\xi\{||\mathbf{Y}_i||\}$, where c_0 and c_1 are finite constants. For the given $c > 0$ and (i, j, k), using the Cauchy-Schwartz Inequality and the Markov inequality yields

$$\begin{aligned}
E_\xi\{h_{ijk}(Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c)\} & \le [E_\xi\{h_{ijk}^2(Y_{ij}, Y_{ik})\}]^{1/2} [E_\xi\{I(||\mathbf{Y}_i|| > c)\}]^{1/2} \\
& \le [\sup_i E_\xi\{h_{ijk}^2(Y_{ij}, Y_{ik})\}]^{1/2} \left[\frac{\sup_i E_\xi(||\mathbf{Y}_i||)}{c} \right]^{1/2} \\
& = c_0^{1/2} c_1^{1/2} c^{-1/2} \\
& = O(c^{-1/2})
\end{aligned} \tag{S2.8}$$

where $I(\cdot)$ is the indicator function.

For the given ϵ, c and any $\mathcal{O}(\boldsymbol{\theta}_r^*)$, $r = 1, \cdots, F$, using the Markov inequality and (S2.8) gives

$$\begin{aligned}
& P_{\xi d}\left\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})| \cdot I(||\mathbf{Y}_i|| > c) > \frac{\epsilon}{2}\right\} \\
\leq \quad & P_{\xi d}\left\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} h_{ijk}(Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) > \frac{\epsilon}{2}\right\} \\
\leq \quad & \left(\frac{2}{\epsilon}\right) E_{\xi d}\left\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} h_{ijk}(Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c)\right\} \\
= \quad & \left(\frac{2}{\epsilon}\right) E_{\xi}\left\{\frac{1}{T}\sum_{i=1}^{N} \sum_{1\leq j<k\leq M_i} h_{ijk}(Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c)\right\} \\
\leq \quad & \left(\frac{2}{\epsilon}\right) \sup_{i,j,k} E_{\xi}\left\{h_{ijk}(Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c)\right\} \\
= \quad & O(c^{-1/2}). \qquad\qquad \text{(S2.9)}
\end{aligned}$$

Noting that

$$-\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik}) = \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \{-\psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})\} \leq \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})|,$$

and using (S2.9), we obtain that

$$\begin{aligned}
& P_{\xi d}\left\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c)\right. \\
& \qquad \left. -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) > \epsilon\right\} \\
= \quad & P_{\xi d}\left[\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \left\{\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) - \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\right\}\right. \\
& \left. \cdot I(||\mathbf{Y}_i|| > c) > \epsilon\right] \\
\leq \quad & P_{\xi d}\left[\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \cdot 2 \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})| \cdot I(||\mathbf{Y}_i|| > c) > \epsilon\right] \\
= \quad & P_{\xi d}\left[\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \cdot \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})| \cdot I(||\mathbf{Y}_i|| > c) > \frac{\epsilon}{2}\right] \\
\to \quad & 0 \quad \text{as } c \to \infty. \qquad\qquad \text{(S2.10)}
\end{aligned}$$

Note that Assumption 6 with V_{ijk} taken as 1 gives

$$P_d\left(T^{-1}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} > K/2\right) \to 0 \quad \text{as } n \to \infty. \qquad \text{(S2.11)}$$

Then, using basic probability properties and (S2.7), we obtain

$$
\begin{aligned}
& P_{\xi d}\Bigg\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) \\
& \quad -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) > \epsilon\Bigg\} \\
\leq \; & P_{\xi d}\Bigg\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) \\
& \quad -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) > \epsilon\Bigg\} \\
& + P_{\xi d}\Bigg\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| \leq c) \\
& \quad -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| \leq c) > \epsilon\Bigg\} \\
\leq \; & P_{\xi d}\Bigg\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) \\
& \quad -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) > \epsilon\Bigg\} \\
& + P_{\xi d}\Bigg\{\left(\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i}\right)\left(\frac{2\epsilon}{K}\right) > \epsilon\Bigg\} \quad (\text{by}(S2.7)) \\
= \; & P_{\xi d}\Bigg\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) \\
& \quad -\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) I(||\mathbf{Y}_i|| > c) > \epsilon\Bigg\} \\
& + P_d\Bigg\{\left(\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i}\right) > K/2\Bigg\} \\
\to \; & 0 \qquad\qquad (S2.12)
\end{aligned}
$$

as $n \to \infty$ and $c \to \infty$, where the zero limit follows from (S2.10) and (S2.11).

Part 3:

For the given ϵ and an open subset $\mathcal{O}(\boldsymbol{\theta}_r^*), r = 1, \cdots, F$, we define

$$
Z_i = \sum_{j<k}\Big[\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) - E_\xi\Big\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\Big\}\Big]
$$

and write it as $\sum_{j<k} U_{i,jk}$ for ease of exposition. Then $\{Z_i : i = 1, \cdots, N\}$ is an independent sequence of random variables and $E_\xi(Z_i) = 0$ for any $i \geq 1$.

For the given $\varepsilon > 0$, we apply Chebyshev's inequality

$$\begin{aligned} P_\xi\left(\left|\frac{1}{\{\sum_{i=1}^N M_i(M_i-1)/2\}}\sum_{i=1}^N Z_i\right| > \varepsilon\right) &\le \frac{1}{\varepsilon^2}E_\xi\left[\left\{\frac{2}{\sum_{i=1}^N M_i(M_i-1)}\sum_{i=1}^N Z_i\right\}^2\right] \\ &= \frac{1}{\varepsilon^2}\left[\frac{2}{\sum_{i=1}^N M_i(M_i-1)}\right]^2\sum_{i=1}^N \mathrm{Var}_\xi(Z_i). \end{aligned}$$

We have $\mathrm{Var}_\xi(Z_i) = \mathrm{Var}_\xi\left(\sum_{j<k} U_{i,jk}\right) = \sum_{j_1<k_1, j_2<k_2} \mathrm{Cov}_\xi(U_{i,j_1k_1}, U_{i,j_2k_2})$. For any i, j, k we write

$$\mathrm{Var}_\xi(U_{i,jk}) = \mathrm{Var}_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\} \le E_\xi[\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\}^2]$$

and, since $|\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})| \le \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})|$, we obtain

$$\begin{aligned} E_\xi[\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\}^2] &\le E_\xi[\{\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})|\}^2] \\ &= E_\xi\{h_{ijk}^2(Y_{ij}, Y_{ik})\} \\ &\le \sup_{ijk} E_\xi\{h_{ijk}^2(Y_{ij}, Y_{ik})\} \\ &\le d \end{aligned}$$

for some constant d using Assumption 2. Hence by Cauchy-Schwartz' inequality, we obtain

$$|\mathrm{Cov}_\xi(U_{i,j_1k_1}, U_{i,j_2k_2})| \le \{\mathrm{Var}_\xi(U_{i,j_1k_1})\}^{1/2}\{\mathrm{Var}_\xi(U_{i,j_2k_2})\}^{1/2} \le d$$

for any i, j_1, j_2, k_1, k_2 and, so $\sup_{i\le N} \mathrm{Var}_\xi(Z_i) \le d \sup_{i\le N}\left\{\frac{M_i(M_i-1)}{2}\right\}^2$. Therefore,

$$\begin{aligned} P_\xi\left(\left|\frac{2}{\sum_{i=1}^N M_i(M_i-1)}\sum_{i=1}^N Z_i\right| > \varepsilon\right) &\le \frac{1}{\varepsilon^2}\left\{\frac{2}{\sum_{i=1}^N M_i(M_i-1)}\right\}^2 N d \sup_{i\le N}\left\{\frac{M_i(M_i-1)}{2}\right\}^2 \\ &= \frac{d}{\varepsilon^2}\left\{\frac{N\sup_i M_i(M_i-1)}{\sum_{i=1}^N M_i(M_i-1)}\right\}^2\frac{1}{N}. \end{aligned}$$

By Assumption 8, we thus have $\frac{2}{\sum_{i=1}^N M_i(M_i-1)}\sum_{i=1}^N Z_i \xrightarrow{p} 0$ as $N\to\infty$. That is,

$$\begin{aligned} &P_\xi\left[\left|\frac{1}{T}\sum_{i=1}^N\sum_{1\le j<k\le M_i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) - \frac{1}{T}\sum_{i=1}^N\sum_{1\le j<k\le M_i} E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\}\right| > \frac{\epsilon}{2}\right] \\ \to\ & 0 \quad \text{as } N\to\infty. \qquad \text{(S2.13)} \end{aligned}$$

Note that $|\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})| \le h_{ijk}(y_{ij}, y_{ik})$, then using Assumptions 2 and 6 we obtain

$$\begin{aligned} &P_d\left\{\left|\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k, j,k\in s(i)} w_{jk|i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik}) - \frac{1}{T}\sum_{i=1}^N\sum_{1\le j<k\le M_i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta}; y_{ij}, y_{ik})\right| > \frac{\epsilon}{2}\right\} \\ \to\ & 0 \quad \text{as } n\to\infty \qquad \text{(S2.14)} \end{aligned}$$

It then follows from (S2.13), (S2.14) and Assumption 7 that

$$\begin{aligned}
& P_{\xi d}\left[\left|\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})-\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}\right|>\epsilon\right] \\
\leq\ & P_{\xi d}\left[\left|\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})-\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};y_{ij},y_{ik})\right|>\frac{\epsilon}{2}\right] \\
& +P_{\xi d}\left[\left|\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};y_{ij},y_{ik})-\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}\right|>\frac{\epsilon}{2}\right] \\
\to\ & 0 \quad \text{as } n\to\infty. \qquad\qquad \text{(S2.15)}
\end{aligned}$$

Hence, combining (S2.12) and (S2.15) gives

$$\begin{aligned}
& P_{\xi d}\Big[\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik}) \\
& -\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}>2\epsilon\Big] \\
\leq\ & P_{\xi d}\Big[\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik}) \\
& -\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})>\epsilon\Big] \\
& +P_{\xi d}\Big[\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik}) \\
& -\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}>\epsilon\Big] \\
\to 0\ & \quad \text{as } n\to\infty. \qquad\qquad \text{(S2.16)}
\end{aligned}$$

Part 4: Let

$$H_n(\boldsymbol{\theta})=\frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})-\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}E_\xi\{\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}.$$

Then for the given open subset $\mathcal{O}(\boldsymbol{\theta}_r^*), r=1,\cdots,F$,

$$\begin{aligned}
& \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}H_n(\boldsymbol{\theta}) \\
\leq\ & \frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})+\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}E_\xi\{-\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\} \\
\leq\ & \frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})+\frac{1}{T}\sum_{i=1}^{N}E_\xi[\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\{-\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}] \\
=\ & \frac{1}{T}\sum_{i\in s}w_i\sum_{j<k,j,k\in s(i)}w_{jk|i}\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})-\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i}E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}.
\end{aligned}$$

Combining this with (S2.16) yields

$$\begin{aligned}
& P_{\xi d}\{\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta}) > 2\epsilon\} \\
\leq\ & P_{\xi d}\Bigg\{\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i} \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik}) \\
& -\frac{1}{T}\sum_{i=1}^{N}\sum_{1\leq j<k\leq M_i} E_\xi\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} \psi_{ijk}(\boldsymbol{\theta}; Y_{ij}, Y_{ik})\} > 2\epsilon\Bigg\} \\
\to\ & 0 \quad \text{as } n\to\infty. \qquad \text{(S2.17)}
\end{aligned}$$

By analogy, we can show that for $r = 1, \cdots, F$,

$$P_{\xi d}\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta}) < -2\epsilon\} \to 0 \quad \text{as } n\to\infty. \qquad \text{(S2.18)}$$

Part 5:

First, we show that

$$\begin{aligned}
& \left\{\mathbf{y} : \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |H_n(\boldsymbol{\theta}; \mathbf{y})| > 2\epsilon\right\} \\
\subset\ & \left\{\mathbf{y} : \sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta}; \mathbf{y}) > 2\epsilon\right\} \cup \left\{\mathbf{y} : \inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta}; \mathbf{y}) < -2\epsilon\right\}. \qquad \text{(S2.19)}
\end{aligned}$$

We prove this result by the contradiction method. Suppose $\mathbf{y}$ belongs to the left hand side of (S2.19), but not in the right hand side of (S2.19), then

$$\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |H_n(\boldsymbol{\theta}; \mathbf{y})| > 2\epsilon, \qquad \text{(S2.20)}$$

$$\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta}; \mathbf{y}) \leq 2\epsilon, \qquad \text{(S2.21)}$$

$$\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta}; \mathbf{y}) \geq -2\epsilon, \qquad \text{(S2.22)}$$

By (S2.21),

$$H_n(\boldsymbol{\theta}; \mathbf{y}) \leq 2\epsilon \text{ for any } \boldsymbol{\theta} \in \mathcal{O}(\boldsymbol{\theta}_r^*),$$

and by (S2.22),

$$H_n(\boldsymbol{\theta}; \mathbf{y}) \geq -2\epsilon \text{ for any } \boldsymbol{\theta} \in \mathcal{O}(\boldsymbol{\theta}_r^*).$$

Therefore,

$$|H_n(\boldsymbol{\theta}; \mathbf{y})| \leq 2\epsilon \text{ for any } \boldsymbol{\theta} \in \mathcal{O}(\boldsymbol{\theta}_r^*),$$

implying that

$$\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |H_n(\boldsymbol{\theta}; \mathbf{y})| \leq 2\epsilon,$$

and this contradicts to (S2.20). Thus, (S2.19) follows.

Now combining (S2.19), (S2.17) and (S2.18), we obtain that

$$\begin{aligned} & P_{\xi d}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |H_n(\boldsymbol{\theta};\mathbf{y})| > 2\epsilon\right\} \\ \leq \quad & P_{\xi d}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta};\mathbf{y}) > 2\epsilon\right\} + P_{\xi d}\left\{\inf_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} H_n(\boldsymbol{\theta};\mathbf{y}) < -2\epsilon\right\} \\ \rightarrow \quad & 0 \text{ as } n\rightarrow\infty \end{aligned} \tag{S2.23}$$

Part 6:

By (S2.5) and (S2.23), we obtain that

$$\begin{aligned} P_{\xi d}\{\sup_{\boldsymbol{\theta}\in\Theta}|H_n(\boldsymbol{\theta})| > 2\epsilon\} \quad & \leq \quad P_{\xi d}\left\{\sup_{\boldsymbol{\theta}\in\cup_{r=1}^F\mathcal{O}(\boldsymbol{\theta}_r^*)} |H_n(\boldsymbol{\theta})| > 2\epsilon\right\} \\ & \leq \quad \sum_{r=1}^F P_{\xi d}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_r^*)} |H_n(\boldsymbol{\theta})| > 2\epsilon\right\} \\ & \rightarrow \quad 0 \quad \text{as } n\rightarrow\infty. \end{aligned}$$

As a result, we obtain

$$\begin{aligned} & \sup_{\boldsymbol{\theta}\in\Theta}|H_n(\boldsymbol{\theta})| \\ = \quad & \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{i\in s} w_i \sum_{j<k,j,k\in s(i)} w_{jk|i}\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik}) - \frac{1}{T}\sum_{i=1}^N \sum_{1\leq j<k\leq M_i} E_\xi\{\psi_{ijk}(\boldsymbol{\theta};Y_{ij},Y_{ik})\}\right| \\ = \quad & \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right| \\ \stackrel{p}{\rightarrow} \quad & 0 \quad \text{as } n\rightarrow\infty, \end{aligned}$$

where "p" represents convergence in probability with respect to joint model ξ and sampling design d.

Proof of Theorem 1: Note that

$$\begin{aligned} \left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta})\right| \quad & = \quad \left|\Delta_T(\boldsymbol{\theta}) + \frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right| \\ & \geq \quad |\Delta_T(\boldsymbol{\theta})| - \left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right|. \end{aligned}$$

By Lemma 1, for any $\epsilon > 0$, we have

$$\begin{aligned}
\inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}\left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta})\right| &\geq \inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}\left\{|\Delta_T(\boldsymbol{\theta})| - \left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right|\right\} \\
&\geq \inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}|\Delta_T(\boldsymbol{\theta})| - \sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}\left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right| \\
&\geq \inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}|\Delta_T(\boldsymbol{\theta})| - \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta}) - \Delta_T(\boldsymbol{\theta})\right| \\
&\geq \inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}|\Delta_T(\boldsymbol{\theta})| + o_p(1).
\end{aligned}$$

It follows from Assumption 4 that, for any $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that

$$P_{\xi d}\left\{\inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\epsilon}\left|\frac{1}{T}\mathbf{U}_w(\boldsymbol{\theta})\right| > \delta_\epsilon\right\} \to 1$$

as $n \to \infty$, where the probability $P_{\xi d}$ is evaluated under the model ξ and sampling design d. Noting that $\mathbf{U}_w(\widehat{\boldsymbol{\theta}}_w) = \mathbf{0}$ by Assumption 5, the limit above implies that, for any $\epsilon > 0$, $P_{\xi d}(||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_w|| \leq \epsilon) \to 1$ as $n \to \infty$. This completes the proof that $\widehat{\boldsymbol{\theta}}_w \xrightarrow{p} \boldsymbol{\theta}$.

S3 Simulation Results

S3.1 Logistic Mixed Effect Model

Table 1: Simulation Results for Estimation of β_1 under Logistic Mixed Effect Model in Case B

δ_B	UML*				PML			
	BR**	RRMSE	100AVE	100MSE***	BR**	RRMSE	100AVE	100MSE***
1/3	7.09	6.06	3.28	3.31	196.38	16.67	5.12	25.03
1/2	8.68	5.81	3.02	3.04	79.54	8.03	3.61	5.81
1	7.03	5.81	2.93	3.04	7.03	5.81	2.93	3.04
δ_B	WPL							
	BR	RRMSE	100AVE	100MSE				
1/3	5.99	6.66	3.91	3.99				
1/2	6.54	6.08	3.36	3.33				
1	7.63	6.01	3.11	3.25				

*: UML denotes the unweighted maximum likelihood method; PML denotes the pseudo maximum likelihood approach; WPL denotes our weighted pairwise likelihood method.
**: BR and RRMSE represent bias ratio (%) and relative root mean square error, respectively.
***: 100AVE and 100MSE represent 100 times of average variance estimates and mean square error, respectively.

S3.2 Linear Mixed Effect Regression Model

Table 2: Simulation Results for Estimation of β_1 under Linear Mixed Effect Regression Model in Case B

	Invariant							
δ_C	UML*				PML			
	BR**	RRMSE	100AVE	100MSE***	BR	RRMSE	100AVE	100MSE
1	0.32	4.70	0.22	0.22	0.64	5.01	0.24	0.25
2	-4.90	4.80	0.23	0.23	-4.89	4.95	0.25	0.25
3	-4.10	4.98	0.24	0.25	-4.54	5.15	0.25	0.26
∞	2.66	4.98	0.24	0.25	2.47	5.10	0.26	0.26
δ_C	WPL							
	BR	RRMSE	100AVE	100MSE				
1	0.45	6.15	0.37	0.38				
2	-4.76	5.70	0.33	0.33				
3	-4.64	5.73	0.32	0.33				
∞	2.15	5.67	0.31	0.32				
	Non-invariant							
δ_C	UML				PML			
	BR	RRMSE	100AVE	100MSE	BR	RRMSE	100AVE	100MSE
1	-0.60	4.62	0.22	0.21	-0.98	4.96	0.25	0.25
2	-2.33	4.84	0.23	0.24	-2.08	4.98	0.25	0.25
3	-1.29	4.98	0.24	0.25	-1.44	5.09	0.25	0.26
∞	0.44	4.85	0.24	0.24	0.31	4.92	0.25	0.24
δ_C	WPL							
	BR	RRMSE	100AVE	100MSE				
1	-0.52	5.98	0.35	0.36				
2	-2.51	5.52	0.30	0.30				
3	-1.15	5.54	0.29	0.31				
∞	-0.74	5.32	0.29	0.28				

$*$: UML denotes the unweighted maximum likelihood method; PML denotes the pseudo maximum likelihood approach; WPL denotes our weighted pairwise likelihood method.
$**$: BR and RRMSE represent bias ratio (%) and relative root mean square error, respectively.
$***$: 100AVE and 100MSE represent 100 times of average variance estimates and mean square error, respectively.

Table 3: Simulation Results for Estimation of σ_e^2 under Linear Mixed Effect Regression Model in Case B

Invariant

δ_C	UML*				PML			
	BR**	RRMSE	100AVE	100MSE***	BR	RRMSE	100AVE	100MSE
1	-222.21	11.20	0.81	5.02	-242.70	12.24	0.83	6.00
2	-51.86	5.48	0.95	1.20	-88.05	6.58	0.97	1.74
3	-23.33	4.94	0.97	0.98	-63.11	5.76	0.99	1.33
∞	-3.13	4.97	0.99	0.99	-43.61	5.47	1.01	1.20

δ_C	WPL			
	BR	RRMSE	100AVE	100MSE
1	-2.05	6.79	1.78	1.84
2	-0.45	5.82	1.33	1.36
3	0.59	5.44	1.24	1.18
∞	-1.50	5.45	1.20	1.19

Non-invariant

δ_C	UML				PML			
	BR	RRMSE	100AVE	100MSE	BR	RRMSE	100AVE	100MSE
1	-184.91	9.63	0.84	3.71	-163.72	9.09	0.90	3.31
2	-45.03	5.33	0.95	1.14	-50.75	5.54	0.99	1.23
3	-22.22	5.04	0.97	1.02	-36.60	5.28	1.00	1.12
∞	-2.39	4.95	0.99	0.98	-26.19	5.13	1.00	1.05

δ_C	WPL			
	BR	RRMSE	100AVE	100MSE
1	-0.85	6.08	1.47	1.48
2	-3.86	5.36	1.15	1.15
3	-3.40	5.31	1.12	1.13
∞	-2.18	5.11	1.11	1.05

$*$: UML denotes the unweighted maximum likelihood method; PML denotes the pseudo maximum likelihood approach; WPL denotes our weighted pairwise likelihood method.

$**$: BR and RRMSE represent bias ratio (%) and relative root mean square error, respectively.

$***$: 100AVE and 100MSE represent 100 times of average variance estimates and mean square error, respectively.

Table 4: Simulation Results for Estimation of θ_2 under Linear Mixed Effect Regression Model in Case B

Invariant								
δ_C	UML*				PML			
	BR**	RRMSE	100AVE	100MSE***	BR	RRMSE	100AVE	100MSE
1	-4.83	12.44	5.97	5.96	32.52	12.91	5.86	6.42
2	-4.62	12.75	6.14	6.26	2.78	12.15	5.54	5.69
3	-5.67	12.62	6.16	6.14	-4.43	11.89	5.45	5.45
∞	-6.09	12.48	6.17	6.00	-9.72	11.79	5.64	5.36
δ_C	WPL							
	BR	RRMSE	100AVE	100MSE				
1	-5.23	13.42	7.00	6.94				
2	-5.64	13.46	6.78	6.99				
3	-7.43	13.36	6.79	6.88				
∞	-8.51	13.17	6.83	6.69				
Non-invariant								
δ_C	UML				PML			
	BR	RRMSE	100AVE	100MSE	BR	RRMSE	100AVE	100MSE
1	-145.43	19.46	4.71	14.60	-17.60	11.87	5.25	5.43
2	-32.42	12.90	5.80	6.42	-17.00	12.01	5.26	5.56
3	-15.84	12.72	6.01	6.24	-15.30	11.95	5.27	5.50
∞	-3.11	12.95	6.19	6.46	-13.35	12.08	5.29	5.63
δ_C	WPL							
	BR	RRMSE	100AVE	100MSE				
1	-5.27	13.92	7.28	7.47				
2	-3.82	13.13	6.50	6.64				
3	-3.40	13.15	6.46	6.66				
∞	-4.26	13.40	6.53	6.92				

$*$: UML denotes the unweighted maximum likelihood method; PML denotes the pseudo maximum likelihood approach; WPL denotes our weighted pairwise likelihood method.
$**$: BR and RRMSE represent bias ratio (%) and relative root mean square error, respectively.
$***$: 100AVE and 100MSE represent 100 times of average variance estimates and mean square error, respectively.