Biometrics 000, 1–35 DOI: 000
000 0000

Supporting Information for
"Grouped Generalized Estimating Equations for Longitudinal Data Analysis"

Tsubasa Ito

Faculty of Economics and Business, Hokkaido University, Hokkaido, Japan

email: tito@econ.hokudai.ac.jp

and

Shonosuke Sugasawa

Center for Spatial Information Science, University of Tokyo, Tokyo, Japan

email: sugasawa@csis.u-tokyo.ac.jp

Summary: Section S.1 presents additional assumptions results for the grouped GEE. In Section S.2, we prove Theorem 1; the existence and weak consistency of the grouped GEE estimators and the classification consistency of the grouping variables. In Section S.3, we prove Theorem 2; asymptotic normality of the grouped GEE estimators. Section S.4 presents some asymptotic properties of the estimated unstructured working correlation matrix.
Key words: Estimating equation; Grouping; k-means algorithm; Unobserved heterogeneity.

This paper has been submitted for consideration for publication in *Biometrics*

S.1. Additional assumptions

We give the following notations similar to those in Xie and Yang (2003), which are needed to provide assumptions assuring a sufficient conditions for the conditions (I*), (L*) and (CC) in Xie and Yang (2003), under which the existence, weak consistency and asymptotic normality of the GEE estimator hold:

$$\pi = \sup_{\boldsymbol{\beta},\boldsymbol{\gamma}} \frac{\lambda_{\max}(\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}))}{\lambda_{\min}(\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}))}, \quad \xi = \tau \max_{1\leqslant i\leqslant n, 1\leqslant t\leqslant T} \max_{1\leqslant g\leqslant G} \boldsymbol{x}_{it}^{\top}\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1}\boldsymbol{x}_{it}.$$

In addition to the Assumption (A1)-(A5), we assume the following regularity assumptions for the grouped GEE:

Assumption S.1:

(A6) For all $i = 1,\dots,n$ and $t = 1,\dots,T$, $a'(\theta_{it})$ is uniformly three times continuously differentiable, $a''(\theta_{it})$ is uniformly bounded away from 0, and $u(\eta_{it})$ is uniformly four times continuously differentiable and $u'(\eta_{it})$ is uniformly bounded away from 0.

(A7) For all $i = 1,\dots,n$, there exist positive constants, b_1, b_2 and b_3, such that $b_1 \leqslant \lambda_{\min}((nT)^{-1}\sum_{i=1}^n \boldsymbol{X}_i^{\top}\boldsymbol{X}_i) \leqslant \lambda_{\max}((nT)^{-1}\sum_{i=1}^n \boldsymbol{X}_i^{\top}\boldsymbol{X}_i) \leqslant b_2$ and $\lambda_{\max}(T^{-1}\boldsymbol{X}_i^{\top}\boldsymbol{X}_i) \leqslant b_3$. For all i, there is q such that $x_{itq} \neq x_{it'q}$ for some $t \neq t'$.

(A8) (i) $\pi^2\xi \to 0$ and (ii) $v\pi\xi \to 0$ for $v = (\sqrt{nT} \wedge T\pi/\min_{1\leqslant i\leqslant n,1\leqslant t\leqslant T}\{\sigma^2(\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_{g_i^0}^0)\})$.

(A9) (i) $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \max_{1\leqslant k,l\leqslant T} |\{\boldsymbol{R}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) - \boldsymbol{R}(\boldsymbol{\alpha},\boldsymbol{\beta}^0,\boldsymbol{\gamma})\}_{k.l}| = O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2})$ for any $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, (ii) for any $\boldsymbol{\gamma}$, $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \max_{1\leqslant k,l\leqslant T} |\{\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})\}_{k.l}| = O_p(n^{-1/2} \vee \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2})$ and $\max_{1\leqslant k,l\leqslant T} |\{\widehat{\boldsymbol{R}}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\}_{k.l}| = O_p(n^{-1/2})$, and (iii) for any $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_{i*}$ whose only ith component differs from that of $\boldsymbol{\gamma}$, $\max_{1\leqslant k,l\leqslant T} |\{R(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}_{i*}) - R(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\}_{kl}| = O_p(1/n)$. (iv) for any $\boldsymbol{\beta} \in \mathcal{B}$ and all $\delta > 0$, $\max_{1\leqslant k,l\leqslant T} |\{\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{kl}| = o_p(T^{-\delta})$ for $\boldsymbol{\gamma} \in \Gamma$, where $\Gamma = \{\boldsymbol{\gamma} = (g_1,\dots,g_n) : n^{-1}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\} = o_p(T^{-\delta}) \ \text{ for all } \ \delta > 0\}$.

Assumption (A6) requires that the marginal variance of y_{it} is uniformly larger than 0 for any $\boldsymbol{\beta} \in \mathcal{B}$ and $\boldsymbol{x}_{it} \in \mathcal{X}$ for all $i = 1,\dots,n$ and $t = 1,\dots,T$. The boundedness of $a^{(k)}(\theta_{it})$ and

$u^{(k)}(\eta_{it})$ for $\boldsymbol{\beta}_g$'s in a local neighborhood around $\boldsymbol{\beta}_g^0$ is also required to ensure the asymptotic properties of GEE estimators, which is satisfied from Assumptions (A1). Assumption (A7) is also imposed well and ensures combined with Assumptions (A2) (i) that $\overline{\boldsymbol{H}}_g(\boldsymbol{\beta}_g)$, $\overline{\boldsymbol{M}}_g(\boldsymbol{\beta}_g)$ and so on are invertible when n or T is sufficiently large. Assumption (A8) is the technical assumption similar to the assumptions in Lemma A.2 (ii), and A.3 (ii) of Xie and Yang (2003), which ensure the sufficient conditions for the conditions (I*) and (CC) in Xie and Yang (2003). The idea behind Assumption (A9) is similar to that of the condition (A4) in Wang (2011), that is, it is essential to approximate $\boldsymbol{S}_g(\boldsymbol{\beta}_g)$ by $\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)$ whose moments are easier to evaluate. For this, Assumption (A9) (i) and (ii) say that the estimated working correlation matrix can be approximated by $\overline{\boldsymbol{R}}(\boldsymbol{\beta}^0, \boldsymbol{\gamma})$ in a local neighborhood of $\boldsymbol{\beta}_g^0$'s and $\overline{\boldsymbol{\alpha}}$. Assumption (A9) (iii) says that each cluster is linearly additive for estimating the working correlation matrix. Then, this is an intuitively reasonable assumption that most of the working correlation matrix estimators satisfy. Assumption (A9) (iv) says that the estimated working correlation matrix can be approximated by $\overline{\boldsymbol{R}}(\boldsymbol{\beta}, \boldsymbol{\gamma}^0)$ if groups are consistently classified to their true groups on average. In Section S.4, we provide the accuracy of these approximations under the unstructured working correlation matrix.

We use the following notations. The notation $a_{nT} \lesssim b_{nT}$ means that $a_{nT} \leqslant Cb_{nT}$ for all n and T, for some constant C that does not depends on n and T. For a column vector $\boldsymbol{a}$, we use $\boldsymbol{a}^\top$ to denote the transpose of $\boldsymbol{a}$ and $||\boldsymbol{a}||$ to denote the Euclidean norm of $\boldsymbol{a}$. For a matrix $\boldsymbol{A}$, $\{\boldsymbol{A}\}_{kl}$ denotes the (k,l)-element of $\boldsymbol{A}$, $\lambda_{\min}(\boldsymbol{A})$ ($\lambda_{\max}(\boldsymbol{A})$) denotes the smallest (largest) eigenvalue of $\boldsymbol{A}$, $\boldsymbol{A}^\top$ denotes the transpose of $\boldsymbol{A}$ and $||\boldsymbol{A}||_F = \{\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{A})\}^{1/2}$ is the Frobenius norm of $\boldsymbol{A}$. We use the notation $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$.

S.2. Proof of Theorem 1

First of all, we need to show the next lemma.

LEMMA 1: *Suppose the Assumptions (A1)-(A9). If $n/T^{\nu} \to 0$ for some $\nu > 0$, it holds that for all $\delta > 0$,*

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}_{nT}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\widehat{g}_i(\boldsymbol{\beta}) \neq g_i^0\} = o_p(T^{-\delta}),$$

where $\widehat{g}_i(\boldsymbol{\beta})$ is obtained by (2.3) in the main text.

Proof. For any $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}_{i0}$ is obtained by replacing only its ith element with g_i^0, that is $\boldsymbol{\gamma}_{i0} = (g_1, \ldots, g_{i-1}, g_i^0, g_{i+1}, \ldots, g_n)$. Note that, from the definition of $\widehat{g}_i(\boldsymbol{\beta})$, we have, for all $g = 1, \ldots, G$,

$$\begin{aligned}\mathbf{1}\{\widehat{g}_i(\boldsymbol{\beta}) = g\} \leqslant &\mathbf{1}\Big\{\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\} \\ &\leqslant \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}\Big\}.\end{aligned}$$

Then, we can write

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\widehat{g}_i(\boldsymbol{\beta}) \neq g_i^0\} = \sum_{g=1}^{G} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\{\widehat{g}_i(\boldsymbol{\beta}) = g\} \leqslant \sum_{g=1}^{G} \frac{1}{n}\sum_{i=1}^{n} Z_{ig}(\boldsymbol{\beta}_g),$$

where

$$\begin{aligned}Z_{ig}(\boldsymbol{\beta}_g) = &\mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\Big\{\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\} \\ &\leqslant \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}\Big\}.\end{aligned}$$

Similar to the proof of Lemma B.4 in Bonhomme and Manresa (2015), we start by bounding $Z_{ig}(\boldsymbol{\beta}_g)$ on $\boldsymbol{\beta} \in \mathcal{B}_{nT}$ by a quantity that does not depend on $\boldsymbol{\beta}$. Denote

$$\begin{aligned}W_{ig}(\boldsymbol{\beta}) = &\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\} \\ &- \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\},\end{aligned}$$

then we have

$$Z_{ig}(\boldsymbol{\beta}_g) = \mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\{W_{ig}(\boldsymbol{\beta}) \leqslant 0\} \leqslant \mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\{W_{ig}(\boldsymbol{\beta}^0) \leqslant |W_{ig}(\boldsymbol{\beta}^0) - W_{ig}(\boldsymbol{\beta})|\}.$$

We have

$$\begin{aligned}|W_{ig}(\boldsymbol{\beta}^0) - W_{ig}(\boldsymbol{\beta})| \leqslant & \Big|\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0, \boldsymbol{\gamma}_{i0})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})\} \\ & \quad - \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}\Big| \\ & + \Big|\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0, \gamma)\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g})\} \\ & \quad - \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \gamma)\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g})\}\Big| \\ \equiv & K^{(1)}_{ig}(\boldsymbol{\beta}) + K^{(2)}_{ig}(\boldsymbol{\beta}).\end{aligned}$$

We can write

$$\begin{aligned}K^{(1)}_{ig}(\boldsymbol{\beta}) \leqslant & |\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})\}^\top \{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0, \boldsymbol{\gamma}_{i0}) - \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})\}| \\ & + 2|\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})\}| \\ & + \{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i0})\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\} \\ \equiv & \sum_{j=1}^{3} I_j.\end{aligned}$$

Since $A_{it}(\boldsymbol{\beta}_g) < \infty$ for all $i = 1, \dots, n$ and $t = 1, \dots, T$, for I_1, we can write From Assumption (A1) , (A5) and (A9) (i), there is a constant C_1, independent of n and T such that

$$\sup_{\beta \in \mathcal{B}_{nT}} I_1 = C_1 C T \lambda^{-1/2}_{\min}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\Big(\frac{1}{T}\sum_{j=1}^{T}\varepsilon^2_{it}\Big).$$

For I_2, from Taylor expansion, for $\boldsymbol{\beta}^*_{g_i^0}$ between $\boldsymbol{\beta}^0_{g_i^0}$ and $\boldsymbol{\beta}_{g_i^0}$, we have

$$m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}) = \phi \boldsymbol{A}_i(\boldsymbol{\beta}^*_{g_i^0})\boldsymbol{\Delta}_i(\boldsymbol{\beta}^*_{g_i^0})\boldsymbol{X}_i(\boldsymbol{\beta}^0_{g_i^0} - \boldsymbol{\beta}_{g_i^0}). \tag{1}$$

Since $\max_{1 \leqslant i \leqslant n}\max_{1 \leqslant t \leqslant T} u'(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_g) < \infty$ from Assumptions (A1) and (A6), we have

$$\begin{aligned}I_2 \lesssim & ||\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0, \boldsymbol{\gamma}_{i0})\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0})\}|| \cdot ||\boldsymbol{\varepsilon}_i|| \\ \lesssim & \lambda_{\max}(\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0, \boldsymbol{\gamma}_{i0}))\{(\boldsymbol{\beta}^0_{g_i^0} - \boldsymbol{\beta}_{g_i^0})\boldsymbol{X}_i^\top\boldsymbol{\Delta}_i(\boldsymbol{\beta}^*_{g_i^0})\boldsymbol{A}_i^2(\boldsymbol{\beta}^*_{g_i^0})\boldsymbol{\Delta}_i(\boldsymbol{\beta}^*_{g_i^0})\boldsymbol{X}_i(\boldsymbol{\beta}^0_{g_i^0} - \boldsymbol{\beta}_{g_i^0})\}^{1/2}||\boldsymbol{\varepsilon}_i|| \\ \lesssim & \lambda_{\max}(\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0, \boldsymbol{\gamma}_{i0}))\lambda^{1/2}_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)||\boldsymbol{\beta}_{g_i^0} - \boldsymbol{\beta}^0_{g_i^0}||(\boldsymbol{\varepsilon}_i^\top\boldsymbol{\varepsilon}_i)^{1/2}.\end{aligned}$$

Then, from Assumptions (A5), (A7) there is a constant C_2, independent of n and T such that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} I_2 \leqslant C_2 CT\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2\Big)^{1/2}.$$

As is the case with I_2, there is a constant C_3, independent of n and T such that $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} I_3 \leqslant C_3C^2T\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau$. For $K_{ig}^{(2)}(\boldsymbol{\beta})$, we can write

$$\begin{aligned}K_{ig}^{(2)}(\boldsymbol{\beta}) \leqslant& |\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}^\top\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}| \\ &+ 2|\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}^\top\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}| \\ &+ \{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}^\top\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}.\end{aligned}$$

From the similar argument for $K_{ig}^{(1)}(\boldsymbol{\beta})$, we can bound $K_{ig}^{(2)}(\boldsymbol{\beta})$ by $C_4(CT\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} + C^2T\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau)$ for some $C_4 > 0$. Next, we will bound $W_{ig}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})$ from below. It can be written as

$$\begin{aligned}W_{ig}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}) =& \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)\}^\top\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}_{i0})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)\} \\ &+ \{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}^\top\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\} \\ &+ 2\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}^\top\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)\} \\ \equiv& \sum_{j=1}^{3} J_j.\end{aligned}$$

From Assumption (A1) , (A5) and (A9) (iii),, there is a constant C_5, independent of C and T, such that $J_1 \geqslant -C_5(T/n)(\sum_{t=1}^{T}\varepsilon_{it}^2/T)$. For J_2, we have

$$\begin{aligned}J_2 =& \{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}^\top\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\} \\ &+ \{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\}^\top\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\}\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)\} \\ \equiv& J_{21} + J_{22}.\end{aligned}$$

For J_{21}, by using (1), we have for $\boldsymbol{\beta}_{g_i}^*$ between $\boldsymbol{\beta}_{g_i^0}^0$ and $\boldsymbol{\beta}_g^0$,

$$J_{21} = (\boldsymbol{\beta}_{g_i^0}^0 - \boldsymbol{\beta}_g^0)^\top\boldsymbol{X}_i^\top\boldsymbol{\Delta}_i(\boldsymbol{\beta}_{g_i}^*)\boldsymbol{A}_i(\boldsymbol{\beta}_{g_i}^*)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\boldsymbol{A}_i(\boldsymbol{\beta}_{g_i}^*)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_{g_i}^*)\boldsymbol{X}_i(\boldsymbol{\beta}_{g_i^0}^0 - \boldsymbol{\beta}_g^0).$$

From Assumption (A7), J_{21} is at least of order $O_p(T)$. Then, from Assumption (A2) (ii) there is a constant C_6^*, independent of C and T, such that $J_{21} \geqslant C_6^*T$. From Assumptions

(A5) and (A9) (ii), it can be shown that J_{22} is dominated by J_{21}, then there is a constant C_6, independent of C and T, such that $J_2 \geqslant C_6 T$. Denote $\widetilde{\boldsymbol{\varepsilon}}_i = (\boldsymbol{R}^0)^{-1/2}\boldsymbol{\varepsilon}_i$. For J_3, we have

$$\begin{aligned}J_3 =&2\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_g)\}^\top \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})(\boldsymbol{R}^0)^{1/2}\widetilde{\boldsymbol{\varepsilon}}_i\\ &+ 2\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_g)\}^\top \{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})\}\\ \equiv& J_{31} + J_{32}.\end{aligned}$$

From Assumption (A9) (ii), J_{32} is dominated by J_{31}. Let $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ be the eigendecomposition of $\overline{\boldsymbol{R}}^{-1/2}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})(\boldsymbol{R}^0)^{1/2}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_T)$ for $\lambda_1 \geqslant,\ldots,\lambda_T$ is a diagonal matrix formed from the eigenvalues and $\boldsymbol{U}$ is the corresponding eigenvectors of $\overline{\boldsymbol{R}}^{-1/2}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})(\boldsymbol{R}^0)^{1/2}$. Then we can write

$$\begin{aligned}J_3 =&\{m^*(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m^*(\boldsymbol{X}_i\boldsymbol{\beta}^0_g)\}^\top \boldsymbol{\Lambda}\widetilde{\boldsymbol{\varepsilon}}^*_i(1+o_p(1))\\ =&C_7\sum_{t=1}^T \lambda_t\{m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_{g_i^0}) - m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_g)\}\widetilde{\varepsilon}^*_{it}(1+o_p(1)),\end{aligned}$$

for $m^*(\boldsymbol{X}_i\boldsymbol{\beta}_g) = \boldsymbol{U}\overline{\boldsymbol{R}}^{-1/2}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})m(\boldsymbol{X}_i\boldsymbol{\beta}_g)$ and $\widetilde{\boldsymbol{\varepsilon}}^*_i = \boldsymbol{U}\widetilde{\boldsymbol{\varepsilon}}_i$. Combined with the above results, we thus obtain

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} Z_{ig}(\boldsymbol{\beta}_g)\\ &\leqslant \mathbf{1}\{g_i^0 \neq g\}\\ &\quad\times \mathbf{1}\Big\{ - C_5\frac{T}{n}\Big(\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}^2\Big) + C_6T + C_7\sum_{t=1}^T\lambda_t\{m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_{g_i^0}) - m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_g)\}\widetilde{\varepsilon}^*_{it}(1+o_p(1))\\ &\qquad \leqslant C_1CT\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\Big(\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}^2\Big) + C_2CT\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\Big(\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}^2\Big)^{1/2}\\ &\qquad\quad + C_3C^2T\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau + C_4(CT\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} + C^2T\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau)\Big\}.\end{aligned}$$

Since the right-hand side of the above inequality does not depend on $\boldsymbol{\beta}_g$ for $g = 1,\ldots,G$, we can denote it as $\widetilde{Z}_{ig}$. As a result, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\widehat{g}_i(\boldsymbol{\beta}) \neq g_i^0\} \leqslant \frac{1}{n}\sum_{i=1}^n\sum_{g=1}^G\widetilde{Z}_{ig}.$$

Using standard probability algebra, we have for all g and M in Assumption (A4) and for any $0 < c < 1$,

$$
\begin{aligned}
&P(\widetilde{Z}_{ig} = 1)\\
\leqslant &P\Big(- C_5 \frac{1}{n}\Big(\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}^2\Big) + C_6 + \frac{1}{T} C_7 \sum_{t=1}^{T} \lambda_t \{ m^*(\boldsymbol{x}_{it}^\top \boldsymbol{\beta}_{g_i^0}^0) - m^*(\boldsymbol{x}_{it}^\top \boldsymbol{\beta}_g^0)\} \widetilde{\varepsilon}_{it}^*(1 + o_p(1))\\
&\qquad \leqslant C_1 C \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2\Big) + C_2 C \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2\Big)^{1/2}\\
&\qquad\quad + C_3 C^2 \lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau + C_4(C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} + C^2\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau)\Big)\\
\leqslant &P\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant n^{1-c}M\Big) + P\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant \lambda_{\min}^{1/2}(\overline{\boldsymbol{H}}^*)\tau^{-1/2}M\Big)\\
&\ + P\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant \lambda_{\min}(\overline{\boldsymbol{H}}^*)\tau^{-1}M\Big)\\
&\qquad + P\Big(\frac{1}{T} C_7 \sum_{t=1}^{T} \lambda_t \{ m^*(\boldsymbol{x}_{it}^\top \boldsymbol{\beta}_{g_i^0}^0) - m^*(\boldsymbol{x}_{it}^\top \boldsymbol{\beta}_g^0)\} \widetilde{\varepsilon}_{it}^*(1 + o_p(1))\\
&\qquad\qquad \leqslant C_5 n^{-c} M - C_6 + C_1 C M + C_2 C\sqrt{M}\\
&\qquad\qquad\quad + C_3 C^2 \lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau + C_4(C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} + C^2\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau)\Big).
\end{aligned}
$$

From Markov's inequality, we have for any $\delta > 0$,

$$
P\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant n^{1-c}M\Big) \leqslant \exp\Big(-n^{1-c}M\Big) E\Big[\exp\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2\Big)\Big].
$$

Since $E[T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^2] = 1$ and $\mathrm{Var}(T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^2) < \infty$ from Assumption (A4), we have $T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^2 = O_p(1)$. Then, we have $P(T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant n^{1-c}M) = o_p(T^{-\delta})$ for any $\delta > 0$. Similarly, we have

$$
\begin{aligned}
&P\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant \lambda_{\min}^{1/2}(\overline{\boldsymbol{H}}^*)\tau^{-1/2}M\Big)\\
&\quad \leqslant \exp\Big(-\lambda_{\min}^{1/2}(\overline{\boldsymbol{H}}^*)\tau^{-1/2}M\Big) E\Big[\exp\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2\Big)\Big] = o_p(T^{-\delta}),
\end{aligned}
$$

where the second inequality follows from Assumption (A3). Similarly, we have

$$
P\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \geqslant \lambda_{\min}(\overline{\boldsymbol{H}}^*)\tau^{-1}M\Big) = o_p(T^{-\delta}).
$$

For the last probability,

$$P\Big(\frac{1}{T}C_7\sum_{t=1}^{T}\lambda_t\{m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{g_i^0}^0)-m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_g^0)\}\widetilde{\varepsilon}_{it}^*(1+o_p(1))$$

$$\leqslant C_5n^{-c}M-C_6+C_1CM+C_2C\sqrt{M}$$

$$+C_3C^2\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau+C_4(C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}+C^2\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau)\Big),$$

the right-hand side of the inequality in the probability, the first and the last two terms are dominated by other terms as $n,T\to\infty$. Then, by taking a sufficiently small C, for $\eta>0$, the probability can be bounded above by

$$P\Big(\Big|C_7\sum_{t=1}^{T}\lambda_t\{m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{g_i^0}^0)-m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_g^0)\}\widetilde{\varepsilon}_{it}^*\Big|\geqslant T\eta\Big).$$

Moreover, it is noted that $m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{g_i^0}^0)-m^*(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_g^0)=O_p(1)$ for all i and t, and λ_t's can be bounded by the eigenvalues of $\overline{\boldsymbol{R}}^{-1/2}(\boldsymbol{\beta}^0,\boldsymbol{\gamma})(\boldsymbol{R}^0)^{1/2}$ multiplied by a constant. Then, the left-hand side of the inequality is a linear combination of $\widetilde{\varepsilon}_{it}^*$, and its expectation is 0, and the order of its variance is at most $O(T+\tau)$. Since $\widetilde{\varepsilon}_{it}^*$ for $t=1,\ldots,T$ are uncorrelated, we can use Theorem 6.2 in Rio (2000), in which the second term of the right-hand side of the equation (6.5) vanishes in this case due to the uncorrelatedness of $\widetilde{\varepsilon}_{it}$'s. Thus, by using the consequence of Theorem 6.2 in Rio (2000) for $\lambda=T\eta/4$, $r=T^{1/2}$ and $s_n^2=T+\tau$, the probability above is bounded above by $4\{1+T^2\eta^2/(16T^{1/2}(T+\tau)))\}^{-T^{1/2}/2}=o(T^{-\delta})$ for any $\delta>0$. This ends the proof.

Similar to Wang (2011), in order to prove the consistency it is essential to approximate $\boldsymbol{S}_g(\boldsymbol{\beta}_g)$, $\boldsymbol{H}_g(\boldsymbol{\beta}_g)$ and so on by $\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)$ and $\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g)$ whose moments are easier to evaluate. The following lemmas 2 - 8 establish the accuracy of these approximations, which play important roles in deriving the asymptotic normality.

Lemma 2: *Suppose the Assumptions (A1)-(A9). If $n/T^\nu\to 0$ for some $\nu>0$, it holds* *that, for all* $g = 1, \ldots, G$ *and all* $\delta > 0$*,*

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma} ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\{\boldsymbol{S}_g(\boldsymbol{\beta}_g) - \boldsymbol{S}_g^*(\boldsymbol{\beta}_g)\}|| = O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)nT)o_p(T^{-\delta}),$$

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma} ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\{\overline{\boldsymbol{S}}_g(\boldsymbol{\beta}_g) - \overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)\}|| = O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)nT)o_p(T^{-\delta}).$$

Proof. We will show the second part of the lemma. Form Assumption (A9) (ii), the first part of the lemma can be shown similarly by replacing $\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})$ and $\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)$ with $\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})$ and $\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)$ respectively. It can be written as

$$\begin{aligned}
&\overline{\boldsymbol{S}}_g(\boldsymbol{\beta}_g) - \overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)\\
=&\sum_{i=1}^n \mathbf{1}\{g_i = g\}\boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}\\
&-\sum_{i=1}^n \mathbf{1}\{g_i^0 = g\}\boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}\\
=&\sum_{i=1}^n \mathbf{1}\{g_i^0 = g\}\boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\{\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}\\
&+\sum_{i=1}^n (\mathbf{1}\{g_i = g\} - \mathbf{1}\{g_i^0 = g\})\boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}\\
\equiv& I_1 + I_2.
\end{aligned}$$

For I_1, we have

$$\begin{aligned}
I_1 =& \sum_{i:g_i^0=g}\sum_{t_1,t_2=1}^T \{\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{t_1,t_2} A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\{y_{it_2} - m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g)\}\boldsymbol{x}_{it_1}\\
=& \sum_{t_1=1}^T\sum_{t_2=1}^T \{\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{t_1,t_2}\\
&\times \Big[\sum_{i:g_i^0=g} A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\{A_{it_2}^{1/2}(\boldsymbol{\beta}_g^0)\varepsilon_{it_2} + m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g^0) - m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g)\}\boldsymbol{x}_{it_1}\Big].
\end{aligned}$$

It is noted that we have

$$E\Big[\Big|\Big|\sum_{i:g_i^0=g} A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)A_{it_2}^{1/2}(\boldsymbol{\beta}_g^0)\varepsilon_{it_2}\boldsymbol{x}_{it_1}\Big|\Big|^2\Big] \lesssim \sum_{i:g_i^0=g} \boldsymbol{x}_{it_1}^\top\boldsymbol{x}_{it_1} = O(n),$$

and

$$
\begin{aligned}
&\sup_{\beta\in\mathcal{B}_{nT}}\Big|\Big|\sum_{i:g_i^0=g} A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\{m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g^0)-m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g)\}\boldsymbol{x}_{it_1}\Big|\Big|^2\\
&\leqslant \sup_{\beta\in\mathcal{B}_{nT}}\sum_{i:g_i^0=g}\Big|\Big|A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\dot{m}(\{\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g\}^*)\boldsymbol{x}_{it_2}^\top(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)\boldsymbol{x}_{it_1}\Big|\Big|^2\\
&\lesssim \sup_{\beta\in\mathcal{B}_{nT}}\sum_{i:g_i^0=g}(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)\boldsymbol{x}_{it_2}\boldsymbol{x}_{it_2}^\top(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)\boldsymbol{x}_{it_1}^\top\boldsymbol{x}_{it_1}\\
&=O_p(n\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau).
\end{aligned}
$$

It is noted that $\max_{1\leqslant,k,l\leqslant T}|\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)-\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}_{kl}|=o_p(T^{-\delta})$ for $\boldsymbol{\gamma}\in\Gamma$ from Assumption (A9) (iv). Then, we have $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}||I_1||=O_p(n^{1/2}T^2)o_p(T^{-\delta})$. For I_2, we have from the triangle inequality

$$
||I_2||^2\leqslant\sum_{i=1}^n\mathbf{1}\{g_i\neq g_i^0\}\sum_{i=1}^n||\boldsymbol{X}_i^\top\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}||^2
$$

Since we have

$$
\begin{aligned}
&||\boldsymbol{X}_i^\top\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}||^2\\
&\lesssim\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)||\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_g)||^2=O_p(T^2),
\end{aligned}
$$

we have $\sup_{\boldsymbol{\gamma}\in\Gamma}||I_2||=O_p(nT)o_p(T^{-\delta})$, which ends the proof.

LEMMA 3: *Suppose the Assumptions (A1)-(A9). It holds that, for all* $g=1,\dots,G$,

$$
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}|||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\{\boldsymbol{S}_g(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g(\boldsymbol{\beta}_g)\}||=O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)T^2).
$$

Proof. From Lemma 2, it is enough to show that

$$
||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\{\boldsymbol{S}_g^*(\boldsymbol{\beta}_g^0)-\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0)\}||=O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)T^2).
$$

The proof is almost the same as that of Lemma 3.1 in Wang (2011). Let $Q=\{q_{j_1,j_2}\}_{1\leqslant j_1,j_2\leqslant T}$

denote the matrix $\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)$. Then,

$$\begin{aligned}
&\boldsymbol{S}_g^*(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)\\
=&\sum_{i=1}^n\sum_{t_1=1}^T\sum_{t_2=1}^T\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{t_1,t_2}A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\{y_{it_2}-m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g)\}\boldsymbol{x}_{it_1}\\
=&\sum_{t_1=1}^T\sum_{t_2=1}^T\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{t_1,t_2}\\
&\qquad\times\Big[\sum_{i=1}^n A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\{A_{it_2}^{1/2}(\boldsymbol{\beta}_g^0)\varepsilon_{it_2}+m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g^0)-m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g)\}\boldsymbol{x}_{it_1}\Big]
\end{aligned}$$

Note that

$$E\Big[\Big|\Big|\sum_{i=1}^n A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)A_{it_2}^{1/2}(\boldsymbol{\beta}_g^0)\varepsilon_{it_2}\boldsymbol{x}_{it_1}\Big|\Big|^2\Big]\lesssim\sum_{i=1}^n\boldsymbol{x}_{it_1}^\top\boldsymbol{x}_{it_1}=O(n),$$

and

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\Big|\Big|\sum_{i=1}^n A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\{m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g^0)-m(\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g)\}\boldsymbol{x}_{it_1}\Big|\Big|^2\\
=&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sum_{i=1}^n\Big|\Big|A_{it_1}^{1/2}(\boldsymbol{\beta}_g)A_{it_2}^{-1/2}(\boldsymbol{\beta}_g)\dot{m}(\{\boldsymbol{x}_{it_2}^\top\boldsymbol{\beta}_g\}^*)\boldsymbol{x}_{it_2}^\top(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)\boldsymbol{x}_{it_1}\Big|\Big|^2\\
\lesssim&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sum_{i=1}^n(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)^\top\boldsymbol{x}_{it_2}\boldsymbol{x}_{it_2}^\top(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)\boldsymbol{x}_{it_1}^\top\boldsymbol{x}_{it_1}\\
=&C^2\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau O_p(n).
\end{aligned}$$

Similar to the proof of Lemma 2, we have $\max_{1\leqslant k,l\leqslant T}\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{kl}=O_p(n^{-1/2})$ from Assumption (A9) (ii). Then, we have $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}||\boldsymbol{S}_g^*(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)||=O_p(T^2)$, which proves the lemma.

The following Lemma is from Remark 1 in Xie and Yang (2003).

Lemma 4: *It holds that, for all $i=1,\ldots,n$,*

$$\overline{\mathscr{D}}_i(\boldsymbol{\beta}_g)=\overline{\boldsymbol{H}}_i(\boldsymbol{\beta}_g)+\overline{\boldsymbol{B}}_i(\boldsymbol{\beta}_g)+\overline{\boldsymbol{\mathcal{E}}}_i(\boldsymbol{\beta}_g),$$

for $\overline{\boldsymbol{B}}_i(\boldsymbol{\beta}_g) = \overline{\boldsymbol{B}}_i^{[1]}(\boldsymbol{\beta}_g) + \overline{\boldsymbol{B}}_i^{[2]}(\boldsymbol{\beta}_g)$ *and* $\overline{\boldsymbol{\mathcal{E}}}_i(\boldsymbol{\beta}_g) = \overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_g) + \overline{\boldsymbol{\mathcal{E}}}_i^{[2]}(\boldsymbol{\beta}_g)$, *where*

$$\overline{\boldsymbol{B}}_i^{[1]}(\boldsymbol{\beta}_g) = \boldsymbol{X}_i^\top \mathrm{diag}[\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i,$$

$$\overline{\boldsymbol{B}}_i^{[2]}(\boldsymbol{\beta}_{g_i}) = \boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}\mathrm{diag}[m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i,$$

$$\overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_g) = \boldsymbol{X}_i^\top \mathrm{diag}[\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{\varepsilon}_i]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i,$$

and

$$\overline{\boldsymbol{\mathcal{E}}}_i^{[2]}(\boldsymbol{\beta}_{g_i}) = \boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\overline{\boldsymbol{R}}^{-1}\mathrm{diag}[\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{\varepsilon}_i]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i.$$

Here, $\boldsymbol{G}_i^{[\ell]}(\boldsymbol{\beta}_g) = \mathrm{diag}({q'}_{it}^{[\ell]}(\boldsymbol{\beta}_g), \dots, {q'}_{it}^{[\ell]}(\boldsymbol{\beta}_g)$, *for* $\ell = 1, 2$, *where*

$$q_{it}^{[1]}(\boldsymbol{\beta}_g) = [a''(\theta_{it}]^{-1/2}m'(\eta_{it}), \quad q_{it}^{[2]}(\boldsymbol{\beta}_g) = [a''(\theta_{it})]^{-1/2},$$

and

$${q'}_{it}^{[1]}(\boldsymbol{\beta}_g) = -\frac{1}{2}\frac{a^{(3)}(\theta_{it})}{[a''(\theta_{it})]^{5/2}}\{m'(\eta_{it})\}^2 + \frac{m''(\eta_{it})}{[a''(\theta_{it})]^{1/2}}, \quad {q'}_{it}^{[2]}(\boldsymbol{\beta}_{g_i}) = -\frac{1}{2}\frac{a^{(3)}(\theta_{it})}{[a''(\theta_{it})]^{5/2}}m'(\eta_{it}).$$

LEMMA 5: *Suppose the Assumptions (A1)-(A9). It holds that, for any* $\boldsymbol{\lambda} \in \mathbb{R}^p$ *and* $g = 1, \dots, G$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top[\mathscr{D}_g^*(\boldsymbol{\beta}_g) - \overline{\mathscr{D}}_g^*(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee n^{-1/2}\}T^2 n).$$

Proof. By Lemma 4, it is sufficient to prove the following three results:

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top[\boldsymbol{H}_g^*(\boldsymbol{\beta}_g) - \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee n^{-1/2}\}T^2 n),$$

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^*(\boldsymbol{\beta}_g) - \overline{\boldsymbol{B}}_g^*(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee n^{-1/2}\}T^2 n),$$

and

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top[\overline{\boldsymbol{\mathcal{E}}}_g^*(\boldsymbol{\beta}_g) - \overline{\boldsymbol{\mathcal{E}}}_g^*(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee n^{-1/2}\}T^2 n).$$

We have

$$\begin{aligned}|\boldsymbol{\lambda}^\top[\boldsymbol{H}_g^*(\boldsymbol{\beta}_g)-\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| =& \Big|\sum_{i:g_i^0=g}\boldsymbol{\lambda}^\top\boldsymbol{X}_i^\top\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g^0)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g^0)\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\{\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\\ &\times\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g^0)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g^0)\boldsymbol{X}_i\boldsymbol{\lambda}\Big|\\ \lesssim& ||\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})||_F\lambda_{\max}\Big(\sum_{i:g_i^0=g}\boldsymbol{X}_i^\top\boldsymbol{X}_i\Big),\end{aligned}$$

which implies that $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{H}_g^*(\boldsymbol{\beta}_g)-\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n)$ from Assumptions (A2) (i), (A7) and (A9) (ii). Next, we will verify

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[1]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{B}}_g^{[1]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n),$$

and

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[2]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{B}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n).$$

We have from Cauchy-Schwarz inequality

$$\begin{aligned}&|\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[1]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{B}}_g^{[1]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|\\ =&\Big|\sum_{i:g_i^0=g}\boldsymbol{\lambda}^\top\boldsymbol{X}_i^\top\mathrm{diag}[\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\\ &\quad\times\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)-m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}\Big|\\ =&\Big|\sum_{i:g_i^0=k}\boldsymbol{\lambda}^\top\boldsymbol{X}_i^\top\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{\lambda}]\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\\ &\quad\times\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)-m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}\Big|\\ \leqslant&\sum_{i:g_i^0=g}||\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{\lambda}]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}||\\ &\quad\times||\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0)-m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}||.\end{aligned}$$

We have

$$\boldsymbol{\lambda}^\top\boldsymbol{X}_i^\top\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\mathrm{diag}^2[\boldsymbol{X}_i\boldsymbol{\lambda}]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}\leqslant\max_{1\leqslant t\leqslant T}|\boldsymbol{x}_{it}^\top\boldsymbol{\lambda}|^2\max_{1\leqslant t\leqslant T}|q'^{[1]}_{it}(\boldsymbol{\beta}_g)|^2\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i),$$

and, by using (1), we have for $\boldsymbol{\beta}_g^*$ between $\boldsymbol{\beta}_g^0$ and $\boldsymbol{\beta}_g$,

$$
\begin{aligned}
&\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}^\top \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}^2 \\
&\quad \times \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\} \\
=&(\boldsymbol{\beta}_g^0 - \boldsymbol{\beta}_g)^\top \boldsymbol{X}_i^\top \boldsymbol{\Delta}(\boldsymbol{\beta}_g^*)\boldsymbol{A}_i(\boldsymbol{\beta}_g^*)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)[\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\{\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\} \\
&\quad \times \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})]^2 \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\boldsymbol{A}_i(\boldsymbol{\beta}_g^*)\boldsymbol{\Delta}(\boldsymbol{\beta}_g^*)\boldsymbol{X}_i(\boldsymbol{\beta}_g^0 - \boldsymbol{\beta}_g) \\
\lesssim&||\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})||_F^2 \lambda_{\max}(\boldsymbol{X}_i^\top \boldsymbol{X}_i)\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)||.
\end{aligned}
$$

Then, from Assumptions (A7) and (A9) (ii), we have

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}_k\in\mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top[\boldsymbol{B}_{nk}^{[1]*}(\boldsymbol{\beta}_k) - \overline{\boldsymbol{B}}_{nk}^{[1]*}(\boldsymbol{\beta}_k)]\boldsymbol{\lambda}| \\
&\quad = nO_p(T^{1/2})O_p(\{T\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee Tn^{-1/2}\})O_p(T^{1/2})\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \\
&\quad = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee n^{-1/2}\}T^2 n)\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2},
\end{aligned}
$$

which proves $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[1]*}(\boldsymbol{\beta}_g) - \overline{\boldsymbol{B}}_g^{[1]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} \vee n^{-1/2}\}T^2 n)$ since $\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau \to 0$. Moreover, we have from Cauchy-Schwarz inequality

$$
\begin{aligned}
&|\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[2]*}(\boldsymbol{\beta}_g) - \overline{\boldsymbol{B}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| \\
&\quad = \Big|\sum_{i:g_i^0=g} \boldsymbol{\lambda}^\top \boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\gamma) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\} \\
&\qquad\qquad \times \mathrm{diag}[m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}\Big| \\
&\quad = \Big|\sum_{i:g_i^0=g} \boldsymbol{\lambda}^\top \boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\gamma) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g) \\
&\qquad\qquad \times \mathrm{diag}[\boldsymbol{X}_i\boldsymbol{\lambda}]\{m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)\}\Big| \\
&\quad \leqslant \sum_{i:g_i^0=g} ||\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{\lambda}]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\{\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\gamma) - \overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}|| \\
&\qquad\qquad \times ||m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)||.
\end{aligned}
$$

We have

$$\boldsymbol{\lambda}^\top \boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\mathrm{diag}^2[\boldsymbol{X}_i\boldsymbol{\lambda}]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)$$
$$\times \{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}$$
$$\lesssim \max_{1\leqslant t\leqslant T}|\boldsymbol{x}_{it}^\top\boldsymbol{\lambda}|^2 \max_{1\leqslant t\leqslant T}|q'^{[2]}_{it}(\boldsymbol{\beta}_g)|^2||\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})||_F^2\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i),$$

and for $\boldsymbol{\beta}_g^*$ between $\boldsymbol{\beta}_g$ and $\boldsymbol{\beta}_g^0$, we have

$$\begin{aligned}||m(\boldsymbol{X}_i\boldsymbol{\beta}_g^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_g)||^2 =&(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)^\top\boldsymbol{X}_i^\top\boldsymbol{A}_i(\boldsymbol{\beta}_g^*)\boldsymbol{\Delta}_i^2(\boldsymbol{\beta}_g^*)\boldsymbol{A}_i(\boldsymbol{\beta}_g^*)\boldsymbol{X}_i(\boldsymbol{\beta}_g^0-\boldsymbol{\beta}_g)\\ \lesssim&\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g-\boldsymbol{\beta}_g^0)||.\end{aligned}$$

Then, from Assumption (A7) and (A9) (ii) we have

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[2]*}(\boldsymbol{\beta}_g) - \overline{\boldsymbol{B}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|\\ &=nO_p(\{T\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee Tn^{-1/2}\})O_p(T^{1/2})O_p(T^{1/2})\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\\ &=O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n)\lambda_{\max}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2},\end{aligned}$$

which proves $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{B}_g^{[2]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{B}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n)$ since $\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau\to 0$. Lastly, we will verify

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[1]*}(\boldsymbol{\beta}_g) - \overline{\boldsymbol{\mathcal{E}}}_g^{[1]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n),$$

and

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[2]*}(\boldsymbol{\beta}_g) - \overline{\boldsymbol{\mathcal{E}}}_g^{[2]*}(\boldsymbol{\beta}_g)]\lambda| = O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n).$$

We have from Cauchy-Schwarz inequality

$$\begin{aligned}&|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[1]*}(\boldsymbol{\beta}_g) - \overline{\boldsymbol{\mathcal{E}}}_g^{[1]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|\\ &=\Big|\sum_{i:g_i^0=g}\boldsymbol{\lambda}^\top\boldsymbol{X}_i^\top\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{\lambda}]\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g^0)\boldsymbol{\varepsilon}_i\Big|\\ &\leqslant\sum_{i:g_i^0=g}||\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_g)\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{\lambda}]\boldsymbol{X}_i\boldsymbol{\lambda}||\cdot||\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g^0)\boldsymbol{\varepsilon}_i||\\ &\lesssim\sum_{i:g_i^0=g}\max_{1\leqslant j\leqslant T}\{||\boldsymbol{x}_{it}^\top\boldsymbol{\lambda}||\}\lambda_{\max}^{1/2}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)||\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma}) - \overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})||_F||\boldsymbol{\varepsilon}_i||.\end{aligned}$$

Then, from Assumption (A7) and (A9) (ii) we have we have

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_{nk}^{[1]}(\boldsymbol{\beta}_k)-\overline{\boldsymbol{\mathcal{E}}}^{[1]*}(\boldsymbol{\beta}_k)]\boldsymbol{\lambda}|\\&=nO_p(T^{1/2})O_p(\{T\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee Tn^{-1/2}\})O_p(T^{1/2})\\&=O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n),\end{aligned}$$

which proves $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[1]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{\mathcal{E}}}_g^{[1]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|=O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n)$.

Moreover, we have from Cauchy-Schwarz inequality

$$\begin{aligned}&|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[2]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{\mathcal{E}}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|\\&=\Big|\sum_{i:g_i^0=g}\boldsymbol{\lambda}^\top\boldsymbol{X}_i^\top\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\mathrm{diag}[\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{\varepsilon}_i]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}\Big|\\&\leqslant\Big(\sum_{i:g_i^0=g}||\{\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}||^2\Big)^{1/2}\\&\quad\times\Big(\sum_{i:g_i^0=g}||\mathrm{diag}[\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{\varepsilon}_i]\boldsymbol{G}_i^{[2]}(\boldsymbol{\beta}_g)\boldsymbol{X}_i\boldsymbol{\lambda}||^2\Big)^{1/2}\\&\lesssim||\widehat{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})||_F\max_{1\leqslant j\leqslant T}\{|A_{it}^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\varepsilon_{it}|\}\lambda_{\max}\Big(\sum_{i:g_i^0=g}\boldsymbol{X}_i^\top\boldsymbol{X}_i\Big).\end{aligned}$$

Then, from Assumption (A7) and (A9) (ii) we have

$$\begin{aligned}&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[2]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{\mathcal{E}}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|\\&=O_p(\{T\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee Tn^{-1/2}\})O_p(nT)\\&=O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n),\end{aligned}$$

which proves $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\sup_{||\boldsymbol{\lambda}||=1}|\boldsymbol{\lambda}^\top[\boldsymbol{\mathcal{E}}_g^{[2]*}(\boldsymbol{\beta}_g)-\overline{\boldsymbol{\mathcal{E}}}_g^{[2]*}(\boldsymbol{\beta}_g)]\boldsymbol{\lambda}|=O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2n)$.

The following three lemmas are from Lemma A.1. (ii), Lemma A.2. (ii), Lemma A.3. (ii) in Xie and Yang (2003), respectively. These three lemmas are hold under the assumption (AH) in Xie and Yang (2003), which is satisfied in our problem from Assumptions (A1).

Lemma 6: *Suppose Assumption (A1) and (A8) (i) hold. It holds that, for any* $\boldsymbol{\lambda}\in\mathbb{R}^p$

and $g = 1, \ldots, G$,

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \boldsymbol{\lambda} - 1| = o_p(1).$$

LEMMA 7: *Suppose Assumptions (A1) and (A8) (i) hold. It holds that, for any $\boldsymbol{\lambda} \in \mathbb{R}^p$ and $g = 1, \ldots, G$,*

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{B}}_g^*(\boldsymbol{\beta}_g) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \boldsymbol{\lambda}| = o_p(1).$$

LEMMA 8: *Suppose Assumptions (A1) and (A8) (ii) hold. It holds that, for any $\boldsymbol{\lambda} \in \mathbb{R}^p$ and $g = 1, \ldots, G$,*

$$\sup_{\boldsymbol{\beta}_g \in \mathcal{B}_{nT}} \sup_{||\boldsymbol{\lambda}||=1} |\boldsymbol{\lambda}^\top \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{\mathcal{E}}}_g^*(\boldsymbol{\beta}_g) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \boldsymbol{\lambda}| = o_p(1).$$

The proof is based on that of Theorem 3.6 in Wang (2011). We will verify the following condition: for any $\epsilon > 0$, there exists a constant $C > 0$ such that for all n and T sufficiently large,

$$P\Big(\sup_{\boldsymbol{\beta} \in \mathcal{B}_{nT}, \boldsymbol{\gamma} \in \Gamma} (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \boldsymbol{S}_g(\boldsymbol{\beta}_g) < 0\Big) \geqslant 1 - \epsilon,$$

where $\mathcal{B}_{nT} = \{\boldsymbol{\beta} : \max_{g=1,\ldots,G} ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)|| = C\tau^{1/2}\}$ and $\Gamma = \{\boldsymbol{\gamma} = (g_1, \ldots, g_n) : n^{-1}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\} = o_p(T^{-\delta}) \ \text{ for all } \ \delta > 0\}$. This is a sufficient condition to ensure the existence of a sequence of roots $\widehat{\boldsymbol{\beta}}_g$ of the equation $\boldsymbol{S}_g(\boldsymbol{\beta}_g) = 0$ for $g = 1, \ldots, G$ such that $\widehat{\boldsymbol{\beta}} \in \mathcal{B}_{nT}$ for $\boldsymbol{\gamma} \in \Gamma$. This is because from Assumption (A5) and (A7), we can estimate each $\boldsymbol{\beta}_i$ consistently by solving $\boldsymbol{S}_i(\boldsymbol{\beta}_i) = \mathbf{0}$, and then, $P(\boldsymbol{\gamma} \notin \Gamma) = o_p(1)$ from Lemma 1.

From Taylor expansion, we can write

$$\begin{aligned}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \boldsymbol{S}_g(\boldsymbol{\beta}_g) =& (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \boldsymbol{S}_g(\boldsymbol{\beta}_g^0) - (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n \mathbf{1}\{g_i = g\} \mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_g^0) \\ \equiv& I_1 + I_2,\end{aligned}$$

where $\boldsymbol{\beta}_{g_i}^*$ lies between $\boldsymbol{\beta}_{g_i}$ and $\boldsymbol{\beta}_g^0$ for $i = 1, \ldots, n$. Next, we write

$$I_1 = (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0) + (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \{\boldsymbol{S}_g(\boldsymbol{\beta}_g^0) - \overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0)\} \equiv I_{11} + I_{12}.$$

For $\ell = 1, \ldots, p$, denote $\boldsymbol{e}_\ell \in \mathbb{R}^p$ with ℓth element equal to 1 and the others equal to 0. Then, we have

$$
\begin{aligned}
&E[\{\boldsymbol{e}_\ell^\top \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0)\}^2] \\
=&\boldsymbol{e}_\ell^\top \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\sum_{i=1}^n \mathbf{1}\{g_i^0 = g\}\boldsymbol{X}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{\beta}_g^0)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g^0)\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}^0)\boldsymbol{R}^0\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}^0) \\
&\quad \times \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_g^0)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_g^0)\boldsymbol{X}_i\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\boldsymbol{e}_\ell \\
\leqslant&\lambda_{\max}(\boldsymbol{R}^0\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}^0,\boldsymbol{\gamma}^0)).
\end{aligned}
$$

Thus, we can bound $|I_{11}|$ by

$$
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} |I_{11}| \leqslant ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)|| \cdot ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0)|| \leqslant C\tau.
$$

From the Lemma 2 and 3, we have

$$
\begin{aligned}
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} |I_{12}| \leqslant&||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)|| \cdot ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\{\boldsymbol{S}_g(\boldsymbol{\beta}_g^0) - \overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0)\}|| \\
\leqslant&\tau^{1/2}O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)T^2).
\end{aligned}
$$

Since $\tau^{-1/2}\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)T^2 \to 0$ from Assumption (A3), $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} |I_{12}| = o_p(\tau)$. Hence, we have $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} |I_1| \leqslant C\tau$. In what follows, we will evaluate I_2. It can be written as

$$
\begin{aligned}
I_2 =& -(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n \mathbf{1}\{g_i = g\}\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_g^0) \\
&-(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n \mathbf{1}\{g_i = g\}\{\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*) - \overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)\}(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_g^0) \\
\equiv& I_{21} + I_{22}.
\end{aligned}
$$

For $g_i^0 = g_i = g$, $\boldsymbol{\beta}_{g_i}^*$ lies between $\boldsymbol{\beta}_g$ and $\boldsymbol{\beta}_g^0$, and then we write $\boldsymbol{\beta}_{g_i}^* \equiv \boldsymbol{\beta}_g^*$ for such i. Hence, we can write

$$
\begin{aligned}
I_{21} =& -(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \overline{\mathscr{D}}_g^*(\boldsymbol{\beta}_g^*)(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0) \\
&-(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n (\mathbf{1}\{g_i = g\} - \mathbf{1}\{g_i^0 = g\})\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_g^0) \\
\equiv& I_{211} + I_{212}.
\end{aligned}
$$

For I_{211}, we write

$$
\begin{aligned}
I_{211} =& -(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^*)(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0) - (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \{\overline{\mathscr{D}}_g^*(\boldsymbol{\beta}_g^*) - \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^*)\}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0) \\
\equiv& I_{2111} + I_{2112}.
\end{aligned}
$$

For I_{2111}, we can write

$$
\begin{aligned}
I_{2111} =& -(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0) \\
& - (\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2} \Big[\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^*) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} - \boldsymbol{I}_p\Big] \\
& \quad \times \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0) \\
\equiv& I_{21111} + I_{21112}.
\end{aligned}
$$

For $\boldsymbol{\beta} \in \mathcal{B}_{nT}$, we have $I_{21111} = -C^2\tau$. Moreover, for $g_i^0 = g_i = g$, $\boldsymbol{\beta}_{g_i}^* \equiv \boldsymbol{\beta}_g^*$ is contained in a local neighborhood of $\boldsymbol{\beta}_g^0$. Then, for I_{21112}, we have from Lemma 6,

$$
\begin{aligned}
|I_{21112}| \leqslant& \sup_{\boldsymbol{\beta} \in \mathcal{B}_{nT}} \max \Big\{ \Big| \lambda_{\min}\Big(\Big[\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^*) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} - \boldsymbol{I}_p \Big] \Big) \Big|, \\
& \qquad \Big| \lambda_{\max}\Big(\Big[\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^*) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} - \boldsymbol{I}_p \Big] \Big) \Big| \Big\} \\
& \times ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)||^2 \\
=& o(1) C^2 \tau,
\end{aligned}
$$

which is dominated by I_{21111}. Hence, for $\boldsymbol{\beta} \in \mathcal{B}_{nT}$ we have $I_{2111} = -C^2\tau$. Next, we verify I_{2112}. For $g_i^0 = g_i = g$, we have from Lemma 4, 7 and 8

$$
\begin{aligned}
|I_{2112}| =& |(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \{\overline{\boldsymbol{B}}_g^*(\boldsymbol{\beta}_g^*) + \overline{\boldsymbol{\mathcal{E}}}_g^*(\boldsymbol{\beta}_g^*)\}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)| \\
\leqslant& \sup_{\boldsymbol{\beta} \in \mathcal{B}_{nT}} \{\lambda_{\max}(\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{B}}_g^*(\boldsymbol{\beta}_g^*) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}) \\
& \quad + \lambda_{\max}(\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} \overline{\boldsymbol{\mathcal{E}}}_g^*(\boldsymbol{\beta}_g^*) \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2})\} ||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)||^2 \\
=& o(1) C^2 \tau,
\end{aligned}
$$

which is dominated by I_{2111}. Hence, for $\boldsymbol{\beta} \in \mathcal{B}_{nT}$ we have $I_{211} = -C^2\tau$. Next, we verify I_{212}.

$$\begin{aligned}
|I_{212}| \leqslant & \Big|(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n (\mathbf{1}\{g_i = g\} - \mathbf{1}\{g_i^0 = g\})\overline{\boldsymbol{H}}_i(\boldsymbol{\beta}_{g_i}^*)(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_{g_i^0}^0)\Big| \\
& + \Big|(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n (\mathbf{1}\{g_i = g\} - \mathbf{1}\{g_i^0 = g\})\overline{\boldsymbol{B}}_i(\boldsymbol{\beta}_{g_i}^*)(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_{g_i^0}^0)\Big| \\
& + \Big|(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n (\mathbf{1}\{g_i = g\} - \mathbf{1}\{g_i^0 = g\})\overline{\boldsymbol{\mathcal{E}}}_i(\boldsymbol{\beta}_{g_i}^*)(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}_{g_i^0}^0)\Big| \\
\equiv & I_{2121} + I_{2122} + I_{2123}.
\end{aligned}$$

From Cauchy-Schwarz inequality, $\boldsymbol{\beta} \in \mathcal{B}_{nT}$ we have

$$\begin{aligned}
|I_{2121}| \lesssim & \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)|| \sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\} n\{\max_{1\leqslant i\leqslant n} \sup_{\boldsymbol{\beta}\in\mathcal{B}} \lambda_{\max}(\overline{\boldsymbol{H}}_i(\boldsymbol{\beta}_{g_i}))\} \\
\lesssim & C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} n\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\}\Big) n\{\max_{1\leqslant i\leqslant n} \sup_{\boldsymbol{\beta}\in\mathcal{B}} \lambda_{\max}(\overline{\boldsymbol{H}}_i(\boldsymbol{\beta}_{g_i}))\}.
\end{aligned}$$

From Assupmtions (A1) and (A6), for $i = 1, \ldots, n$ we have

$$\max_{\boldsymbol{\beta}\in\mathcal{B}}\{\lambda_{\max}(\overline{\boldsymbol{H}}_i(\boldsymbol{\beta}_{g_i}))\} \lesssim \max_{\boldsymbol{\beta}\in\mathcal{B}} \max_{t=1,\ldots,T}[a''(\theta_{it}(\boldsymbol{\beta}_{g_i}))\{u'(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{g_i})\}^2]\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i) = O_p(T),$$

for $\boldsymbol{\beta}_{g_i}^*$ between $\boldsymbol{\beta}_{g_i^0}^0$ and $\boldsymbol{\beta}_{g_i}$, which implies that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma} |I_{2121}| = C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}n^2 T o_p(T^{-\delta}) = o_p(\tau).$$

Similarly, Cauchy-Schwarz inequality we have

$$\begin{aligned}
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma} |I_{2122}| \lesssim & C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2} n\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\}\Big) \\
& \times n[\{\max_{1\leqslant i\leqslant n} \sup_{\boldsymbol{\beta}\in\mathcal{B}} ||\overline{\boldsymbol{B}}_i^{[1]}(\boldsymbol{\beta}_{g_i})||_F\} + \{\max_{1\leqslant i\leqslant n} \sup_{\boldsymbol{\beta}\in\mathcal{B}} ||\overline{\boldsymbol{B}}_i^{[2]}(\boldsymbol{\beta}_{g_i})||_F\}].
\end{aligned}$$

It is noted that we have from Cauchy-Schwarz inequality

$$
\begin{aligned}
&\{\overline{\boldsymbol{B}}_i^{[1]}(\boldsymbol{\beta}_{g_i})\}_{jk}\\
=&\boldsymbol{e}_j^\top \boldsymbol{X}_i^\top \mathrm{diag}[\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_{g_i})\boldsymbol{X}_i\boldsymbol{e}_k\\
\leqslant&\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)\lambda_{\max}(\mathrm{diag}[\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}])\lambda_{\max}(\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_{g_i}))\\
=&\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)\max_{1\leqslant k\leqslant T}\Big\{\sum_{t=1}^T\{\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\}_{kj}A_{it}^{-1/2}(\boldsymbol{\beta}_{g_i})\{m(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_{g_i^0}) - m(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{g_i})\}\Big\}\\
&\times\lambda_{\max}(\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_{g_i}))\\
=&O_p(T^2).
\end{aligned}
$$

Similarly $\{\overline{\boldsymbol{B}}_i^{[2]}(\boldsymbol{\beta}_{g_i})\}_{jk} = O_p(T^2)$, then we have

$$
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma}|I_{2122}| = \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}n o_p(T^{-\delta})nT^{5/2} = o_p(\tau).
$$

Similarly, we have

$$
\begin{aligned}
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}|I_{2123}| \lesssim& C\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}n\Big(\frac{1}{n}\sum_{i=1}^n\mathbf{1}\{g_i\neq g_i^0\}\Big)\\
&\times n[\{\max_{1\leqslant i\leqslant n}\sup_{\boldsymbol{\beta}\in\mathcal{B}}||\overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_{g_i})||_F\} + \{\max_{1\leqslant i\leqslant n}\sup_{\boldsymbol{\beta}\in\mathcal{B}}||\overline{\boldsymbol{\mathcal{E}}}_i^{[2]}(\boldsymbol{\beta}_{g_i})||_F\}].
\end{aligned}
$$

It is noted that we have

$$
\begin{aligned}
&E[||\overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_{g_i})||_F^2]\\
=&\sum_{\ell=1}^T E[\boldsymbol{e}_\ell^\top\overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_{g_i})^\top\overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_{g_i})\boldsymbol{e}_\ell]\\
=&\sum_{\ell=1}^T E\Big[\boldsymbol{\varepsilon}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{e}_\ell]\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta}_{g_i})\boldsymbol{X}_i\\
&\qquad\times\boldsymbol{X}_i^\top\boldsymbol{G}_i^{[1]}(\boldsymbol{\beta})\mathrm{diag}[\boldsymbol{X}_i\boldsymbol{e}_\ell]\overline{\boldsymbol{R}}^{-1}(\boldsymbol{\beta},\boldsymbol{\gamma})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{\varepsilon}_i\Big]\\
\leqslant&\sum_{\ell=1}^T\lambda_{\max}(\boldsymbol{X}_i^\top\boldsymbol{X}_i)\max_{1\leqslant i\leqslant n,1\leqslant t\leqslant T}\max_{\boldsymbol{\beta}\in\mathcal{B}}|{q'}_{it}^{[1]}(\boldsymbol{\beta}_{g_i})|\max_{1\leqslant i\leqslant n,1\leqslant t\leqslant T}|\boldsymbol{x}_{it}^\top\boldsymbol{e}_\ell|^2\\
&\quad\times\max_{\boldsymbol{\beta}\in\mathcal{B}}\{\max_{1\leqslant i\leqslant n,1\leqslant t\leqslant T}A_{it}^{-1}(\boldsymbol{\beta}_{g_i})A_{it}(\boldsymbol{\beta}^0_{g_i^0})\}E[\boldsymbol{\varepsilon}_i^\top\boldsymbol{\varepsilon}_i]\\
=&O(T^3),
\end{aligned}
$$

which implies that $||\overline{\boldsymbol{\mathcal{E}}}_i^{[1]}(\boldsymbol{\beta}_{g_i})||_F = O_p(T^{3/2})$. Similarly $||\overline{\boldsymbol{\mathcal{E}}}_i^{[2]}(\boldsymbol{\beta}_{g_i})||_F = O_p(T^{3/2})$, then we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma} |I_{2123}| = \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^{*})\tau^{1/2} n o_p(T^{-\delta}) n T^{3/2} = o_p(\tau).$$

Thus, I_{2121}, I_{2122} and I_{2123} are dominated by I_{211} for $\boldsymbol{\beta}\in\mathcal{B}_{nT}$ and $\boldsymbol{\gamma}\in\Gamma$. Hence $I_{21} = -C^2\tau$ for $\boldsymbol{\beta}\in\mathcal{B}_{nT}$ and $\boldsymbol{\gamma}\in\Gamma$. Lastly, we verify I_{22}. We can write

$$\begin{aligned} I_{22} = &-(\boldsymbol{\beta}_g-\boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n \mathbf{1}\{g_i=g\}\mathbf{1}\{g_i=g_i^0\}\{\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)-\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)\}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_g^0)\\ &-(\boldsymbol{\beta}_g-\boldsymbol{\beta}_g^0)^\top \sum_{i=1}^n \mathbf{1}\{g_i=g\}\mathbf{1}\{g_i\neq g_i^0\}\{\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)-\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)\}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_g^0)\\ \equiv& I_{221}+I_{222}. \end{aligned}$$

For I_{221}, we can write, from Lemma 5,

$$\begin{aligned} |I_{221}| \leqslant& \sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} \max\{|\lambda_{\max}(\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)-\overline{\mathscr{D}}_g^*(\boldsymbol{\beta}_g))|,|\lambda_{\min}(\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)-\overline{\mathscr{D}}_g^*(\boldsymbol{\beta}_g))|\}\\ &\times \lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g-\boldsymbol{\beta}_g^0)||^2\\ =&O_p(\{\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}\vee n^{-1/2}\}T^2 n)C^2\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}_g^*)\tau. \end{aligned}$$

Since $\lambda_{\min}(\overline{\boldsymbol{H}}^*)$ is at least of order larger than $O_p(nT)$, and from definition, we have $\tau = \sup_{\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\gamma}}\lambda_{\max}(\{\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})\}^{-1}\boldsymbol{R}^0) \leqslant \sup_{\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\gamma}}\lambda_{\max}(\{\overline{\boldsymbol{R}}(\boldsymbol{\beta},\boldsymbol{\gamma})\}^{-1})\lambda_{\max}(\boldsymbol{R}^0) \leqslant O_p(T)$ form Assumption (A5), the order of $\tau\lambda_{\min}^{-2}(\overline{\boldsymbol{H}}^*)n^2$ is at most $O_p(T^{-1})$. Then, from Assumption (A3) we have $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}|I_{221}| = \tau o_p(1)$. As for I_{222}, we have

$$\begin{aligned} |I_{222}| \leqslant& \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)||\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\boldsymbol{\beta}_g-\boldsymbol{\beta}_g^0)||\\ &\times\sum_{i=1}^n \mathbf{1}\{g_i=g\}\mathbf{1}\{g_i\neq g_i^0\}\cdot||\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)-\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)||_F\cdot||\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_g^0||\\ \leqslant& \lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau n\Big(\sum_{i=1}^n\mathbf{1}\{g_i\neq g_i^0\}\Big)\sum_{i=1}^n||\mathscr{D}_i(\boldsymbol{\beta}_{g_i}^*)-\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i}^*)||_F\cdot||\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_g^0||. \end{aligned}$$

It is noted that he order of $||\mathscr{D}_i(\boldsymbol{\beta}_{g_i})-\overline{\mathscr{D}}_i(\boldsymbol{\beta}_{g_i})||_F$ is at most $O_p(T)$. Then, form Lemma 1, $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma}|I_{222}| = o_p(T^{-\delta})$, which implies that I_{22} is dominated by I_{21}. Thus, $(\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0)^\top\boldsymbol{S}_g(\boldsymbol{\beta}_g)$ on $\boldsymbol{\beta}\in\mathcal{B}_{nT}$ and $\boldsymbol{\gamma}\in\Gamma$ is asymptotically dominated in probability by $I_{11}+I_{21} = C\tau - C^2\tau$, which is negative for C large enough, which proves the first part of the Theorem.

Next, we show the second part of the theorem. We have

$$
\begin{aligned}
&P\Big(\max_{1\leqslant i\leqslant n}|\widehat{g}_i(\widehat{\boldsymbol{\beta}})-g_i^0|>0\Big)\\
&\quad\leqslant G\max_{1\leqslant g\leqslant G}P(\widehat{\boldsymbol{\beta}}_g\notin\mathcal{B}_{nT})+n\max_{1\leqslant i\leqslant n}P\Big(\widehat{\boldsymbol{\beta}}_g\in\mathcal{B}_{nT},\widehat{g}_i(\widehat{\boldsymbol{\beta}})\neq g_i^0\Big).
\end{aligned}
$$

The order of the first term is $o(1)$ from the first part of the Theorem. We have $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\mathbf{1}\{\widehat{g}_i(\boldsymbol{\beta})\neq g_i^0\}\leqslant\sum_{g=1}^{G}\widetilde{Z}_{ig}$. Then,

$$
\begin{aligned}
\max_{1\leqslant i\leqslant n}P\Big(\widehat{\boldsymbol{\beta}}_g\in\mathcal{B}_{nT},\widehat{g}_i(\widehat{\boldsymbol{\beta}})\neq g_i^0\Big)&=\max_{1\leqslant i\leqslant n}E[\mathbf{1}\{\widehat{\boldsymbol{\beta}}_g\in\mathcal{B}_{nT}\}\mathbf{1}\{\widehat{g}_i\neq g_i^0\}]\\
&\leqslant\max_{1\leqslant i\leqslant n}E\Big[\mathbf{1}\{\widehat{\boldsymbol{\beta}}_g\in\mathcal{B}_{nT}\}\sum_{g=1}^{G}\widetilde{Z}_{ig}\Big]\\
&\leqslant\max_{1\leqslant i\leqslant n}\sum_{g=1}^{G}P(\widetilde{Z}_{ig}=1)=o(T^{-\delta}),
\end{aligned}
$$

which proves the theorem.

S.3. Proof of Theorem 2

To show Theorem 2, we need to show the next lemmas.

Let $\widetilde{\boldsymbol{\beta}}_g$ denote a root of $\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)=0$. The next result shows that the grouped GEE estimator and the infeasible estimator with known population groups are asymptotically equivalent.

Lemma 9: *Suppose the Assumptions (A1)-(A9) hold. As n and T tend to infinity such that $n/T^{\nu}\to 0$ for some $\nu>0$, we have $\widehat{\boldsymbol{\beta}}_g=\widetilde{\boldsymbol{\beta}}_g+o_p(1)$ for $g=1,\dots,G$.*

Proof. We have

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma}||\boldsymbol{S}_g(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)||\\
&\quad\leqslant\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}||\boldsymbol{S}_g(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g(\boldsymbol{\beta}_g)||+\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma}||\overline{\boldsymbol{S}}_g(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)||.
\end{aligned}
$$

Then, we have $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT},\boldsymbol{\gamma}\in\Gamma}||\boldsymbol{S}_g(\boldsymbol{\beta}_g)-\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g)||=O_p(T^2)$ from Lemmas 2 and 3. Since $\widehat{\boldsymbol{\beta}}_g\in\mathcal{B}_{nT}$ for $\boldsymbol{\gamma}\in\Gamma$ from Theorem 1 and $\widetilde{\boldsymbol{\beta}}_g\in\mathcal{B}_{nT}$ from Theorem 2 in Xie and Yang (2003), this implies

$$\sup_{\gamma\in\Gamma}|(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)^\top\{\boldsymbol S_g(\widehat{\boldsymbol\beta}_g)-\overline{\boldsymbol S}^*_g(\widehat{\boldsymbol\beta}_g)\}|=|(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)^\top\overline{\boldsymbol S}^*_g(\widehat{\boldsymbol\beta}_g)|=O_p(T^2).$$

From Taylor expansion, for $\boldsymbol\beta_g^*$ between $\widehat{\boldsymbol\beta}_g$ and $\widetilde{\boldsymbol\beta}_g$ we have

$$\begin{aligned}\overline{\boldsymbol S}^*_g(\widehat{\boldsymbol\beta}_g)=&\overline{\boldsymbol S}^*_g(\widetilde{\boldsymbol\beta}_g)-\mathscr{D}^*_g(\boldsymbol\beta^*_g)(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)\\=&-\overline{\boldsymbol H}^*_g(\boldsymbol\beta^*_g)(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)-\{\overline{\mathscr{D}}^*_g(\boldsymbol\beta^*_g)-\overline{\boldsymbol H}^*_g(\boldsymbol\beta^*_g)\}(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g).\end{aligned}$$

Then, we have, from Lemmas 6 - 8,

$$|(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)^\top\{\boldsymbol S_g(\widehat{\boldsymbol\beta}_g)-\overline{\boldsymbol S}^*_g(\widehat{\boldsymbol\beta}_g)\}|=(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)^\top\overline{\boldsymbol H}^*_g(\boldsymbol\beta^*_g))(\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g)+o_p(1).$$

Hence, we have

$$\sup_{\gamma\in\Gamma}\inf_{\boldsymbol\beta\in\mathcal{B}_{nT}}\lambda_{\min}(\overline{\boldsymbol H}^*_g(\boldsymbol\beta))||\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g||^2\leqslant O_p(T^2)+o_p(1),$$

which implies $||\widehat{\boldsymbol\beta}_g-\widetilde{\boldsymbol\beta}_g||=o_p(1)$, since the order of $\lambda_{\min}(\overline{\boldsymbol H}^*_g(\boldsymbol\beta))$ is at least $O_p(nT)$. The Lemma follows from Lemma 1.

Next lemma is almost the same with Lemma 2 in Xie and Yang (2003).

LEMMA 10: *Suppose the Assumptions (A1)-(A9) hold. Moreover, suppose that, for all $g=1,\ldots,G$, there exists a constant ζ such that $(c^*T)^{1+\zeta}\gamma^*\to 0$ as $n\to\infty$. Moreover, suppose the marginal distribution of each observation has a density of the form from (2.1) in the main text. Then, when $n\to\infty$, we have*

$$\{\overline{\boldsymbol M}^*_g(\boldsymbol\beta^0_g)\}^{-1/2}\overline{\boldsymbol S}^*_g(\boldsymbol\beta^0_g)\to N(0,\boldsymbol I_p)\quad\textit{in distribution.}$$

Proof. For any $p\times 1$ vector $\boldsymbol\lambda$ such that $||\boldsymbol\lambda||=1$, let $\boldsymbol\lambda^\top\{\overline{\boldsymbol M}^*_g(\boldsymbol\beta^0_g)\}^{-1/2}\overline{\boldsymbol S}^*_g(\boldsymbol\beta^0_g)=\sum_{i:g_i^0=g}Z_{nTi}$, where $Z_{nTi}=\boldsymbol\lambda^\top\{\overline{\boldsymbol M}^*_g(\boldsymbol\beta^0_g)\}^{-1/2}\boldsymbol X_i^\top\boldsymbol\Delta_i(\boldsymbol\beta^0_g)\boldsymbol A_i^{1/2}(\boldsymbol\beta^0_g)\overline{\boldsymbol R}_i^{-1}(\boldsymbol\beta^0,\boldsymbol\gamma^0)\boldsymbol\varepsilon_i$. To establish the asymptotic normality, it suffices to check the Lindeberg condition for $\boldsymbol\lambda^\top\{\overline{\boldsymbol M}^*_g(\boldsymbol\beta^0_g)\}^{-1/2}\overline{\boldsymbol S}^*_g(\boldsymbol\beta^0_g)$, that is, for any $\epsilon>0$,

$$\sum_{i:g_i^0=g}E[Z^2_{nTi}\mathbf{1}\{|Z_{nTi}|>\epsilon\}]\to 0,$$

which is shown in the proof of Lemma 2 in Xie and Yang (2003).

We will show

$$\{\overline{\boldsymbol{M}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)(\widetilde{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g^0) \to N(0, \boldsymbol{I}_p) \quad \textit{in distribution.}$$

The theorem follows from Lemma 9.

For $\boldsymbol{\beta}_g^* \in \mathcal{B}_{nT}$ between $\widetilde{\boldsymbol{\beta}}_g$ and $\boldsymbol{\beta}_g^0$, from Theorem 1, we have

$$\begin{aligned}
&\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0) \\
&\quad = -\{\boldsymbol{H}_g(\widehat{\boldsymbol{\beta}}_g^0)\}^{1/2}(\widetilde{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g^0) + \Big[\{\boldsymbol{H}_g(\widehat{\boldsymbol{\beta}}_g)\}^{1/2} - \{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}\Big](\widetilde{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g^0) \\
&\quad\quad - \Big[\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\mathscr{D}}_g^*(\boldsymbol{\beta}_g^*)\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2} - \boldsymbol{I}_p\Big]\{\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)\}^{1/2}(\widetilde{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g^0).
\end{aligned}$$

From Lemmas 4 and 6 - 8, the second term in the right hand side of the above equation is $o_p(1)$, which implies that $\{\overline{\boldsymbol{M}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\boldsymbol{S}}_g^*(\boldsymbol{\beta}_g^0)$ and $\{\overline{\boldsymbol{M}}_g^*(\boldsymbol{\beta}_g^0)\}^{-1/2}\overline{\boldsymbol{H}}_g^*(\boldsymbol{\beta}_g^0)(\widetilde{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g^0)$ are asymptotically identically distributed. Hence, the theorem follows from Lemma 10.

S.4. Property of $\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma})$

In this section, we denote the estimated unstructured working correlation matrix as $\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{R}^*(\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\gamma}), \boldsymbol{\beta}, \boldsymbol{\gamma})$ for $\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ given in (2.4) in the main text. Then, it follows that

$$\begin{aligned}
\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_{g_i^0}^0)\boldsymbol{R}^0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_{g_i^0}^0)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}) \\
&+ \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0) - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^\top \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}).
\end{aligned}$$

The next lemma shows that $\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma})$ satisfies Assumption (A5) (ii).

LEMMA 11: *Suppose Assumptions (A1)-(A8) hold. It holds that* $\lambda_{\max}(\{\overline{\boldsymbol{R}}^{mo}(\boldsymbol{\beta}^0, \boldsymbol{\gamma})\}^{-2}\boldsymbol{R}^0) = O_p(1)$ *for any* $\boldsymbol{\gamma}$.

Proof. Since the eigenvalues of $\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}^0, \boldsymbol{\gamma})(\boldsymbol{R}^0)^{-1/2}$ and $(\boldsymbol{R}^0)^{-1/4}\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}^0, \boldsymbol{\gamma})(\boldsymbol{R}^0)^{-1/4}$ are the same, we will show that $\lambda_{\min}((\boldsymbol{R}^0)^{-1/4}\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}^0, \boldsymbol{\gamma})(\boldsymbol{R}^0)^{-1/4})$ is bounded away from zero. It can be written as

$$
\begin{aligned}
&\lambda_{\min}((\boldsymbol{R}^0)^{-1/4}\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})(\boldsymbol{R}^0)^{-1/4})\\
&\quad\geqslant\lambda_{\min}\Big((\boldsymbol{R}^0)^{-1/4}\frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{R}^0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})(\boldsymbol{R}^0)^{-1/4}\Big)\\
&\qquad+\lambda_{\min}\Big((\boldsymbol{R}^0)^{-1/2}\frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})-m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i})\}\\
&\qquad\qquad\times\{m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i^0})-m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i})\}^\top\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})(\boldsymbol{R}^0)^{-1/2}\Big).
\end{aligned}
$$

Since the smallest eigenvalue does not diverge to infinity, it is enough to show that the first term of the right-hand side of the above inequality is bounded away from zero. Then, we have

$$
\begin{aligned}
&\lambda_{\min}\Big((\boldsymbol{R}^0)^{-1/4}\frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{R}^0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})(\boldsymbol{R}^0)^{-1/2}\Big)\\
&\quad\geqslant\frac{1}{n}\sum_{i=1}^n\lambda_{\min}\Big((\boldsymbol{R}^0)^{-1/4}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{R}^0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})(\boldsymbol{R}^0)^{-1/4}\Big)\\
&\quad\geqslant\frac{1}{n}\sum_{i=1}^n\lambda^2_{\min}\Big((\boldsymbol{R}^0)^{-1/4}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}^0_{g_i^0})(\boldsymbol{R}^0)^{1/2}\Big)\\
&\quad\geqslant\frac{1}{n}\sum_{i=1}^n\min_{1\leqslant t\leqslant T}\{A_{it}^{-1}(\boldsymbol{\beta}^0_{g_i})A_{it}(\boldsymbol{\beta}^0_{g_i^0})\}\lambda^{1/4}_{\min}(\boldsymbol{R}^0)>0,
\end{aligned}
$$

where the last inequality follows from Assumption (A5) (i).

The next lemma shows that $\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma})$ satisfies Assumption (A9) (i).

LEMMA 12: *Under Assumptions (A1)-(A8), it holds that for any $\boldsymbol{\gamma}$,*

$$
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\max_{1\leqslant k,l\leqslant T}\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma})-\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\}_{kl}=O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2}).
$$

Proof. For any $\boldsymbol{\gamma}$, we can write

$$
\begin{aligned}
&\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^\top\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i})\\
&-\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i})\}^\top\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\\
=&\frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}) - \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\\
&\qquad\times\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^\top\{\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}) - \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\}\\
&+\frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}) - \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^\top\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\\
&+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^\top\{\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}) - \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\}\\
&+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\Big[\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^\top\\
&\qquad\qquad - \{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i})\}\{\boldsymbol{y}_i - m(\boldsymbol{X}_i\boldsymbol{\beta}^0_{g_i})\}^\top\Big]\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}^0_{g_i})\\
\equiv&\sum_{j=1}^{4} I_j.
\end{aligned}
$$

From Taylor expansion, for $\boldsymbol{\beta}^*_{g_i}$ between $\boldsymbol{\beta}_{g_i}$ and $\boldsymbol{\beta}^0_{g_i}$, we have

$$
\begin{aligned}
1 - A_{it}^{1/2}(\boldsymbol{\beta}_{g_i})A_{it}^{-1/2}(\boldsymbol{\beta}^0_{g_i}) =& 1 - \sqrt{\frac{a''(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{g_i})}{a''(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_{g_i})}}\\
=& -\frac{1}{2}\{a''(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^*_{g_i})a''(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^0_{g_i})\}^{-1/2}u'(\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^*_{g_i})\boldsymbol{x}_{it}^\top(\boldsymbol{\beta}_{g_i} - \boldsymbol{\beta}^0_{g_i}).
\end{aligned}
$$

Then, the (k,l)-element of I_1 can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\{A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i})-A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\}\{A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i})-A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\}\\
&\qquad\times\{y_{ik}-m(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i})\}\{y_{il}-m(\boldsymbol{x}_{il}^{\top}\boldsymbol{\beta}_{g_i})\}\\
=&\frac{1}{4n}\sum_{i=1}^{n}\{a''(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i}^{*})a''(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i}^{0})\}^{-1/2}u'(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i}^{*})\boldsymbol{x}_{ik}^{\top}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0})\\
&\qquad\quad\times\{a''(\boldsymbol{x}_{il}^{\top}\boldsymbol{\beta}_{g_i}^{*})a''(\boldsymbol{x}_{il}^{\top}\boldsymbol{\beta}_{g_i}^{0})\}^{-1/2}u'(\boldsymbol{x}_{il}^{\top}\boldsymbol{\beta}_{g_i}^{*})\boldsymbol{x}_{il}^{\top}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0})\\
&\qquad\quad\times A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i})\{y_{ik}-m(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i})\}\{y_{il}-m(\boldsymbol{x}_{il}\boldsymbol{\beta}_{g_i})\}A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i})\\
\lesssim&\Big(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0})^{\top}\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^{\top}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0})\{y_{ik}-m(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i})\}^2\Big)^{1/2}\\
&\quad\times\Big(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0})^{\top}\boldsymbol{x}_{il}\boldsymbol{x}_{il}^{\top}(\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0})\{y_{il}-m(\boldsymbol{x}_{il}^{\top}\boldsymbol{\beta}_{g_i})\}^2\Big)^{1/2}\\
\leqslant&\{\max_{1\leqslant t\leqslant T}\lambda_{\max}(\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\top})\}\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{\beta}_{g_i}-\boldsymbol{\beta}_{g_i}^{0}||^2\{y_{ik}-m(\boldsymbol{x}_{ik}^{\top}\boldsymbol{\beta}_{g_i})\}^2,
\end{aligned}$$

where the second last inequality follows from Cauchy-Schwarz inequality. Since we have for all $t=1,\ldots,T$, $\lambda_{\max}(\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\top})=O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n}\{y_{it}-m(\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_{g_i})\}^2=O_p(1)$, this implies that the order of $\{I_1\}_{k.l}$ is $O_p(\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^{*})\tau)$ for $\boldsymbol{\beta}\in\mathcal{B}_{nT}$. Similarly, the order of $\{I_2\}_{kl}$ and $\{I_3\}_{kl}$ are $O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^{*})\tau^{1/2})$ for $\boldsymbol{\beta}\in\mathcal{B}_{nT}$. For I_4, we can write

$$\begin{aligned}
I_4=&\frac{1}{n}\sum_{i=1}^{n}A_i^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^{0})-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^{0})\}^{\top}A_i^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\\
&+\frac{1}{n}\sum_{i=1}^{n}A_i^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^{0})\}\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^{0})-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^{\top}A_i^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\\
&+\frac{1}{n}\sum_{i=1}^{n}A_i^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^{0})-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^{0})-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i})\}^{\top}A_i^{-1/2}(\boldsymbol{\beta}_{g_i}^{0})\\
\equiv&\sum_{j=1}^{3}I_{4j}.
\end{aligned}$$

By using (1) for $\boldsymbol{\beta}_{g_i}^{*}$ between $\boldsymbol{\beta}_{g_i}$ and $\boldsymbol{\beta}_{g_i}^{0}$, the (k,l)-element of I_{41} can be written as from

Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i}^0)A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\{m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i}^0)-m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i})\}\{y_{il}-m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i}^0)\}\\
=&\frac{1}{n}\sum_{i=1}^{n} A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i}^0)A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\phi A_{ik}(\boldsymbol{\beta}_{g_i}^*)u'(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i}^*)\boldsymbol{x}_{ik}^\top(\boldsymbol{\beta}_{g_i}^0-\boldsymbol{\beta}_{g_i})\{y_{il}-m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i}^0)\}\\
\lesssim&\Big(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta}_{g_i}^0-\boldsymbol{\beta}_{g_i})^\top\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^\top(\boldsymbol{\beta}_{g_i}^0-\boldsymbol{\beta}_{g_i})\Big)^{1/2}\\
&\times\Big(\frac{1}{n}\sum_{i=1}^{n} A_{ik}^{-1}(\boldsymbol{\beta}_{g_i}^0)A_{il}^{-1}(\boldsymbol{\beta}_{g_i}^0)A_{ik}^2(\boldsymbol{\beta}_{g_i}^*)\{u'(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i}^*)\}^2\{y_{il}-m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i}^0)\}^\top\{y_{il}-m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i}^0)\}\Big)^{1/2},
\end{aligned}
$$

which implies that the order of $\{I_{41}\}_{kl}$ is $O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2})$ for $\boldsymbol{\beta}\in\mathcal{B}_{nT}$. Similarly, the order of $\{I_{42}\}_{kl}$ and $\{I_{43}\}_{kl}$ are $O_p(\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*)\tau^{1/2})$ and $O_p(\lambda_{\min}^{-1}(\overline{\boldsymbol{H}}^*)\tau)$, respectively for $\boldsymbol{\beta}\in\mathcal{B}_{nT}$, which proves the lemma.

The next lemma shows that $\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma})$ satisfies Assumption (A9) (ii).

LEMMA 13: *Under Assumptions (A1)-(A8), it holds that for any* $\boldsymbol{\gamma}$,

$$
\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}}\max_{1\leqslant k,l\leqslant T}|\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma})\}_{k.l}|=O_p(n^{-1/2}\vee\lambda_{\min}^{-1/2}(\overline{\boldsymbol{H}}^*(\boldsymbol{\beta}^0))\tau^{1/2}),
$$

Proof. From Lemma 12, it is enough to show that

$$
\max_{1\leqslant k,l\leqslant T}\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\}_{kl}=O_p(n^{-1/2}).
$$

We can write

$$
\begin{aligned}
&\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})-\overline{\boldsymbol{R}}^*(\boldsymbol{\beta}^0,\boldsymbol{\gamma})\\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\Big\{\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)\}\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)\}^\top-\boldsymbol{\Sigma}_i\Big\}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\\
&+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^0)\}\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^0)\}^\top\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\\
&+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\{\boldsymbol{y}_i-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^0)\}\{m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i^0}^0)-m(\boldsymbol{X}_i\boldsymbol{\beta}_{g_i}^0)\}^\top\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_{g_i}^0)\\
=&I_1+I_2+I_3.
\end{aligned}
$$

For $\sigma_{ikl} = \{\boldsymbol{\Sigma}_i\}_{kl}$, the (k,l)-element of I_1 can be written as

$$\{I_1\}_{kl} = \frac{1}{n}\sum_{i=1}^{n} A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i}^0) A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i}^0)[\{y_{ik} - m(\boldsymbol{x}_{ik}^\top \boldsymbol{\beta}_{g_i^0}^0)\}\{y_{il} - m(\boldsymbol{x}_{il}^\top \boldsymbol{\beta}_{g_i^0}^0)\} - \sigma_{ikl}].$$

Then, it is obvious $E[\{I_1\}_{kl}] = 0$ and

$$\begin{aligned}\mathrm{Var}(\{I_1\}_{kl}) =& \frac{1}{n^2}\sum_{i=1}^{n} A_{ik}^{-1}(\boldsymbol{\beta}_{g_i}^0) A_{il}^{-1}(\boldsymbol{\beta}_{g_i}^0) A_{ik}(\boldsymbol{\beta}_{g_i^0}^0) A_{il}(\boldsymbol{\beta}_{g_i^0}^0) \mathrm{Var}(\varepsilon_{ik}\varepsilon_{il}) \\ \leqslant& \frac{1}{n^2}\sum_{i=1}^{n} A_{ik}^{-1}(\boldsymbol{\beta}_{g_i}^0) A_{il}^{-1}(\boldsymbol{\beta}_{g_i}^0) A_{ik}(\boldsymbol{\beta}_{g_i^0}^0) A_{il}(\boldsymbol{\beta}_{g_i^0}^0)(E[\varepsilon_{ik}^4]E[\varepsilon_{il}^4])^{1/2} = O_p(1/n),\end{aligned}$$

where the last equality follows from Assumptions (A1) and (A4). Then, this implies that the order of the (k,l)-element of I_1 is $O_p(n^{-1/2})$. Similarly, both of the (k,l)-elements of I_2 and I_3 are $O_p(n^{-1/2})$, which implies the lemma.

The next lemma shows that $\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma})$ satisfies Assumption (A9) (iii).

LEMMA 14: *Under Assumptions (A1)-(A8), it holds that for any $\boldsymbol{\beta} \in \mathcal{B}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_{i*}$ whose only ith component differs from that of $\boldsymbol{\gamma}$,*

$$\max_{1\leqslant k,l\leqslant T} |\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i*}) - \widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma})\}_{kl}| = O_p(1/n).$$

Proof. The lemma immediately holds since we can write

$$\begin{aligned}&\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}_{i*}) - \widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma})\}_{kl} \\ &\quad= \frac{1}{n}\Big\{A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i^*}) A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i^*})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top \boldsymbol{\beta}_{g_i^*})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top \boldsymbol{\beta}_{g_i^*})\} \\ &\qquad - A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i^*}) A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i^*})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top \boldsymbol{\beta}_{g_i^*})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top \boldsymbol{\beta}_{g_i^*})\}\Big\},\end{aligned}$$

which is of order $O_p(1/n)$.

The next lemma shows that $\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma})$ satisfies Assumption (A9) (iv).

LEMMA 15: *Under Assumptions (A1)-(A8), it holds that for any $\boldsymbol{\beta} \in \mathcal{B}$, any $\boldsymbol{\gamma}$ satisfying $\sup_{\boldsymbol{\beta}\in\mathcal{B}_{nT}} n^{-1}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\} = o_p(T^{-\delta})$ and all $\delta > 0$,*

$$\max_{1\leqslant k,l\leqslant T} |\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}^0)\}_{kl}| = o_p(T^{-\delta}).$$

Proof. From Cauchy-Schwarz inequality, we can write

$$
\begin{aligned}
&\{\widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widehat{\boldsymbol{R}}^*(\boldsymbol{\beta},\boldsymbol{\gamma}^0)\}_{kl} \\
=&\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\}\Big\{A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i})A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i})\} \\
&\qquad - A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i^0})A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i^0})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i^0})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i^0})\}\Big\} \\
\leqslant&\Big(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{g_i \neq g_i^0\}\Big)^{1/2} \\
&\times\Big(\frac{1}{n}\sum_{i=1}^n \Big\{A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i})A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i})\} \\
&\qquad - A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i^0})A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i^0})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i^0})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i^0})\}\Big\}^2\Big)^{1/2}.
\end{aligned}
$$

Since we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \Big\{A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i})A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i})\} \\
&\qquad - A_{ik}^{-1/2}(\boldsymbol{\beta}_{g_i^0})A_{il}^{-1/2}(\boldsymbol{\beta}_{g_i^0})\{y_{ik} - m(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta}_{g_i^0})\}\{y_{il} - m(\boldsymbol{x}_{il}^\top\boldsymbol{\beta}_{g_i^0})\}\Big\}^2 = O_p(1),
\end{aligned}
$$

the lemma follows from Lemma 1.

S.5. Additional numerical results

S.5.1 *Details of competing methods in simulation studies*

We here provide details of competing methods used in the simulation study in Section 4.

- (RC; random coefficient model) Fit the following logistic random coefficient model:

$$
y_{it} \sim \mathrm{Ber}(p_{it}), \quad \mathrm{logit}(p_{it}) = \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_i, \quad \boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}_0, \boldsymbol{V}).
$$

  The model is fitted by using the R package "lme4" (Bates et al., 2016).

- (GMM; growth mixture model) Fit the following growth mixture model:

$$
f(y_{it}|\boldsymbol{x}_{it}) = \sum_{\ell=1}^L \pi_\ell \mathrm{Be}(y_{it}; \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_\ell), \quad \sum_{\ell=1}^L \pi_\ell = 1,
$$

  where $\mathrm{Be}(y_{it}; x_{it}^\top\boldsymbol{\beta}_\ell)$ denotes the Bernoulli distribution with success probability being $1/\{1+\exp(-\boldsymbol{x}_{it}^\top\boldsymbol{\beta}_\ell)\}$, and L is set to the same number of groups used in the GGEE method. The model parameters are estimated via an EM algorithm. The subject-specific estimates of coefficients are given by $\widehat{\boldsymbol{\beta}}_i = \sum_{\ell=1}^{L} \widehat{p}_{i\ell}\widehat{\boldsymbol{\beta}}_\ell$, where $\widehat{p}_{i\ell}$ is the posterior probability that the ith subject is classified to the ℓth group.

- (PWP; pair-wise penalization method) Consider the subject-wise logistic regression, $y_{it} \sim \mathrm{Ber}(p_{it})$ with $\mathrm{logit}(p_{it}) = \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_i$, and estimate $\boldsymbol{\beta}_i$ by maximizing the following objective function:

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\{y_{it}\log p_{it} + (1-y_{it})\log(1-p_{it})\} - \lambda \sum_{i\sim j}\sum_{k=1}^{p}|\beta_{ik}-\beta_{jk}|,$$

where $i \sim j$ denotes contingency between ith and jth subjects and λ is a tuning parameter. Based on the output of RC, we first computed the pair-wise difference of estimated regression coefficients and obtained a minimum spanning tree over n subjects. Then, pairs of connected subjects in the minimum spanning tree are regarded as "adjacent" in the above penalty term. The above objective function is easily optimized, and λ can be selected via cross-validation by using the R package "glmnet" (Friedman et al., 2010). This method can be regarded as an alternative and scalable version of the pair-wise penalization method by Zhu et al. (2018).

S.5.2 *Performance of confidence intervals*

We carry out simulation studies to investigate the performance of the Wald-type confidence intervals based on the estimated variance-covariance matrices using the form given in Theorem 2 (plug-in method) and the clustered bootstrap. We adopted the same data generating process used in the first simulation study in Section 4. We estimate variance-covariance matrices of $\boldsymbol{\beta}_g$ for $g = 1, 2, 3$, based on the plug-in and clustered bootstrap (with 100 bootstrap samples) methods, and then obtain Wald-type 95% confidence intervals, denoted by CI_{gk} for $k = 1, \ldots, p$. The performance of the intervals are evaluated by coverage probability (CP), $(pG)^{-1}\sum_{g=1}^{G}\sum_{k=1}^{p} \mathrm{I}(\beta_{gk} \in \mathrm{CI}_{gk})$, and average length (AL), $(pG)^{-1}\sum_{g=1}^{G}\sum_{k=1}^{p} |\mathrm{CI}_{gk}|$, which are averaged over 500 Monte Carlo replications. The results are shown in Table 1. It shows that the plug-in method tends to exhibits under-coverage probability when T is small. On the other hand, the bootstrap approach produces desirable confidence intervals with coverage probability close to the nominal level and longer interval lengths than those of the plug-in method.

[Table 1 about here.]

S.5.3 *Additional results in Section 5*

In Figure 1, we provided the CVA values for candidate values of G. It shows that the CVA value basically decreases from $G = 2$ and attains the minimum value at $G = 8$.

[Figure 1 about here.]

References

Bates, D., Machler, M., Bolker, B., and Walker, S. (2016). Fitting linear mixe-effects models using lme4. *Journal of Statistical Software* **67,** 1–48.

Bonhomme, S. and Manresa, E. (2015). Grouped pattern of heterogeneity in panel data. *Econometrica* **83,** 1147–1184.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* **33,** 1.

Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*, pages 1158–1176. Springer, Berlin.

Wang, L. (2011). Gee analysis of clustered binary data with diverging number of covariates. *The Annals of Statistics* **39,** 389–417.

Xie, M. and Yang, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *The Annals of Statistics* **31,** 310–347.

Zhu, X., Tang, X., and Qu, A. (2018). Longitudinal clustering for heterogeneous binary data. *Statistica Sinica* page to appear.

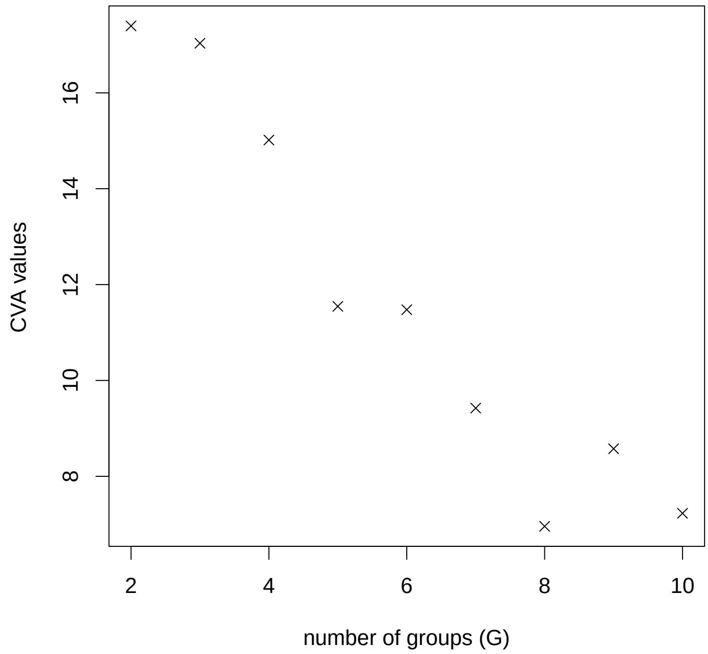

Figure 1. The CVA value for each G (the number of groups).

Table 1

Coverage probability (CP) and average length (AL) of 95% confidence intervals of group-specific parameters based on the plug-in and clustered bootstrap methods under exchangeable correlation (EX), first-order autoregressive (AR) and unstructured (US) working correlation matrices, averaged over 500 Monte Carlo replications.

		Plug-in			Bootstrap		
(n, T)		EX	AR	US	EX	AR	US
(180, 10)	CP	90.7	87.3	88.4	95.3	93.8	95.3
	AL	0.67	0.66	0.65	0.95	1.04	0.95
(180, 20)	CP	92.9	90.4	88.0	95.2	94.6	96.5
	AL	0.56	0.56	0.55	0.68	0.74	1.08
(270, 10)	CP	90.5	86.0	88.5	94.7	92.4	94.5
	AL	0.55	0.54	0.54	0.71	0.78	0.73
(270, 20)	CP	93.1	91.1	89.7	95.4	95.1	95.7
	AL	0.46	0.46	0.46	0.54	0.60	0.66