Supplemental Materials for "Variable Selection and Model Averaging for Longitudinal Data Incorporating GEE Approach"

Hui Yang, Peng Lin, Guohua Zou and Hua Liang

Amgen Inc., Shandong University of Technology, Capital Normal University, and George Washington University

Abstract: In this document, we present the assumptions and the proofs for Theorems 1-4.

Let $f(\mathbf{y};\boldsymbol{\theta},\boldsymbol{\gamma})$ be the density function. Denote the corresponding score functions, evaluated at $(\boldsymbol{\theta}_0,\mathbf{0})$, by

$$\mathbf{T}=\begin{bmatrix}\mathbf{T}_1\\ \mathbf{T}_2\end{bmatrix}=\begin{bmatrix}\partial\log f(\mathbf{y};\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}\\ \partial\log f(\mathbf{y};\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}\end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}}.$$

Denote the quasi-score functions, evaluated at $(\boldsymbol{\theta}_0,\mathbf{0})$, by

$$\mathbf{U}=\begin{bmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{bmatrix}=\begin{bmatrix}\partial Q(\boldsymbol{\theta},\boldsymbol{\gamma};\mathbf{y})/\partial\boldsymbol{\theta}\\ \partial Q(\boldsymbol{\theta},\boldsymbol{\gamma};\mathbf{y})/\partial\boldsymbol{\gamma}\end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}}.$$

The corresponding second derivatives are denoted as

$$\mathbf{H}=\begin{bmatrix}\partial^2 Q/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top & \partial^2 Q/\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top\\ \partial^2 Q/\partial\boldsymbol{\gamma}\partial\boldsymbol{\theta}^\top & \partial^2 Q/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top\end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}}.$$

To study the large sample properties of the proposed model selection criterion ΔAIC, we need some regularity conditions.

A.1 Assumptions and two preliminary lemmas

(C.1): The log density function $\log f(\mathbf{y};\boldsymbol{\theta},\boldsymbol{\gamma})$ has continuous partial derivatives with respect to $(\boldsymbol{\theta},\boldsymbol{\gamma})$ in a neighborhood around $(\boldsymbol{\theta}_0,\mathbf{0})$, which are dominated by functions with finite means under $f_{\mathcal{N}}(\mathbf{y})=f(\mathbf{y};\boldsymbol{\theta}_0,\mathbf{0})$. The true density $f_0(\mathbf{y})=f(\mathbf{y};\boldsymbol{\theta}_0,\boldsymbol{\delta}n^{-1/2})$ can be represented by $f_{\mathcal{N}}(\mathbf{y})$ as

$$f_0(\mathbf{y})=f_{\mathcal{N}}(\mathbf{y})\big\{1+\mathbf{T}_2^\top(\mathbf{y})\boldsymbol{\delta}n^{-1/2}+r(\mathbf{y},\boldsymbol{\delta}n^{-1/2})\big\},$$

where $r(\mathbf{y},\mathbf{t})$ is small enough to make $f_{\mathcal{N}}(\mathbf{y})r(\mathbf{y},\mathbf{t})$ is of order $o(\|\mathbf{t}\|)$ uniformly in $\mathbf{y}$.

(C.2): The log quasi-likelihood function $Q(\boldsymbol{\theta}, \boldsymbol{\gamma}; \mathbf{y})$ has third continuous derivatives with respect to $(\boldsymbol{\theta}, \boldsymbol{\gamma})$ in a neighborhood around $(\boldsymbol{\theta}_0, \mathbf{0})$, which are dominated by functions with finite means under $f_{\mathcal{N}}(\mathbf{y})$. The quasi-information matrix $\boldsymbol{\Sigma}$ (defined below) exists and is non-singular under $f_{\mathcal{N}}(\mathbf{y})$.

$$\boldsymbol{\Sigma} = \mathrm{E}_{\mathcal{N}}(-\mathbf{H}) = \mathrm{var}_{\mathcal{N}}(\mathbf{U}) = \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{00} & \boldsymbol{\Sigma}^{01} \\ \boldsymbol{\Sigma}^{10} & \boldsymbol{\Sigma}^{11} \end{bmatrix}.$$

(C.3): The integrals $\int \mathbf{U}(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y}) r(\mathbf{y}, \mathbf{t}) \mathrm{d}\mathbf{y}$ and $\int \|\mathbf{U}(\mathbf{y})\|^2 f_{\mathcal{N}}(\mathbf{y}) r(\mathbf{y}, \mathbf{t}) \mathrm{d}\mathbf{y}$ are of order $o\left(\|\mathbf{t}\|^2\right)$.

(C.4): For some $\xi > 0$, the integrals $\int \|\mathbf{U}(\mathbf{y})\|^{2+\xi} f_{\mathcal{N}}(\mathbf{y}) \mathrm{d}\mathbf{y}$ and $\int \|\mathbf{U}(\mathbf{y})\|^{2+\xi} f_{\mathcal{N}}(\mathbf{y}) r(\mathbf{y}, \mathbf{t}) \mathrm{d}\mathbf{y}$ are of order $O(\mathbf{1})$. Also the variables $|U_{1k}^{2+\xi}(\mathbf{y}) T_{2r}(\mathbf{y})|$ and $|U_{2l}^{2+\xi}(\mathbf{y}) T_{2r}(\mathbf{y})|$ have finite mean under the null density $f_{\mathcal{N}}(\mathbf{y})$, for $k \in \{1, \cdots, p\}$ and $r, l \in \{1, \cdots, q\}$ with $U_{1k} = \partial Q / \partial \theta_k, U_{2l} = \partial Q / \partial \gamma_l$ and $T_{2r} = \partial \mathrm{log} f / \partial \gamma_r$.

The similar assumptions have customarily been assumed in the literature on quasi-likelihood function, GEE and local misspecification framework. See, for example, Wedderburn (1974), McCullagh (1983), Liang and Zeger (1986) and Hjort and Claeskens (2003) .

Lemma A.1. *Under the misspecification framework and the regularity conditions given in the Assumptions, we have*

$$\begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} \xrightarrow{d} N_{p+q}\left(\begin{bmatrix} \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{11} \end{bmatrix} \boldsymbol{\delta}, \boldsymbol{\Sigma} \right),$$

where

$$\mathbf{R}_{1,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{U}_1(\mathbf{y}_i) \quad \text{and} \quad \mathbf{R}_{2,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{U}_2(\mathbf{y}_i).$$

In particular, for the submodel S:

$$\begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,s,n} \end{bmatrix} \xrightarrow{d} N_{p+q_s}\left(\begin{bmatrix} \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\pi}_s \boldsymbol{\Sigma}_{11} \end{bmatrix} \boldsymbol{\delta}, \boldsymbol{\Sigma}_s \right).$$

Here "$\xrightarrow{d}$" denotes convergence in distribution under the sequence of $f_0(\mathbf{y})$.

Proof. We shall finish the proof by three steps. In the first two steps, we calculate the expectation and variance of the quasi-score under $f_0(\mathbf{y})$, respectively. In the third step, we verify the requirement for the Lyapounov central limit theorem, and complete the proof.

Step 1. Consider $\mathrm{E}_0(\mathbf{U}_1)$ first. $\mathrm{E}_0(\mathbf{U}_2)$ can be manipulated by the similar arguments. A direct calculation yields that

$$\mathrm{E}_0(\mathbf{U}_1) = \int \mathbf{U}_1(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y})\mathrm{d}\mathbf{y} + \int \mathbf{U}_1(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y}) \mathbf{T}_2^\top(\mathbf{y}) \boldsymbol{\delta} n^{-1/2}\mathrm{d}\mathbf{y} + \int \mathbf{U}_1(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y}) r(\mathbf{y}, \boldsymbol{\delta} n^{-1/2})\mathrm{d}\mathbf{y}. \tag{A.1}$$

It is easy to see that the first term in (A.1) equals to zero by the fact $\mathbf{U}(\mathbf{y}) = \mathbf{D}^\top \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ with $\boldsymbol{\mu} = \mathrm{E}_{\mathcal{N}}(\mathbf{y})$. Note that

$$\begin{aligned}
\int \mathbf{U}(\mathbf{y}) f_N(\mathbf{y}) \mathbf{T}^\top(\mathbf{y})\mathrm{d}\mathbf{y} &= \int \mathbf{U}(\mathbf{y}) f_N(\mathbf{y}) \left[\partial \log f_N(\mathbf{y})/\partial \boldsymbol{\beta}\right]^\top \mathrm{d}\mathbf{y} \\
&= \int \mathbf{D}^\top \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \left[\partial f_N(\mathbf{y})/\partial \boldsymbol{\beta}\right]^\top \mathrm{d}\mathbf{y} \\
&= \mathbf{D}^\top \mathbf{V}^{-1} \int \mathbf{y} \left[\partial f_N(\mathbf{y})/\partial \boldsymbol{\beta}^\top\right] \mathrm{d}\mathbf{y} - \mathbf{D}^\top \mathbf{V}^{-1} \boldsymbol{\mu} \int \left[\partial f_N(\mathbf{y})/\partial \boldsymbol{\beta}^\top\right] \mathrm{d}\mathbf{y} \\
&= \mathbf{D}^\top \mathbf{V}^{-1} \frac{\partial}{\partial \boldsymbol{\beta}^\top} \int \mathbf{y} f_N(\mathbf{y})\mathrm{d}\mathbf{y} - \mathbf{D}^\top \mathbf{V}^{-1} \boldsymbol{\mu} \frac{\partial}{\partial \boldsymbol{\beta}^\top} \int f_N(\mathbf{y})\mathrm{d}\mathbf{y} \\
&= \mathbf{D}^\top \mathbf{V}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}^\top} - \mathbf{0} = \mathbf{D}^\top \mathbf{V}^{-1}\mathbf{D} = \boldsymbol{\Sigma},
\end{aligned}$$

where the interchanges are justified by Assumption (C.1) that $|\mathbf{T}(\mathbf{y})|$ is dominated by function with finite mean under $f_{\mathcal{N}}(\mathbf{y})$ and (C.4) that $|\mathbf{U}(\mathbf{y})\mathbf{T}(\mathbf{y})|$ has finite mean under $f_{\mathcal{N}}(\mathbf{y})$. So the second term in (A.1) is $\boldsymbol{\Sigma}_{01}\boldsymbol{\delta} n^{-1/2}$. Also by Assumption (C.3), we conclude that the third term in (A.1) is of order $o(\mathbf{1}/\sqrt{n})$.

By the similar arguments, $\mathrm{E}_0(\mathbf{U}_2) = \boldsymbol{\Sigma}_{11}\boldsymbol{\delta}/\sqrt{n} + o(\mathbf{1}/\sqrt{n})$. As a result, the expectation of the quasi-score under $f_0(\mathbf{y})$ becomes

$$\mathrm{E}_0\begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{11} \end{bmatrix} \frac{\boldsymbol{\delta}}{\sqrt{n}} + o(\mathbf{1}/\sqrt{n}).$$

Step 2. Similarly to calculating the expectation of the quasi-score, we first consider $\mathrm{var}_0(\mathbf{U}_1)$. The rest terms can be manipulated by the similar arguments. Note that

$$\mathrm{E}_0(\mathbf{U}_1\mathbf{U}_1^\top) = \int \mathbf{U}_1(\mathbf{y})\mathbf{U}_1^\top(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y})\mathrm{d}\mathbf{y} + \int \mathbf{U}_1(\mathbf{y})\mathbf{U}_1^\top(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y}) \mathbf{T}_2^\top(\mathbf{y}) \boldsymbol{\delta} n^{-1/2}\mathrm{d}\mathbf{y} + \int \mathbf{U}_1(\mathbf{y})\mathbf{U}_1^\top(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y}) r(\mathbf{y}, \boldsymbol{\delta} n^{-1/2})\mathrm{d}\mathbf{y}. \tag{A.2}$$

The first term is $\mathrm{E}_{\mathcal{N}}(\mathbf{U}_1\mathbf{U}_1^\top)$. By Assumption (C.4), we see that

$$\left|\int U_{1k}^2(\mathbf{y}) T_{2r}(\mathbf{y}) f_{\mathcal{N}}(\mathbf{y})\mathrm{d}\mathbf{y}\right| \leq \int \left|U_{1k}^2(\mathbf{y}) T_{2r}(\mathbf{y})\right| f_{\mathcal{N}}(\mathbf{y})\mathrm{d}\mathbf{y} = O(1)$$

and for $k_1, k_2 \in \{1, \cdots, p\}$,

$$\left|\int U_{1k_1}(\mathbf{y})U_{1k_2}(\mathbf{y})T_{2r}(\mathbf{y})f_{\mathcal{N}}(\mathbf{y})\mathrm{d}\mathbf{y}\right|$$
$$\leq \frac{1}{2}\left[\int \left|U_{1k_1}^2(\mathbf{y})T_{2r}(\mathbf{y})f_{\mathcal{N}}(\mathbf{y})\right|\mathrm{d}\mathbf{y} + \int \left|U_{1k_2}^2(\mathbf{y})T_{2r}(\mathbf{y})f_{\mathcal{N}}(\mathbf{y})\right|\mathrm{d}\mathbf{y}\right] = O(1).$$

Therefore $\int \mathbf{U}_1(\mathbf{y})\mathbf{U}_1^\top(\mathbf{y})f_{\mathcal{N}}(\mathbf{y})\mathbf{T}_2^\top(\mathbf{y})\mathrm{d}\mathbf{y}$ is of order $O(\mathbf{1})$. It follows that the second term in (A.2) is of order $O(\mathbf{1}/\sqrt{n})$. By Assumption (C.3), we conclude that the third term in (A.2) is of order $o(1/\sqrt{n})$.

A direct simplification yields that $\mathrm{var}_0(\mathbf{U}_1) = \mathrm{var}_{\mathcal{N}}(\mathbf{U}_1) + O(\mathbf{1}/\sqrt{n}) = \mathbf{\Sigma}_{00} + O(\mathbf{1}/\sqrt{n})$. Go through the similar arguments for $\mathrm{var}_0(\mathbf{U}_2)$, $\mathrm{cov}_0(\mathbf{U}_1, \mathbf{U}_2^\top)$ and $\mathrm{cov}_0(\mathbf{U}_2, \mathbf{U}_1^\top)$. The variance of the quasi-score can be expressed under $f_0(\mathbf{y})$ as

$$\mathrm{var}_0\begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma}_{00} & \mathbf{\Sigma}_{10} \\ \mathbf{\Sigma}_{10} & \mathbf{\Sigma}_{11} \end{bmatrix} + O(\mathbf{1}/\sqrt{n}) = \mathbf{\Sigma} + O(\mathbf{1}/\sqrt{n}).$$

Step 3. Because $\mathbf{y}_i$'s are independent, the corresponding quasi-scores, denoted by $\mathbf{U}_{\mathrm{F},i} = \mathbf{U}(\mathbf{y}_i)$, are independent too. By Assumption (C.4), for some $\xi > 0$

$$\mathrm{E}_0\big(\|\mathbf{U}(\mathbf{y})\|^{2+\xi}\big) = \int \|\mathbf{U}(\mathbf{y})\|^{2+\xi} f_{\mathcal{N}}(\mathbf{y})\mathrm{d}\mathbf{y} + \int \|\mathbf{U}(\mathbf{y})\|^{2+\xi} f_{\mathcal{N}}(\mathbf{y})\mathbf{T}_2^\top(\mathbf{y})\boldsymbol{\delta}n^{-1/2}\mathrm{d}\mathbf{y}$$
$$+ \int \|\mathbf{U}(\mathbf{y})\|^{2+\xi} f_{\mathcal{N}}(\mathbf{y})r(\mathbf{y}, \boldsymbol{\delta}n^{-1/2})\mathrm{d}\mathbf{y} = O(1).$$

Therefore $\|\mathbf{U}_{\mathrm{F},i}\|^{2+\xi}$ has bounded mean under the true density $f_0(\mathbf{y})$. So is $\|\mathbf{U}_{\mathrm{F},i} - \mathrm{E}_0(\mathbf{U}_{\mathrm{F},i})\|^{2+\xi}$. Denote the true distribution of $\mathbf{U}_{\mathrm{F},i}$ by $F_{0,i}(\mathbf{u})$. Then

$$\lim_{n\to\infty} n^{-(1+\xi/2)} \sum_{i=1}^{n} \int \|\mathbf{u} - \mathrm{E}_0(\mathbf{U}_{\mathrm{F},i})\|^{2+\xi} \mathrm{d}F_{0,i}(\mathbf{u}) \to 0.$$

Thus Lyapounov condition is guaranteed. Applying Lyapounov central limit theorem to the quasi-score $\mathbf{U}_{\mathrm{F},i}$ indicates that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{\mathbf{U}_{\mathrm{F},i} - \mathrm{E}_0(\mathbf{U}_{\mathrm{F},i})\right\} \xrightarrow{d} N_{p+q}(\mathbf{0}, \mathbf{\Sigma}).$$

Therefore

$$\begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} \xrightarrow{d} N_{p+q}\left(\begin{bmatrix} \mathbf{\Sigma}_{01} \\ \mathbf{\Sigma}_{11} \end{bmatrix}\boldsymbol{\delta}, \mathbf{\Sigma}\right).$$

Q.E.D.

Lemma A.2. *Under the misspecification framework and the regularity conditions given the Assumptions, the GEE estimates have the following equivalence in distribution form:*

$$\sqrt{n}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \boldsymbol{\Sigma}^{-1}\begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} + o_p(\mathbf{1}),$$

In particular, with the submodel S:

$$\sqrt{n}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}}_s \end{bmatrix} = \boldsymbol{\Sigma}_s^{-1}\begin{bmatrix} \mathbf{R}_{1,n} \\ \boldsymbol{\pi}_s\mathbf{R}_{2,n} \end{bmatrix} + o_p(\mathbf{1}).$$

Proof. Consider a Taylor series expansion of the quasi-score around $(\boldsymbol{\theta}_0, \mathbf{0})$:

$$\begin{bmatrix} \mathbf{R}_{1,n}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}) \\ \mathbf{R}_{2,n}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}) \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} + \begin{bmatrix} \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \\ \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}} \times \begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} - \mathbf{0} \end{bmatrix}$$
$$+\frac{1}{2}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} - \mathbf{0} \end{bmatrix}^\top \times \begin{bmatrix} \partial^2\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta} & \partial^2\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top\partial\boldsymbol{\theta} \\ \partial^2\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\gamma} & \partial^2\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top\partial\boldsymbol{\gamma} \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}^*,\boldsymbol{\gamma}=\boldsymbol{\gamma}^*} \times \begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} - \mathbf{0} \end{bmatrix}, \tag{A.3}$$

with $\boldsymbol{\theta}^*$ being between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$, and $\boldsymbol{\gamma}^*$ between $\mathbf{0}$ and $\widehat{\boldsymbol{\gamma}}$. Recalling the consistency of the GEE estimates, it is easy to see $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 + o_p(\mathbf{1})$ and $\boldsymbol{\gamma}^* = o_p(\mathbf{1})$. Also Assumption (C.1) indicates the matrix of the second derivative in the third term of (A.3) is stochastic bounded, so the third term is of order $o_p(\mathbf{1})$. Thus, (A.3) becomes

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} + \begin{bmatrix} \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \\ \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}} \times \begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} - \mathbf{0} \end{bmatrix} + o_p(\mathbf{1}).$$

Therefore

$$\sqrt{n}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} - \mathbf{0} \end{bmatrix} = -\sqrt{n}\begin{bmatrix} \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \\ \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \end{bmatrix}^{-1}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}} \times \begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} + o_p(\mathbf{1}).$$

Again Assumption (C.2) and the law of large number yield

$$\frac{1}{\sqrt{n}}\begin{bmatrix} \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \\ \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}} = -\boldsymbol{\Sigma} + o_p(\mathbf{1})$$

and

$$\sqrt{n}\begin{bmatrix} \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{1,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \\ \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{2,n}(\boldsymbol{\theta},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top \end{bmatrix}^{-1}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\gamma}=\mathbf{0}} = -\boldsymbol{\Sigma}^{-1} + o_p(\mathbf{1}).$$

Consequently,

$$\begin{aligned}\sqrt{n}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}} - \mathbf{0} \end{bmatrix} &= \left\{\boldsymbol{\Sigma}^{-1} + o_p(\mathbf{1})\right\}\begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} + o_p(\mathbf{1}) \\ &= \boldsymbol{\Sigma}^{-1}\begin{bmatrix} \mathbf{R}_{1,n} \\ \mathbf{R}_{2,n} \end{bmatrix} + o_p(\mathbf{1}).\end{aligned}$$

This completes the proof. Q.E.D.

A.2 Proof of Theorem 1

Based on Lemma A.2, the estimator of the uncertain parameters under the full model becomes

$$\begin{aligned}\sqrt{n}\widehat{\boldsymbol{\gamma}} &= \boldsymbol{\Sigma}^{10}\mathbf{R}_{1,n} + \boldsymbol{\Sigma}^{11}\mathbf{R}_{2,n} + o_p(\mathbf{1})\\ &= \boldsymbol{\Sigma}^{11}(\mathbf{R}_{2,n} - \boldsymbol{\Sigma}_{10}\boldsymbol{\Sigma}_{00}^{-1}\mathbf{R}_{1,n}) + o_p(\mathbf{1}).\end{aligned}$$

The estimator of the uncertain parameters under the submodel S can be written as

$$\begin{aligned}\sqrt{n}\widehat{\boldsymbol{\gamma}}_{\rm s} &= \boldsymbol{\Sigma}_{\rm s}^{11}(\boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} - \boldsymbol{\Sigma}_{10,{\rm s}}\boldsymbol{\Sigma}_{00}^{-1}\mathbf{R}_{1,n}) + o_p(\mathbf{1})\\ &= \boldsymbol{\Sigma}_{\rm s}^{11}\boldsymbol{\pi}_{\rm s}(\mathbf{R}_{2,n} - \boldsymbol{\Sigma}_{10}\boldsymbol{\Sigma}_{00}^{-1}\mathbf{R}_{1,n}) + o_p(\mathbf{1}).\end{aligned} \tag{A.4}$$

A direct calculation indicates the relationship between $\widehat{\boldsymbol{\gamma}}_{\rm s}$ and $\widehat{\boldsymbol{\gamma}}$ as follows

$$\sqrt{n}\widehat{\boldsymbol{\gamma}}_{\rm s} = \sqrt{n}\boldsymbol{\Sigma}_{\rm s}^{11}\boldsymbol{\pi}_{\rm s}\big(\boldsymbol{\Sigma}^{11}\big)^{-1}\widehat{\boldsymbol{\gamma}} + o_p(\mathbf{1}). \tag{A.5}$$

Also the large sample behavior of the GEE estimators can be derived by Lemmas A.1 and A.2:

$$\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}}\end{bmatrix} \xrightarrow{d} N_{p+q}\left(\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\boldsymbol{\Sigma}_{01}\\ \boldsymbol{\Sigma}_{11}\end{bmatrix}\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}\right). \tag{A.6}$$

Now, we are going to prove the main theorem. To derive the specific form of ΔAIC, consider a Taylor series expansion of the log quasi-likelihood around $(\boldsymbol{\theta}_0, \mathbf{0})$:

$$\begin{aligned}Q(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}; \mathcal{D}) &= Q(\boldsymbol{\theta}_0, \mathbf{0}; \mathcal{D}) + \sqrt{n}\begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix}^\top \times \begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}} - \mathbf{0}\end{bmatrix}\\ &+ \frac{\sqrt{n}}{2}\begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}} - \mathbf{0}\end{bmatrix}^\top \begin{bmatrix}\partial\mathbf{R}_{1,n}(\boldsymbol{\theta}, \boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{1,n}(\boldsymbol{\theta}, \boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top\\ \partial\mathbf{R}_{2,n}(\boldsymbol{\theta}, \boldsymbol{\gamma})/\partial\boldsymbol{\theta}^\top & \partial\mathbf{R}_{2,n}(\boldsymbol{\theta}, \boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^\top\end{bmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}^*, \boldsymbol{\gamma}=\boldsymbol{\gamma}^*}\begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}} - \mathbf{0}\end{bmatrix},\end{aligned}$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\gamma}^*$ between $\mathbf{0}$ and $\widehat{\boldsymbol{\gamma}}$. It follows that

$$\begin{aligned}&Q(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}; \mathcal{D}) - Q(\boldsymbol{\theta}_0, \mathbf{0}; \mathcal{D})\\ &= \begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix}^\top \times \sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}} - \mathbf{0}\end{bmatrix} + \frac{\sqrt{n}}{2}\begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}} - \mathbf{0}\end{bmatrix}^\top \big\{-\sqrt{n}\{\boldsymbol{\Sigma} + o_p(\mathbf{1})\}\big\}\begin{bmatrix}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\\ \widehat{\boldsymbol{\gamma}} - \mathbf{0}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix}^\top \times \left\{\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix} + o_p(\mathbf{1})\right\}\\ &\quad - \frac{1}{2}\left\{\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix} + o_p(\mathbf{1})\right\}^\top \big\{\boldsymbol{\Sigma} + o_p(\mathbf{1})\big\}\left\{\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix} + o_p(\mathbf{1})\right\}\\ &= \frac{1}{2}\begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix}^\top \boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{R}_{1,n}\\ \mathbf{R}_{2,n}\end{bmatrix} + o_p(\mathbf{1}),\end{aligned}$$

where the second equality follows from Lemma A.2. In particular,

$$Q(\widehat{\boldsymbol{\theta}}, \widehat{\gamma}_{\rm s}; \mathcal{D}) - Q(\boldsymbol{\theta}_0, \mathbf{0}; \mathcal{D}) = \frac{1}{2}\begin{bmatrix} \mathbf{R}_{1,n} \\ \boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} \end{bmatrix}^\top \boldsymbol{\Sigma}_{\rm s}^{-1} \begin{bmatrix} \mathbf{R}_{1,n} \\ \boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} \end{bmatrix} + o_p(\mathbf{1}). \tag{A.7}$$

For the narrow model, it becomes

$$Q(\widehat{\boldsymbol{\theta}}, \mathbf{0}; \mathcal{D}) - Q(\boldsymbol{\theta}_0, \mathbf{0}; \mathcal{D}) = \frac{1}{2}\,\mathbf{R}_{1,n}^\top \boldsymbol{\Sigma}_{00}^{-1}\mathbf{R}_{1,n} + o_p(\mathbf{1}). \tag{A.8}$$

Recall the definition of $\Delta\text{AIC}_{n,\text{s}}$, which gives

$$\begin{aligned} \Delta\text{AIC}_{n,\text{s}} &= -2\sum_{i=1}^{n} Q(\widehat{\boldsymbol{\theta}}, \widehat{\gamma}_{\rm s}; \mathbf{y}_i) + 2\sum_{i=1}^{n} Q(\widehat{\boldsymbol{\theta}}, \mathbf{0}; \mathbf{y}_i) + 2|\text{S}/\mathcal{N}| \\ &= -2\big[Q(\widehat{\boldsymbol{\theta}}, \widehat{\gamma}_{\rm s}; \mathcal{D}) - Q(\widehat{\boldsymbol{\theta}}, \mathbf{0}; \mathcal{D})\big] + 2|\text{S}/\mathcal{N}|. \end{aligned}$$

(A.7) and (A.8) indicate that

$$\begin{aligned} \Delta\text{AIC}_{n,\text{s}} &= -2\big[Q(\widehat{\boldsymbol{\theta}}, \widehat{\gamma}_{\rm s}; \mathcal{D}) - Q(\boldsymbol{\theta}_0, \mathbf{0}; \mathcal{D})\big] + 2\big[Q(\widehat{\boldsymbol{\theta}}_{\mathcal{N}}, \mathbf{0}; \mathcal{D}) - Q(\boldsymbol{\theta}_0, \mathbf{0}; \mathcal{D})\big] + 2|\text{S}/\mathcal{N}| \\ &= -\begin{bmatrix} \mathbf{R}_{1,n} \\ \boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} \end{bmatrix}^\top \boldsymbol{\Sigma}_{\rm s}^{-1} \begin{bmatrix} \mathbf{R}_{1,n} \\ \boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} \end{bmatrix} + \mathbf{R}_{1,n}^\top \Sigma_{00}^{-1}\mathbf{R}_{1,n} + 2|\text{S}/\mathcal{N}| + o_p(\mathbf{1}). \end{aligned}$$

Using the expressions given in (A.4) and (A.5), $\Delta\text{AIC}_{n,\text{s}}$ can be further expressed as

$$\begin{aligned} &-\big(\boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} - \boldsymbol{\pi}_{\rm s}\boldsymbol{\Sigma}_{10}\boldsymbol{\Sigma}_{00}^{-1}\mathbf{R}_{1,n}\big)^\top \boldsymbol{\Sigma}_{\rm s}^{11} \big(\boldsymbol{\pi}_{\rm s}\mathbf{R}_{2,n} - \boldsymbol{\pi}_{\rm s}\boldsymbol{\Sigma}_{10}\boldsymbol{\Sigma}_{00}^{-1}\mathbf{R}_{1,n}\big) + 2|\text{S}/\mathcal{N}| + o_p(\mathbf{1}) \\ &= -\sqrt{n}\widehat{\gamma}_{\rm s}^\top \big(\boldsymbol{\Sigma}_{\rm s}^{11}\big)^{-1}\sqrt{n}\widehat{\gamma}_{\rm s} + 2|\text{S}/\mathcal{N}| + o_p(\mathbf{1}) \\ &= -\,n\widehat{\gamma}^\top \big(\boldsymbol{\Sigma}^{11}\big)^{-1}\boldsymbol{\pi}_{\rm s}^\top \boldsymbol{\Sigma}_{\rm s}^{11}\boldsymbol{\pi}_{\rm s}\big(\boldsymbol{\Sigma}^{11}\big)^{-1}\widehat{\gamma} + 2|\text{S}/\mathcal{N}| + o_p(\mathbf{1}). \end{aligned}$$

Recalling (A.6), we see that $\sqrt{n}\widehat{\gamma}\xrightarrow{d}N_q(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{11})$. Thus, the first term of $\Delta\text{AIC}_{n,\text{s}}$ converges to a noncentral chi-squared distribution and so

$$\Delta\text{AIC}_{n,\text{s}}\xrightarrow{d} -\chi^2_{|\text{S}/\mathcal{N}|}(\boldsymbol{\lambda}_{\rm s}) + 2|\text{S}/\mathcal{N}|$$

with $\boldsymbol{\lambda}_{\rm s} = n\gamma_0^\top \big(\boldsymbol{\Sigma}^{11}\big)^{-1}\boldsymbol{\pi}_{\rm s}^\top \boldsymbol{\Sigma}_{\rm s}^{11}\boldsymbol{\pi}_{\rm s}\big(\boldsymbol{\Sigma}^{11}\big)^{-1}\gamma_0$. This completes the proof. Q.E.D.

A.3 Proof of Theorem 2

From Lemmas A.1 and A.2, we have

$$\begin{aligned} \sqrt{n}\begin{bmatrix} \widehat{\boldsymbol{\theta}}_{\rm s} - \boldsymbol{\theta}_0 \\ \widehat{\gamma}_{\rm s} \end{bmatrix} &\to_d \begin{bmatrix} (\boldsymbol{\Sigma}^{00,s}\boldsymbol{\Sigma}_{01} + \boldsymbol{\Sigma}^{01,s}\boldsymbol{\pi}_{\rm s}\boldsymbol{\Sigma}_{11})\boldsymbol{\delta} + \boldsymbol{\Sigma}^{00,s}\mathbf{M}_1 + \boldsymbol{\Sigma}^{01,s}\boldsymbol{\pi}_{\rm s}\mathbf{M}_2 \\ (\boldsymbol{\Sigma}^{10,s}\boldsymbol{\Sigma}_{01} + \boldsymbol{\Sigma}^{11,s}\boldsymbol{\pi}_{\rm s}\boldsymbol{\Sigma}_{11})\boldsymbol{\delta} + \boldsymbol{\Sigma}^{10,s}\mathbf{M}_1 + \boldsymbol{\Sigma}^{11,s}\boldsymbol{\pi}_{\rm s}\mathbf{M}_2 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{\delta} + \boldsymbol{\Sigma}_{00}^{-1}\mathbf{M}_1 - \boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{\pi}_{\rm s}^\top \boldsymbol{\Sigma}_{\rm s}^{11}\boldsymbol{\pi}_{\rm s}\big(\boldsymbol{\Sigma}^{11}\big)^{-1}\boldsymbol{\Delta} \\ \boldsymbol{\Sigma}_{\rm s}^{11}\boldsymbol{\pi}_{\rm s}\big(\boldsymbol{\Sigma}^{11}\big)^{-1}\boldsymbol{\Delta} \end{bmatrix}. \end{aligned}$$

Since $\boldsymbol{\zeta}$ is a function of $(\boldsymbol{\theta}, \boldsymbol{\gamma})$, $\sqrt{n}(\widehat{\boldsymbol{\zeta}}_{\text{s}} - \boldsymbol{\zeta}_{0})$ can be expanded by Taylor expansion and a delta method as:

$$\begin{aligned}
&\sqrt{n}(\widehat{\boldsymbol{\zeta}}_{\text{s}} - \boldsymbol{\zeta}_{0}) = \sqrt{n}\{\boldsymbol{\zeta}(\widehat{\boldsymbol{\theta}}_{\text{s}}, \widehat{\gamma}_{\text{s}}) - \boldsymbol{\zeta}(\boldsymbol{\theta}_{0}, \boldsymbol{\delta}/\sqrt{n})\} \\
&= \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\theta}}\right)^{\top} \sqrt{n}(\widehat{\boldsymbol{\theta}}_{\text{s}} - \boldsymbol{\theta}_{0}) + \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\gamma}_{\text{s}}}\right)^{\top} \sqrt{n}(\widehat{\gamma}_{\text{s}} - \gamma_{0}) - \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\gamma}}\right)^{\top} \boldsymbol{\delta} + o_p(\mathbf{1}) \\
&\xrightarrow{d} \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\theta}}\right)^{\top} \{\boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{\delta} + \boldsymbol{\Sigma}_{00}^{-1}\mathbf{M}_1 - \boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{\pi}_{\text{s}}^{\top}\boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\Delta}\} \\
&\qquad + \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\gamma}_{\text{s}}}\right)^{\top} \boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\Delta} - \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\gamma}}\right)^{\top} \boldsymbol{\delta} \\
&= \left\{\left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\theta}}\right)^{\top} \boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{01} - \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\gamma}}\right)^{\top}\right\} \boldsymbol{\delta} - \left\{\left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\theta}}\right)^{\top} \boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{\pi}_{\text{s}}^{\top} - \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\gamma}_{\text{s}}}\right)^{\top}\right\} \\
&\qquad \boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\Delta} + \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\theta}}\right)^{\top} \boldsymbol{\Sigma}_{00}^{-1}\mathbf{M}_1 \\
&= \left(\frac{\partial \boldsymbol{\zeta}}{\partial \boldsymbol{\theta}}\right)^{\top} \boldsymbol{\Sigma}_{00}^{-1}\mathbf{M}_1 + \boldsymbol{\omega}^{\top}\boldsymbol{\delta} - \boldsymbol{\omega}^{\top}\boldsymbol{\pi}_{\text{s}}^{\top}\boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\Delta}.
\end{aligned}$$

Therefore,

$$\sqrt{n}(\widehat{\boldsymbol{\zeta}}_{\text{s}} - \boldsymbol{\zeta}_{0}) \xrightarrow{d} \boldsymbol{\Omega}_{\text{s}} = \boldsymbol{\Omega}_{0} + \boldsymbol{\omega}^{\top}\boldsymbol{\delta} - \boldsymbol{\omega}^{\top}\boldsymbol{\pi}_{\text{s}}^{\top}\boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\Delta}$$

where $\boldsymbol{\Omega}_{0} \sim N_p(\mathbf{0}, \boldsymbol{\tau}_{0}^2)$. The limiting variable $\boldsymbol{\Omega}_{\text{s}}$ follows Normal distribution with mean $\boldsymbol{\omega}^{\top}\boldsymbol{\delta} - \boldsymbol{\omega}^{\top}\boldsymbol{\pi}_{\text{s}}^{\top}\boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\delta}$ and variance $\boldsymbol{\tau}_0^2 + \boldsymbol{\omega}^{\top}\boldsymbol{\pi}_{\text{s}}^{\top}\boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}\boldsymbol{\omega}$.

Q.E.D.

A.4 Proof of Theorem 3

Since the compromise estimator has the form of $\widehat{\boldsymbol{\zeta}} = \sum_{\text{s}} p(\text{S}|\boldsymbol{\Delta})\widehat{\boldsymbol{\zeta}}_{\text{s}}$ and $(\widehat{\boldsymbol{\theta}}_{\text{s}}, \widehat{\gamma}_{\text{s}})$ is a linear function of $(\widehat{\boldsymbol{\theta}}, \widehat{\gamma})$ in addition to a term of $o_p(1)$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_{0}) \xrightarrow{d} \boldsymbol{\Omega} = \sum_{\text{s}} p(\text{S}|\boldsymbol{\Delta})\boldsymbol{\Omega}_{\text{s}} = \boldsymbol{\Omega}_0 + \boldsymbol{\omega}^{\top}\boldsymbol{\delta} - \boldsymbol{\omega}^{\top}\sum_{\text{s}} p(\text{S}|\boldsymbol{\Delta})\boldsymbol{\pi}_{\text{s}}^{\top}\boldsymbol{\Sigma}_{\text{s}}^{11}\boldsymbol{\pi}_{\text{s}}(\boldsymbol{\Sigma}^{11})^{-1}\boldsymbol{\Delta}$$

The limiting variable $\boldsymbol{\Omega}$ has mean $\boldsymbol{\omega}^{\top}\boldsymbol{\delta} - \boldsymbol{\omega}^{\top}\text{E}[\widehat{\boldsymbol{\delta}}(\boldsymbol{\Delta})]$ and variance $\boldsymbol{\tau}_{0}^2 + \boldsymbol{\omega}^{\top}\text{var}[\widehat{\boldsymbol{\delta}}(\boldsymbol{\Delta})]\boldsymbol{\omega}$.

Bibliography

Hjort, N. and Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association* **98**, 879–899.

Liang, K. Y. and Zeger, S. L. (1986), Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics* **11**, 59–67.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447.

Global Biostatistical Science, Amgen Inc., Thousand Oaks, California 91320, U.S.A.
E-mail: (huiy@amgen.com)
Department of Mathematics, Shandong University of Technology, Zibo 255000, China.
E-mail: (mathlinpeng@163.com)
School of Mathematical Science, Capital Normal University, Beijing 100037, China.
E-mail: (ghzou@amss.ac.cn)

Department of Statistics, George Washington University, Washington, D.C. 20052, U.S.A.
E-mail: (hliang@gwu.edu)