Supplementary Material for “Improved empirical likelihood inference and variable selection for generalized linear models with longitudinal nonignorable dropouts”

Lei Wang[1] **, Wei Ma**[1]

(C1) $\{(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{r}_i) : i = 1, \ldots, n\}$ are independent and identically distributed random vectors. The parameter vector $\boldsymbol{\beta}^0$ is an interior point of the parameter space, a compact subset of R^p.
(C2) $E\|\boldsymbol{y}_i - \boldsymbol{\mu}_i\|^3 < \infty$, the function $g^{-1}(\cdot)$ is bounded and has continuous second derivative, the marginal variance function $v(\cdot)$ is continuous and bounded, and have first derivative.
(C3) The probability function $\pi_{ij}(\boldsymbol{\Theta}_j)$ satisfies (a) it is twice differentiable with respect to $\boldsymbol{\Theta}_j$; (b) $0 < c_0 < \pi_{ij}(\boldsymbol{\Theta}_j) < 1$ for a positive constant c_0; (c) $\partial\pi_{ij}(\boldsymbol{\Theta}_j)/\partial\boldsymbol{\Theta}_j$ is uniformly bounded.
(C4) The random vectors $\boldsymbol{x}_{ij}$ are bounded in probability for all i and j, the matrices $\boldsymbol{\Sigma}_{\boldsymbol{g}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{h}}$ are nonsingular, $\boldsymbol{\Lambda}_{\boldsymbol{g}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{h}}$ are positive definite.
(C5) The $p_\nu(\cdot)$ satisfies $\max_{j\in\mathcal{A}} p'_\nu(|\beta_{j0}|) = o_p(n^{-1/3})$ and $\max_{j\in\mathcal{A}} p''_\nu(|\beta_{j0}|) = o(1)$.
(C6) As $n \to \infty$, $\nu \to 0$, $n^{1/3}\nu \to \infty$, and $\liminf_{n\to 0} \liminf_{\beta\to 0^+} p'_\nu(|\beta|)/\nu > 0$.

In the following, we mainly derive the results for $\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})$ and the similar arguments for $\hat{\boldsymbol{h}}_i(\boldsymbol{\beta})$ can be obtained.

Lemma 1 *Under assumptions (C1)-(C4), we have*

$$(1)\quad \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) \longrightarrow N(0, \boldsymbol{\Sigma}_{\boldsymbol{g}}), \quad (2)\quad \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \xrightarrow{p} \boldsymbol{\Lambda}_{\boldsymbol{g}};$$

$$(3)\quad \frac{1}{n}\sum_{i=1}^n \frac{\partial\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\Delta}_{\boldsymbol{g}}, \qquad (4)\quad \max_i \|\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\| = o_p(n^{1/2}).$$

Proof of Lemma 1. Note that

$$\begin{aligned} n^{-1/2}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) &= n^{-1/2}\sum_{i=1}^n \boldsymbol{g}_i(\boldsymbol{\Theta}_m^0, \boldsymbol{\beta}^0) - \Big[n^{-1}\sum_{i=1}^n \partial\boldsymbol{g}_i(\boldsymbol{\Theta}_m^0, \boldsymbol{\beta}^0)/\partial\boldsymbol{\Theta}_m\Big] n^{1/2}(\hat{\boldsymbol{\Theta}}_m - \boldsymbol{\Theta}_m^0) + o_p(1) \\ &= I_n + I_n^*. \end{aligned}$$

It can be seen that $I_n \to N(0, \boldsymbol{\Lambda}_{\boldsymbol{g}})$ and

$$I_n^* \to N(0, E[\partial\boldsymbol{g}_i(\boldsymbol{\Theta}_m^0, \boldsymbol{\beta}^0)/\partial\boldsymbol{\Theta}_m]\boldsymbol{\Sigma} E[\partial\boldsymbol{g}_i(\boldsymbol{\Theta}_m^0, \boldsymbol{\beta}^0)/\partial\boldsymbol{\Theta}_m]^T).$$

[1] School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin, 300071, China

It is not difficult to verify that $E(I_n + I_n^*) = o_p(1)$. Direct calculation yields $E(I_n I_n^*) = O_p(n^{-1/2})$ and $Cov(I_n, I_n^*) = o_p(1)$. Then it follows

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) \longrightarrow N(0, \boldsymbol{\Lambda_g}+E[\partial \boldsymbol{g}_i(\boldsymbol{\Theta}_m^0, \boldsymbol{\beta}^0)/\partial \boldsymbol{\Theta}_m]\boldsymbol{\Sigma} E[\partial \boldsymbol{g}_i(\boldsymbol{\Theta}_m^0, \boldsymbol{\beta}^0)/\partial \boldsymbol{\Theta}_m]^T).$$

Lemma 2 *Under assumptions (C1)-(C4), we have*

$$(1)\quad \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\boldsymbol{h}}_i(\boldsymbol{\beta}^0) \longrightarrow N(0, \boldsymbol{\Sigma_h}), \quad (2)\quad \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{h}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{h}}_i(\boldsymbol{\beta}^0)^T \xrightarrow{p} \boldsymbol{\Lambda_h};$$

$$(3)\quad \frac{1}{n}\sum_{i=1}^n \frac{\partial \hat{\boldsymbol{h}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\Delta_h}, \qquad (4)\quad \max_i \|\hat{\boldsymbol{h}}_i(\boldsymbol{\beta}^0)\| = o_p(n^{1/2}).$$

Proof of Lemma 2. Use the similar arguments as in the proof of Lemma 1.

Proof of Theorem 1. Noting that the Lagrange multiplier method leads to the empirical log-likelihood ratio function for $\boldsymbol{\beta}^0$

$$\hat{R}_Q(\boldsymbol{\beta}^0) = 2\sum_{i=1}^n \log\{1 + \boldsymbol{\lambda}^T(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\},$$

where vector $\rho = \boldsymbol{\lambda}(\boldsymbol{\beta}^0)$ is the solution to

$$\mathcal{D}(\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \frac{\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{1 + \boldsymbol{\lambda}^T\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)} = 0.$$

It can be seen that $\hat{\boldsymbol{\beta}}_Q$ is the solution satisfying the following two equations:

$$\boldsymbol{T}_{1n}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \frac{\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}^T\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})} = 0,$$
$$\boldsymbol{T}_{2n}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \frac{\{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}\}^T\boldsymbol{\lambda}}{1 + \boldsymbol{\lambda}^T\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})} = 0.$$

Note that $\boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0) = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)$ and $\boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0) = 0$. Using the similar arguments in the proof of Lemma 1 in Qin and Lawless (1994), we have $\hat{\boldsymbol{\beta}}_Q \to \boldsymbol{\beta}^0$. Applying Taylor expansions to $\boldsymbol{T}_{1n}(\hat{\boldsymbol{\beta}}_Q, \hat{\boldsymbol{\lambda}})$ and $\boldsymbol{T}_{2n}(\hat{\boldsymbol{\beta}}_Q, \hat{\boldsymbol{\lambda}})$ at $(\boldsymbol{\beta}^0, 0)$, we get

$$0 = \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0) + \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) + \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}^T}\hat{\boldsymbol{\lambda}} + o_p(u_n),$$
$$0 = \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0) + \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) + \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}^T}\hat{\boldsymbol{\lambda}} + o_p(u_n),$$

where $u_n = ||\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0|| + ||\hat{\boldsymbol{\lambda}}||$. Also, the above two equations can be rewritten as

$$\begin{pmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0 \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}^T} & \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} \\ \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}^T} & 0 \end{pmatrix}^{-1}_{(\boldsymbol{\beta}^0,0)} \begin{pmatrix} -\frac{1}{n}\sum\limits_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) + o_p(u_n) \\ o_p(u_n) \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{n}\sum\limits_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T & \frac{1}{n}\sum\limits_{i=1}^{n} \frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \\ \frac{1}{n}\sum\limits_{i=1}^{n} \{\frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\}^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{n}\sum\limits_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) + o_p(u_n) \\ o_p(u_n) \end{pmatrix}.$$

From this and $\boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0) = n^{-1}\sum\limits_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) = O_p(n^{-1/2})$, we know that $u_n = O_p(n^{-1/2})$. Then we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) = \{\boldsymbol{\Delta}_{\boldsymbol{g}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{g}}\}^{-1} \boldsymbol{\Delta}_{\boldsymbol{g}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) + o_p(1).$$

Together with Lemma 1, we can derive that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) \longrightarrow N(0, \{\boldsymbol{\Delta}_{\boldsymbol{g}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{g}}\}^{-1} \boldsymbol{\Delta}_{\boldsymbol{g}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{g}} \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{g}} \{\boldsymbol{\Delta}_{\boldsymbol{g}}^T \boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{g}}\}^{-1}).$$

Proof of Theorem 2. Based on the proof of Lemma 1 and applying the same idea in the proof of (2.14) in Owen (1990), we first show that

$$||\boldsymbol{\lambda}|| = O_p(n^{-1/2}).$$

Write $\boldsymbol{\lambda} \equiv \boldsymbol{\lambda}(\boldsymbol{\beta}^0) = \rho \boldsymbol{u}$, where $\rho = ||\boldsymbol{\lambda}||$, $\boldsymbol{u} = \boldsymbol{\lambda}/||\boldsymbol{\lambda}||$ and $||\boldsymbol{u}|| = 1$. We have

$$\begin{aligned} 0 =& \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{1 + \boldsymbol{\lambda}^T(\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0))} \\ =& \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{1 + \rho \boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)} \\ =& \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) - \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{u}\rho}{1 + \rho \boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}. \end{aligned}$$

By multiplicating $\boldsymbol{u}^T$, we obtain

$$\begin{aligned} \left\| \boldsymbol{u}^T \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) \right\| =& \frac{1}{n}\sum_{i=1}^{n} \frac{\boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{u}\rho}{1 + \rho \boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)} \\ \geq& \frac{1}{1 + \rho \max\limits_{1 \leq i \leq n} |\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)|} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{u}\rho, \end{aligned}$$

where the inequality follows from positivity of $1 + \rho \boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)$. Then

$$\rho \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{u}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{u} \leq \left\| \boldsymbol{u}^T \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) \right\| [1 + \rho \max_{1 \leq i \leq n} |\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)|].$$

Note that $\max_{1\le i\le n} ||\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)|| = o(n^{1/2})$. By WLLN and Lemma 1, we have

$$\rho[\boldsymbol{u}^T \boldsymbol{\Lambda_g} \boldsymbol{u} + o_p(1)] \le \Big\| \boldsymbol{u}^T \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)) \Big\| + \rho o(n^{1/2}),$$

which leads to $\rho = O_p(n^{-1/2})$, i.e., $||\boldsymbol{\lambda}|| = O_p(n^{-1/2})$. Naturally,

$$\max_{1\le i\le n} ||\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)|| \le ||\boldsymbol{\lambda}|| \max_{1\le i\le n} ||\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)|| = O_p(n^{-1/2})o(n^{1/2}) = o_p(1).$$

Expanding $\mathcal{D}(\boldsymbol{\lambda})$, we get

$$\begin{aligned} 0 = \mathcal{D}(\boldsymbol{\lambda}) =& \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\Big\{1 - \boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) + \frac{[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^2}{(1+\xi_i)^3}\Big\} \\ =& \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) - \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\boldsymbol{\lambda} + \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\frac{[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^2}{(1+\xi_i)^3}, \end{aligned}$$

where $\xi_i \in (0, \boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0))$. Using the fact that $\max_{1\le i\le n} ||\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)|| = o_p(1)$, then $|\xi_i| = o_p(1)$. Noting that

$$\begin{aligned} \Big\|\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\frac{[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^2}{(1+\xi_i)^3}\Big\| \le& \frac{\max_{1\le i\le n} ||\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)||}{1 - \max_{1\le i\le n}|\xi_i|}\Big\|\boldsymbol{\lambda}^T \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\boldsymbol{\lambda}\Big\| \\ =& o(n^{1/2})O_p(n^{-1}) \\ =& o_p(n^{-1/2}), \end{aligned}$$

thus

$$\boldsymbol{\lambda} = \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\Big] + \boldsymbol{\zeta}, \tag{1}$$

where $||\boldsymbol{\zeta}|| = o_p(n^{-1/2})$. A Taylor expansion of $\hat{R}_Q(\boldsymbol{\beta}^0)$ yields

$$\begin{aligned} \hat{R}_Q(\boldsymbol{\beta}^0) =& 2\sum_{i=1}^n \Big\{\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) - \frac{1}{2}[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^2 + \frac{1}{3}\frac{[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^3}{(1+\xi_i)^3}\Big\} \\ =& 2\boldsymbol{\lambda}^T\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) - \sum_{i=1}^n \boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\boldsymbol{\lambda} + \frac{2}{3}\sum_{i=1}^n \frac{[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^3}{(1+\xi_i)^3}. \end{aligned}$$

Similarly,

$$\begin{aligned} \Big\|\sum_{i=1}^n \frac{[\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)]^3}{(1+\xi_i)^3}\Big\| \le& \frac{\max_{1\le i\le n} ||\boldsymbol{\lambda}^T\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)||}{1 - \max_{1\le i\le n}|\xi_i|}\Big\|\boldsymbol{\lambda}^T \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\boldsymbol{\lambda}\Big\| \\ =& o_p(1)nO_p(n^{-1}) \\ =& o_p(1), \end{aligned}$$

therefore,

$$\hat{R}_Q(\boldsymbol{\beta}^0) = 2\boldsymbol{\lambda}^T \sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) - \sum_{i=1}^n \boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{\lambda} + o_p(1). \qquad (2)$$

Substituting (1) in (2), it holds that

$$\begin{aligned}\hat{R}_Q(\boldsymbol{\beta}^0) =& n\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\Big] \\ & - n\boldsymbol{\zeta}^T \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{\zeta} + o_p(1).\end{aligned}$$

A simple calculation shows that

$$n\boldsymbol{\zeta}^T \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T \boldsymbol{\zeta} = n o_p(n^{-1/2}) o_p(n^{-1/2}) = o_p(1).$$

Therefore,

$$\hat{R}_Q(\boldsymbol{\beta}^0) = \Big[\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]^{-1}\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\Big] + o_p(1).$$

Together with the proof of Lemma 1, we can conclude that

$$\hat{R}_Q(\boldsymbol{\beta}^0) \longrightarrow \rho_1 w_1 + \rho_2 w_2 + \ldots + \rho_{p\times q} w_{p\times q},$$

where $w_l,\ l = 1, \ldots, p \times q$, are independent and follow the standard χ^2 distribution with one degree, the weights ρ_l are eigenvalues of $\boldsymbol{\Lambda}_{\boldsymbol{g}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{g}}$. For the asymptotic theories of $\hat{\boldsymbol{\beta}}_H$, it can be proved in a similar way.

Lemma 3 *Assume conditions (C1)-(C6) hold and denote* $\mathcal{D}_n = \{\boldsymbol{\beta} : ||\boldsymbol{\beta} - \boldsymbol{\beta}^0|| \leq d_n\}$, $d_n = n^{-1/3}$, *then as* $n \to \infty$, *with probability to 1,* $\hat{R}_p(\boldsymbol{\beta})$ *has a minimum in* $\mathcal{D}_n$.

Proof of Lemma 3. Let $\boldsymbol{\beta} = \boldsymbol{\beta}^0 + \boldsymbol{u} d_n$, where $||\boldsymbol{u}|| = 1$. For simplicity, we denote $\boldsymbol{\lambda} \equiv \lambda(\boldsymbol{\beta})$. First, we give a lower bound for $\hat{R}_p(\boldsymbol{\beta})$ on the surface of the ball. Similar to the proof of Owen (1990), when $||\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})||^3 < \infty$, we have

$$\begin{aligned}\boldsymbol{\lambda} =& \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})^T\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\Big] + o(n^{-1/3}) \\ =& O(n^{-1/3})\ (a.s.),\end{aligned}$$

uniformly about $\boldsymbol{\beta} \in \mathcal{D}_n$. By this and the Taylor expansion, similar to the proof of Lemma 1 in Qin and Lawless (1994), we obtain

$$\begin{aligned}
\hat{R}_p(\boldsymbol{\beta}) =& 2\sum_{i=1}^n \boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}) - \sum_{i=1}^n \boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})^T \boldsymbol{\lambda} + n\sum_{j=1}^p p_\nu(|\beta_j|) + o(n^{1/3}) \\
=& n\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\Big]^T \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})^T\Big]^{-1} \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\Big] + n\sum_{j=1}^p p_\nu(|\beta_j|) + o(n^{1/3}) \\
=& n\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) + \frac{1}{n}\sum_{i=1}^n \frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \boldsymbol{u} n^{-1/3}\Big]^T \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]^{-1} \\
\times& \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) + \frac{1}{n}\sum_{i=1}^n \frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \boldsymbol{u} n^{-1/3}\Big] + n\sum_{j=1}^p p_\nu(|\beta_j|) + o(n^{1/3}) \\
=& n\Big\{O(n^{-1/2}(\log\log n)^{1/2}) + E\Big[\frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\Big]\boldsymbol{u} n^{-1/3}\Big\}^T \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]^{-1} \\
\times& \Big\{O(n^{-1/2}(\log\log n)^{1/2}) + E\Big[\frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\Big]\boldsymbol{u} n^{-1/3}\Big\} \\
&+ n\sum_{j=1}^p p_\nu(|\beta_{j0}|) + n\sum_{j=1}^d p'_\nu(|\beta_{j0}|)\mathrm{sign}(\beta_{j0}) u_j d_n + n\sum_{j=1}^d p''_\nu(|\beta_{j0}|) u_j^2 d_n^2 + o(n^{1/3}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\Big\|\sum_{j=1}^d n p'_\nu(|\beta_{j0}|)\mathrm{sign}(\beta_{j0}) u_j d_n\Big\| \le n\sqrt{d} a_n d_n = O(n a_n d_n) = o(n^{1/3}), \\
&n\sum_{j=1}^d p''_\nu(|\beta_{j0}|) u_j^2 d_n^2 \le n p b_n d_n^2 = o(n^{1/3}).
\end{aligned}$$

This yields $\hat{R}_p(\boldsymbol{\beta}) \ge (c-\epsilon)n^{1/3} + n\sum_{j=1}^d p_\nu(|\beta_{j0}|),\ a.s.$, where $c-\epsilon > 0$ and c is the smallest eigenvalue of $E[\partial\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}]^T[E\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T]^{-1}E[\partial\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}]$. Similarly,

$$\begin{aligned}
\hat{R}_p(\boldsymbol{\beta}^0) =& n\Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\Big]^T \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)^T\Big]^{-1} \Big[\frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)\Big] \\
&+ n\sum_{j=1}^d p_\nu(|\beta_{j0}|) + o(1) \\
=& O(\log\log n) + n\sum_{j=1}^d p_\nu(|\beta_{j0}|),\ a.s..
\end{aligned}$$

Since $\hat{R}_p(\boldsymbol{\beta})$ is a continuous function about $\boldsymbol{\beta}$ as $\boldsymbol{\beta}$ belongs to the ball $||\boldsymbol{\beta} - \boldsymbol{\beta}^0|| \le n^{-1/3}$, $\hat{R}_p(\boldsymbol{\beta})$ has a minimum value in the interior of this ball, and $\hat{\boldsymbol{\beta}}_Q$

satisfies

$$\frac{\partial \hat{R}_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_Q} = 2\sum_{i=1}^{n} \frac{[\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}]^T \boldsymbol{\lambda}}{1+\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})} + nb(\boldsymbol{\beta}) = 0,$$

with $b(\boldsymbol{\beta}) = \{p'_\nu(|\beta_1|)\text{sign}(\beta_1),\ p'_\nu(|\beta_2|)\text{sign}(\beta_2), \ldots, p'_\nu(|\beta_p|)\text{sign}(\beta_p)\}^T$. This completes the proof.

Proof of Theorem 3. As Lemma 3 implies that there is a local minimizer $\hat{\boldsymbol{\beta}}_Q$ of $\hat{R}_p(\boldsymbol{\beta})$ uniformly for $\boldsymbol{\beta} \in \mathcal{D}_n$, then by Taylor expansion, we have

$$\begin{aligned}\frac{1}{n}\frac{\partial \hat{R}_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} =& \frac{2}{n}\sum_{i=1}^{n} \frac{[\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}]^T \boldsymbol{\lambda}}{1+\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})} + b(\boldsymbol{\beta}) \\ =& 2\Big[\boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0) + \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta} - \boldsymbol{\beta}^0) + \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}}(\boldsymbol{\lambda} - 0)\Big] + b(\boldsymbol{\beta}) + o(n^{-1/3}),\end{aligned}$$

where $b(\boldsymbol{\beta})$ is defined the same as in the proof of Lemma 3, and

$$\begin{aligned}\boldsymbol{T}_{1n}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})}{1+\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})}, \\ \boldsymbol{T}_{2n}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n}\sum_{i=1}^{n} \frac{[\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}]^T \boldsymbol{\lambda}}{1+\boldsymbol{\lambda}^T \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})}.\end{aligned}$$

In addition, the last term holds due to condition (C6). Note that $\boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0) = 0$ and $\partial T_{2n}(\boldsymbol{\beta}^0, 0)/\partial \boldsymbol{\beta} = 0$. By standard arguments, $\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0)/\partial \boldsymbol{\lambda} = E[\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0)/\partial \boldsymbol{\beta}]^T + o_p(1)$. Thus for $j = 1, \ldots, p$, we have

$$\frac{\partial \hat{R}_p(\boldsymbol{\beta})}{\partial \beta_j} = n\nu\{p'_\nu(|\beta_j|)\text{sign}(\beta_j)/\nu + O_P(n^{-1/3}/\nu)\}.$$

Under condition (C5) on $p'_\nu(\cdot)$, it can be seen that $p'_\nu(|\beta_j|)\text{sign}(\beta_j)$ dominates the sign of $\partial \hat{R}_p(\boldsymbol{\beta})/\partial \beta_j$ asymptotically for all $j \notin \mathcal{A}$. Therefore, it is sufficient to show that with probability tending to 1, as $n \to \infty$ for any $\boldsymbol{\beta}_1$ satisfying $\beta_1 - \beta_{10} = O_P(n^{-1/3})$ and for some small $\varepsilon_n = Cn^{-1/3}$ and $j \notin \mathcal{A}$,

$$\frac{\partial \hat{R}_p(\boldsymbol{\beta})}{\partial \beta_j} > 0,\ \beta_j \in (0, \varepsilon_n) \text{ and } \frac{\partial \hat{R}_p(\boldsymbol{\beta})}{\partial \beta_j} < 0,\ \beta_j \in (-\varepsilon_n, 0).$$

This completes the proof of part (i).

Denote $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \text{diag}\{p''_\nu(|\beta_1|), \ldots, p''_\nu(|\beta_p|)\}$, $\boldsymbol{\Sigma}^d = \text{diag}\{p''_\nu(|\beta_{10}|), \ldots, p''_\nu(|\beta_{d0}|)\}$, $b^d(\boldsymbol{\beta}) = \{p'_\nu(|\beta_{10}|)\text{sign}(\beta_{10}), \ldots, p'_\nu(|\beta_{d0}|)\text{sign}(\beta_{d0})\}^T$. Next, we establish part (ii). Taking derivation of $\boldsymbol{T}_{1n}(\boldsymbol{\beta}, \lambda)$ and $\boldsymbol{T}_{2n}(\boldsymbol{\beta}, \lambda)$ about $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ at $(\boldsymbol{\beta}, 0)$ respectively, we obtain

$$\begin{aligned}\Big[\frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\beta}}\Big]^T =& \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\lambda}^T} = \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\partial \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big]^T, \\ \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\lambda}^T} =& -\frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{g}}_i(\boldsymbol{\beta})\hat{\boldsymbol{g}}_i(\boldsymbol{\beta})^T, \quad \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\beta}} = 0.\end{aligned}$$

Expanding $\boldsymbol{T}_{1n}(\hat{\boldsymbol{\beta}}_Q, \hat{\boldsymbol{\lambda}})$ and $2\boldsymbol{T}_{2n}(\hat{\boldsymbol{\beta}}_Q, \hat{\boldsymbol{\lambda}}) + b(\hat{\boldsymbol{\beta}}_Q)$ about at $(\boldsymbol{\beta}^0, 0)$ yields

$$\begin{aligned} 0 =& \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0) + \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) + \frac{\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}^T}\hat{\boldsymbol{\lambda}} + o_p(r_n), \\ 0 =& \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0) + \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) + \frac{\partial \boldsymbol{T}_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}^T}\hat{\boldsymbol{\lambda}} + \frac{1}{2}b(\boldsymbol{\beta}^0) \\ &+ \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0) + o_p(r_n), \end{aligned}$$

where $r_n = ||\hat{\boldsymbol{\beta}}_Q - \boldsymbol{\beta}^0|| + ||\hat{\boldsymbol{\lambda}}||$. As $\hat{\boldsymbol{\beta}}_{Q_2} = 0$ with probability tending to 1, we consider the components $\hat{\boldsymbol{\beta}}_{Q_1}$ and $\hat{\boldsymbol{\lambda}}_1$ and immediately derive that

$$\begin{pmatrix} \hat{\boldsymbol{\lambda}}_1 \\ \hat{\boldsymbol{\beta}}_{Q_1} - \boldsymbol{\beta}_1^0 \end{pmatrix} = \boldsymbol{S}^{-1} \begin{pmatrix} -\boldsymbol{T}_{1n}^{(1)}(\boldsymbol{\beta}_1^0, 0) + o_p(r_n) \\ -\frac{1}{2}b^d(\boldsymbol{\beta}) + o_p(r_n) \end{pmatrix}$$

with

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{\Lambda}_{\boldsymbol{g}}^{(11)} & \boldsymbol{\Delta}_{\boldsymbol{g}}^{(12)} \\ \boldsymbol{\Delta}_{\boldsymbol{g}}^{(21)} & \boldsymbol{\Sigma}^d \end{pmatrix}_{(\boldsymbol{\beta}_1^0, 0)} \to S = \begin{pmatrix} \boldsymbol{\Lambda}_{\boldsymbol{g}}^{(11)} & \boldsymbol{T}_{(12)} \\ \boldsymbol{T}_{21} & \boldsymbol{\Sigma}^d \end{pmatrix},$$

where $\boldsymbol{T}_{1n}^{(1)}(\boldsymbol{\beta}_1^0, 0)$ is the $d \times q$ sub-vector of $\boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)$. More precisely, $\boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)$ has q different parts, and we take the first d sub-vector of each part to combine $\boldsymbol{T}_{1n}^{(1)}(\boldsymbol{\beta}_1^0, 0)$. It can be verified that the corresponding $dq \times dq$ submatrix of $\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)/\partial \boldsymbol{\lambda}^T$ converges to $\boldsymbol{\Lambda}_{\boldsymbol{g}}^{(11)}$, which is the corresponding $dq \times dq$ submatrix of $\boldsymbol{\Lambda}_{\boldsymbol{g}}$; the corresponding $dq \times d$ submatrix of $\partial \boldsymbol{T}_{1n}(\boldsymbol{\beta}^0, 0)/\partial \boldsymbol{\beta}$ converges to $\boldsymbol{T}_{(12)} = \boldsymbol{T}_{21}^T$, which is the corresponding $dq \times d$ submatrix of $\boldsymbol{\Delta}_{\boldsymbol{g}}$. Note

$$\boldsymbol{T}_{1n}^{(1)}(\boldsymbol{\beta}_1^0, 0) = (1_{1\times d}, 0_{1\times(p-d)}, \cdots, 1_{1\times d}, 0_{1\times(p-d)})_{p\times q} n^{-1} \sum_{i=1}^n \hat{\boldsymbol{g}}_i(\boldsymbol{\beta}^0) = O_p(n^{-1/2}),$$

which leads to $r_n = O_p(n^{-1/2})$. Therefore,

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}_{Q_1} - \boldsymbol{\beta}_1^0) =& (\boldsymbol{\Sigma}^d + \boldsymbol{V}')^{-1}\boldsymbol{\Delta}_{\boldsymbol{g}}^{21}(-\{\boldsymbol{\Lambda}_{\boldsymbol{g}}^{(11)}\}^{-1})\sqrt{n}\boldsymbol{T}_{1n}^{(1)}(\boldsymbol{\beta}_1^0, 0) \\ &- \frac{1}{2}(\boldsymbol{\Sigma}^d + \boldsymbol{V}')^{-1}\sqrt{n}b^d(\boldsymbol{\beta}) + o_p(1), \end{aligned}$$

where $\boldsymbol{V}' = \boldsymbol{\Delta}_{\boldsymbol{g}}^{21}\{\boldsymbol{\Lambda}_{\boldsymbol{g}}^{(11)}\}^{-1}\boldsymbol{\Delta}_{\boldsymbol{g}}^{(12)}$. Furthermore, under condition (C6), for $n \to \infty$, $\nu \to 0$, $a\nu \to 0$ (a is the constant in the SCAD penalty), we have $P(\min_{j=1,2,\ldots,d}|\beta_{j0}| > a\nu) \to 1$, which implies $P(|b^d(\boldsymbol{\beta})| = 0) \to 1$. Under condition (C5), we have $b_n = o(1)$ and $P(\boldsymbol{\Sigma}^d = 0) \to 1$, and $\sqrt{n}\boldsymbol{T}_{1n}^{(1)}(\boldsymbol{\beta}_1^0, 0) \to N(0, \boldsymbol{\Sigma}_{\boldsymbol{g}}^{(11)})$, where $\boldsymbol{\Sigma}_{\boldsymbol{g}}^{(11)}$ is the corresponding $dq \times dq$ submatrix of $\boldsymbol{\Sigma}_{\boldsymbol{g}}$. By Slutsky's theorem and the Central Limit theorem, we have the result.

References

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

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300–325.