Supporting information for: "Functional Mixed Effects Model for Small Area Estimation"

Tapabrata Maiti[1], Samiran Sinha[2] and Ping-Shou Zhong[1]

[1] Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824
maiti@stt.msu.edu, pszhong@stt.msu.edu
and

[2] Department of Statistics, Texas A&M University, College Station, TX 77843
sinha@stat.tamu.edu

This document contains some lemmas and their proof that are key in proving Theorems 1 and 2 stated in the main paper. First, we re-state some notations that we have already introduced in the main paper. Model (3.3) in the paper is

$$Y = Z_F b_F + W\nu + \Upsilon,$$

where $W = (Z_R, M_0, \cdots, M_p)$ and $\nu = (b_R^T, U_0^T, U_1^T, \cdots, U_p^T)^T$. Denote $G = \text{var}(\nu) = \text{diag}(\text{cov}(b_R), I_n \otimes \Sigma_{u0}, \cdots, I_n \otimes \Sigma_{up})$ and $\text{cov}(Y) = \Sigma$. Also, $\Sigma_i = \Sigma_i(\delta) = Z_{Fi}\text{cov}(b_R)Z_{Fi}^T + \text{cov}(U_{i0}) + \sum_{k=1}^p \text{Diag}(X_{ik})\ \text{cov}(U_{ik})\text{Diag}(X_{ik}) + \Omega_i$ for $i = 1, \ldots, n$. The estimator of b_F is $\widehat{b}_F = (Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}Y$ and the predictor of ν is $\widehat{\nu} = GW^T\Sigma^{-1}(Y - Z_F\widehat{b}_F)$. The covariance matrix $\Sigma(\delta)$ is involved with parameters $\delta = (\sigma_{b_0}^2, \cdots, \sigma_{b_p}^2, \psi_0, \cdots, \psi_p, \rho_0, \cdots, \rho_p)^T$. Let $s = (p+1)(q+2)$ be the number of parameters in δ. These parameters are estimated through restricted maximum likelihood method by maximizing

$$\ell(\delta) = -\frac{1}{2}\log|Z_F^T\Sigma^{-1}Z_F| - \frac{1}{2}\sum_{i=1}^n \log|\Sigma_i| - \frac{1}{2}(Y - Z_F\widehat{b}_F)^T\Sigma^{-1}(Y - Z_F\widehat{b}_F). \tag{S1}$$

Then there exist some $\mathcal{T}$ such that $\mathcal{T}^T Z_F = 0$ and $\text{rank}(\mathcal{T}) = mn - L_1(p+1)$, and define

$$P = \mathcal{T}(\mathcal{T}^T\Sigma\mathcal{T})^{-1}\mathcal{T}^T = \Sigma^{-1} - \Sigma^{-1}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}$$

such that the likelihood can be written as

$$\ell(\delta) = -\frac{1}{2}\log|\mathcal{T}^T\Sigma\mathcal{T}| - \frac{1}{2}Y^TPY. \tag{S2}$$

and define the REML estimator $\widehat{\delta}$ as the solution to the score equation $\partial\ell(\delta)/\partial\delta = 0$.

Now we shall state the lemmas. Lemma 1 is needed for Lemma 2 which is needed for Theorem 2. Lemmas 3 and 4 are needed for proving Theorems 1 and 2.

Lemma 1 *Let* $\Sigma_i^* = \Sigma_i(\delta^*)$ *and* $\Sigma_i = \Sigma_i(\delta)$. *If* $\max_{i,k}\|X_{ik}\| < \infty$, *then* $\text{tr}\{(\Sigma_i^* - \Sigma_i)^2\} \leq C_\delta\|\delta^* - \delta\|^2$, *where* C_δ *is a constant and* $\|\delta^* - \delta\|^2 = (\delta^* - \delta)^T(\delta^* - \delta)$.

Proof: We write $\Sigma_i^* - \Sigma_i = J_1 + J_2 + J_3 + J_4 + J_5$, where

$$\begin{aligned}
J_1 &= Z_{F_i}\mathrm{Diag}\{({\sigma_{b_0}^*}^2 - \sigma_{b_0}^2)\mathrm{Diag}(\lambda_{L_1}^{-1}), \cdots, ({\sigma_{b_p}^*}^2 - \sigma_{b_p}^2)\mathrm{Diag}(\lambda_{L_1}^{-1})\} Z_{F_i}^T, \\
J_2 &= (\psi_0^* - \psi_0)A_m(\rho_0), \\
J_3 &= \psi_0^* A_m(\rho_0^*) - \psi_0^* A_m(\rho_0), \\
J_4 &= \sum_{k=1}^p \mathrm{Diag}(X_{ik})[\psi_k^*\{A_m(\rho_k^*) - A_m(\rho_k)\}]\mathrm{Diag}(X_{ik}), \\
J_5 &= \sum_{k=1}^p \mathrm{Diag}(X_{ik})\{(\psi_k^* - \psi_k)A_m(\rho_k)\}\mathrm{Diag}(X_{ik}).
\end{aligned}$$

Therefore,

$$\mathrm{tr}\{(\Sigma_i^* - \Sigma_i)^T(\Sigma_i^* - \Sigma_i)\} \leq C[\mathrm{tr}(J_1^T J_1) + \mathrm{tr}(J_2^T J_2) + \mathrm{tr}(J_3^T J_3) + \mathrm{tr}(J_4^T J_4) + \mathrm{tr}(J_5^T J_5)].$$

We see that

$$\mathrm{tr}(J_1^T J_1) \leq C \sum_{k=1}^p ({\sigma_{b_k}^*}^2 - \sigma_{b_k}^2)^2 \sum_{j,l=1}^m \{Z_{F_i}^{(k)^T}(t_j)\mathrm{Diag}(\lambda_{L_1}^{-2}) Z_{F_i}^{(k)}(t_l)\}^2.$$

Also, notice that $\mathrm{tr}(J_2^T J_2) = \mathrm{tr}\{(\psi_0^* - \psi_0)A_m(\rho_0)\}^2 \leq C(\psi_0^* - \psi_0)^2 \mathrm{tr}\{A_m(\rho_0)\}^2$ and

$$\begin{aligned}
\mathrm{tr}[\{\psi_0^* A_m(\rho_0^*) - \psi_0^* A_m(\rho_0)\}^2] &= \mathrm{tr}\left\{[(\psi_0^* - \psi_0)\{A_m(\rho_0^*) - A_m(\rho_0)\} + \psi_0\{A_m(\rho_0^*) - A_m(\rho_0)\}]^2\right\} \\
&\leq 2\{(\psi_0^* - \psi_0)^2 + \psi_0^2\}\mathrm{tr}\{A_m(\rho_0^*) - A_m(\rho_0)\}^2 \\
&\leq C\{(\psi_0^* - \psi_0)^2 + \psi_0^2\}\|\rho_0^* - \rho_0\|^2.
\end{aligned}$$

Hence, $\mathrm{tr}(J_3^T J_3) \leq C\|\delta^* - \delta\|^2$. Next,

$$\begin{aligned}
\mathrm{tr}(J_4^T J_4) &\leq C\sum_{k=1}^p \mathrm{tr}\big\{\mathrm{Diag}(X_{ik})[\psi_k^*\{A_m(\rho_k^*) - A_m(\rho_k)\}]\mathrm{Diag}(X_{ik})\big\}^2 \\
&\leq C\sum_{k=1}^p \Big[\mathrm{tr}\big\{\mathrm{Diag}(X_{ik})[(\psi_k^* - \psi_k)\{A_m(\rho_k^*) - A_m(\rho_k)\}]\mathrm{Diag}(X_{ik})\big\}^2 \\
&\quad + \mathrm{tr}\big\{\mathrm{Diag}(X_{ik})[\psi_k\{A_m(\rho_k^*) - A_m(\rho_k)\}]\mathrm{Diag}(X_{ik})\big\}^2\Big] \\
&\leq C\sum_{k=1}^p [(\psi_k^* - \psi_k)^2 \mathrm{tr}\{A_m(\rho_k^*) - A_m(\rho_k)\}^2 \mathrm{tr}\{\mathrm{Diag}(X_{ik})\}^4 \\
&\quad + \psi_k^2 \mathrm{tr}\{A_m(\rho_k^*) - A_m(\rho_k)\}^2 \mathrm{tr}\{\mathrm{Diag}(X_{ik})\}^4].
\end{aligned}$$

If $A_m(\rho_k)$ has bounded second derivatives with respect to ρ_k, then $\mathrm{tr}\{A_m(\rho_k^*) - A_m(\rho_k)\}^2 \leq C\|\rho_k^* - \rho_k\|$. And $\mathrm{tr}(J_5^T J_5) \leq C\sum_{k=1}^p (\psi_k^* - \psi_k)^2 \mathrm{tr}\big\{\mathrm{Diag}(X_{ik})A_m(\rho_k)\mathrm{Diag}(X_{ik})\big\}^2$. In summary, $\mathrm{tr}\{(\Sigma_i^* - \Sigma_i)^2\} \leq C_\delta\|\delta^* - \delta\|^2$.

Lemma 2 *Let C_δ be the constant defined in Lemma 1 and assume that the smallest eigenvalue of Σ_i is bounded below by $c_0 > 0$. Suppose that $\|\delta^* - \delta\|^2 \leq \Delta$ such that $2C_\delta \Delta / c_0^2 < 1$ when n is large enough. Then*

$$\mathrm{tr}\{(\Sigma_i^{*-1})^2\} \leq (1 - 2c_0^{-4} C_\delta^2 \Delta^2)^{-1} \Big[2\mathrm{tr}\{(\Sigma_i^{-1})^2\} + 2C_\delta \Delta / c_0^4 \Big].$$

Proof: By the matrix inverse formula, ${\Sigma^*}_i^{-1} = \Sigma_i^{-1} - \Sigma_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1} + \Sigma_i^{-1}(\Sigma_i^* - \Sigma_i){\Sigma^*}_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1}$, we then have

$$\begin{aligned}\mathrm{tr}\{(\Sigma_i^{*-1})^2\} \leq\ & 2\mathrm{tr}\{(\Sigma_i^{-1})^2\} + 2\mathrm{tr}[\{\Sigma_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1}\}^2] \\ & + 2\mathrm{tr}[\{\Sigma_i^{-1}(\Sigma_i^* - \Sigma_i){\Sigma^*}_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1}\}^2]. \qquad \text{(S3)}\end{aligned}$$

Because $(\Sigma_i^* - \Sigma_i)\Sigma_i^{-2}(\Sigma_i^* - \Sigma_i)$ is non-negative definite,

$$\begin{aligned}\mathrm{tr}[\{\Sigma_i^{-1}(\Sigma_i^* - \Sigma_i){\Sigma^*}_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1}\}^2] &\leq \mathrm{tr}^2\{\Sigma_i^{-1}(\Sigma_i^* - \Sigma_i){\Sigma^*}_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1}\} \\ &\leq \mathrm{tr}\{(\Sigma_i^{*-1})^2\}\mathrm{tr}[\{(\Sigma_i^* - \Sigma_i)\Sigma_i^{-2}(\Sigma_i^* - \Sigma_i)\}^2] \\ &\leq \mathrm{tr}^2\{(\Sigma_i^* - \Sigma_i)\Sigma_i^{-2}(\Sigma_i^* - \Sigma_i)\} \\ &\leq \mathrm{tr}\{(\Sigma_i^{*-1})^2\}\lambda_{\min}^{-4}(\Sigma_i)\mathrm{tr}^2\{(\Sigma_i^* - \Sigma_i)^2\} \\ &\leq \mathrm{tr}\{(\Sigma_i^{*-1})^2\}c_0^{-4}C_\delta^2\Delta^2,\end{aligned}$$

where the last inequality follows from Lemma 1 and the assumption in this Lemma. In addition, $\mathrm{tr}\{(\Sigma_i^{-1}(\Sigma_i^* - \Sigma_i)\Sigma_i^{-1})^2\} \leq \lambda_{\min}^{-4}(\Sigma_i)\mathrm{tr}\{(\Sigma_i^* - \Sigma_i)^2\}$. Hence, from (S3),

$$\mathrm{tr}\{(\Sigma_i^{*-1})^2\} \leq (1 - 2c_0^{-4}C_\delta^2\Delta^2)^{-1}(2\mathrm{tr}\{(\Sigma_i^{-1})^2\} + 2C_\delta\Delta/c_0^4).$$

This completes the proof of Lemma 2.

Lemma 3 *Let*

$$d_i^2 = \max_{j,k}\Big\{\mathrm{tr}(P\mathcal{V}_i P\mathcal{V}_i), \mathrm{tr}\Big(P\frac{\partial \mathcal{V}_i}{\partial \delta_j}P\frac{\partial \mathcal{V}_i}{\partial \delta_j}\Big), \mathrm{tr}\Big(P\frac{\partial^2 \mathcal{V}_i}{\partial \delta_j \partial \delta_k}P\frac{\partial^2 \mathcal{V}_i}{\partial \delta_j \partial \delta_k}\Big)\Big\}$$

and $d_* = \min_i d_i$. Then there exists $\widehat{\delta}$ such that for any $0 < q_0 < 1$ and large n,

$$\widehat{\delta} - \delta = -A^{-1}a + o_p(d_*^{-2q_0}),$$

where $a = \partial \ell(\delta)/\partial \delta$ and $A = E\{\partial^2 \ell(\delta)/\partial \delta^2\}$, on the set $\mathcal{B}$ with $P(\mathcal{B})$ converging to 1.

Proof: We will apply Theorem 2.1 of Das et al. (2004). Let us first verify the following conditions.

The gth moment of the following quantities are bounded for some d_i and $d_* = \min_i d_i$,

$$\frac{1}{d_i}\left|\frac{\partial\ell(\delta)}{\partial\delta}\Big|_{\delta_0}\right|,\quad \frac{1}{\sqrt{d_i d_j}}\left|\frac{\partial^2\ell(\delta)}{\partial\delta_i\delta_j}\Big|_{\delta_0} - E(\frac{\partial^2\ell(\delta)}{\partial\delta_i\delta_j}\Big|_{\delta_0})\right|,\quad \frac{d_*}{d_i d_j d_k}M_{ijk},$$

where $M_{ijk} = \sup_{\delta\in S_\delta(\delta_0)} |\partial^3\ell(\delta)/(\partial\delta_i\partial\delta_j\partial\delta_k)|$ with $S_\gamma(\delta_0) = \{\delta : |\delta_i - \delta_{0i}| \le \gamma d_*/d_i \ 1 \le i \le s\}$.

Using the likelihood given in (S2), we obtain the first derivative of $\ell(\delta)$ with respect to δ

$$\frac{\partial\ell(\delta)}{\partial\delta_i} = \frac{1}{2}\{\epsilon^T P\mathcal{V}_i P\epsilon - \mathrm{tr}(P\mathcal{V}_i)\}, \tag{S4}$$

where $\epsilon = Y - Z_F b_F$ and $\mathcal{V}_i = \mathrm{diag}(\mathcal{V}_{i1},\cdots,\mathcal{V}_{in})$. Let $\epsilon = \Sigma^{1/2}u$ and $u \sim N(0, I_{mn})$. Then for any $g \ge 2$,

$$\begin{aligned} E\Big|\frac{\partial\ell(\delta)}{\partial\delta_i}\Big|^g &= 2^{-g}E\Big|u^T\Sigma^{1/2}P\mathcal{V}_iP\Sigma^{1/2}u - E(u^T\Sigma^{1/2}P\mathcal{V}_iP\Sigma^{1/2}u)\Big|^g \\ &\le c\|\Sigma^{1/2}P\mathcal{V}_iP\Sigma^{1/2}\|_2^g = c\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP)^{g/2}. \end{aligned}$$

Thus, if we take $d_i = \mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP)^{1/2}$, the gth moment of $(1/d_i)\,|\partial\ell(\delta_0)/\partial\delta|$ is bounded. Because $\partial P/\partial\delta_j = -\mathcal{T}(\mathcal{T}^T\Sigma\mathcal{T})^{-1}\mathcal{T}^T(\partial\Sigma/\partial\delta_j)\mathcal{T}(\mathcal{T}^T\Sigma\mathcal{T})^{-1}\mathcal{T}^T = -P(\partial\Sigma/\partial\delta_j)P = -P\mathcal{V}_jP$, we have

$$\frac{\partial^2\ell(\delta)}{\partial\delta_i\partial\delta_j} = \frac{1}{2}\Big\{-\epsilon^TQ_{ij}\epsilon + \mathrm{tr}(P\mathcal{V}_jP\mathcal{V}_i) - \mathrm{tr}(P\frac{\partial\mathcal{V}_i}{\partial\delta_j})\Big\}, \tag{S5}$$

where $Q_{ij} = P\{\mathcal{V}_jP\mathcal{V}_i + \mathcal{V}_iP\mathcal{V}_j - (\partial\mathcal{V}_i/\partial\delta_j)\}P := PK_{ij}P$. Then we have

$$\begin{aligned} E\Big|\frac{\partial^2\ell(\delta)}{\partial\delta_i\partial\delta_j} - E(\frac{\partial^2\ell(\delta)}{\partial\delta_i\partial\delta_j})\Big|^g &= 2^{-g}E\Big|u^T\Sigma^{1/2}Q_{ij}\Sigma^{1/2}u - E(u^T\Sigma^{1/2}Q_{ij}\Sigma^{1/2}u)\Big|^g \\ &\le c\|\Sigma^{1/2}Q_{ij}\Sigma^{1/2}\|_2^g = c\mathrm{tr}(K_{ij}PK_{ij}P)^{g/2}, \end{aligned}$$

where

$$\begin{aligned} \mathrm{tr}(K_{ij}PK_{ij}P) &= \mathrm{tr}((\mathcal{V}_jP\mathcal{V}_i + \mathcal{V}_iP\mathcal{V}_j - \frac{\partial\mathcal{V}_i}{\partial\delta_j})P(\mathcal{V}_jP\mathcal{V}_i + \mathcal{V}_iP\mathcal{V}_j - \frac{\partial\mathcal{V}_i}{\partial\delta_j})P) \\ &= 2\mathrm{tr}(\mathcal{V}_jP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_iP) + 2\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_jP) \\ &\quad - 2\mathrm{tr}(\mathcal{V}_jP\mathcal{V}_iP\frac{\partial\mathcal{V}_i}{\partial\delta_j}P) - 2\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_jP\frac{\partial\mathcal{V}_i}{\partial\delta_j}P) + \mathrm{tr}(\frac{\partial\mathcal{V}_i}{\partial\delta_j}P\frac{\partial\mathcal{V}_i}{\partial\delta_j}P) \end{aligned}$$

and applying Lemma 5.2 of Das et al. (2004), we have

$$\begin{aligned} |\mathrm{tr}(\mathcal{V}_jP\mathcal{V}_iP\frac{\partial\mathcal{V}_i}{\partial\delta_j}P)| &\le \mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_jP)^{1/2}\mathrm{tr}(\frac{\partial\mathcal{V}_i}{\partial\delta_j}P\frac{\partial\mathcal{V}_i}{\partial\delta_j}P)^{1/2}; \\ |\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_jP\frac{\partial\mathcal{V}_i}{\partial\delta_j}P)| &\le \mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_jP)^{1/2}\mathrm{tr}(\frac{\partial\mathcal{V}_i}{\partial\delta_j}P\frac{\partial\mathcal{V}_i}{\partial\delta_j}P)^{1/2}; \\ |\mathrm{tr}(\mathcal{V}_jP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_iP)| &\le \mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_jP). \end{aligned}$$

Therefore,

$$\mathrm{tr}(K_{ij}PK_{ij}P) \le \big\{2\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_jP)^{1/2} + \mathrm{tr}(\frac{\partial \mathcal{V}_i}{\partial \delta_j}P\frac{\partial \mathcal{V}_i}{\partial \delta_j}P)^{1/2}\big\}^2.$$

Notice that $\mathrm{tr}(A^2) \le \mathrm{tr}^2(A)$ for any non-negative matrix A. Since $P^{1/2}\mathcal{V}_iP\mathcal{V}_iP^{1/2}$ is a non-negative definite matrix, we have

$$\begin{aligned}\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_jP\mathcal{V}_jP) &\le \mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_iP\mathcal{V}_iP)^{1/2}\mathrm{tr}(\mathcal{V}_jP\mathcal{V}_jP\mathcal{V}_jP\mathcal{V}_jP)^{1/2} \\ &\le \mathrm{tr}(P^{1/2}\mathcal{V}_iP\mathcal{V}_iP^{1/2})\mathrm{tr}(P^{1/2}\mathcal{V}_jP\mathcal{V}_jP^{1/2}) = \mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP)\mathrm{tr}(\mathcal{V}_jP\mathcal{V}_jP).\end{aligned}$$

Hence if we take $d_i = \max_j[\mathrm{tr}(\mathcal{V}_iP\mathcal{V}_iP)^{1/2}, \mathrm{tr}\{\partial\mathcal{V}_i/\partial\delta_j)P(\partial\mathcal{V}_i/\partial\delta_j)P\}^{1/2}]$ the gth moment of

$$\frac{1}{\sqrt{d_id_j}}\left|\frac{\partial^2\ell(\delta)}{\partial\delta_i\delta_j}\Big|_{\delta_0} - E(\frac{\partial^2\ell(\delta)}{\partial\delta_i\delta_j}\Big|_{\delta_0})\right|$$

is bounded for any $g \ge 2$.

Next, we compute the third derivatives,

$$\begin{aligned}&\frac{\partial^3\ell(\delta)}{\partial\delta_i\partial\delta_j\partial\delta_k} \\ &= -2^{-1}\epsilon^T\Big\{-P\mathcal{V}_kP(\mathcal{V}_jP\mathcal{V}_i + \mathcal{V}_iP\mathcal{V}_j - \frac{\partial\mathcal{V}_i}{\partial\delta_j})P \\ &\quad + P(\frac{\partial\mathcal{V}_j}{\partial\delta_k}P\mathcal{V}_i - \mathcal{V}_jP\mathcal{V}_kP\mathcal{V}_i - \mathcal{V}_jP\frac{\partial\mathcal{V}_i}{\partial\delta_k} + \frac{\partial\mathcal{V}_i}{\partial\delta_k}P\mathcal{V}_j - \mathcal{V}_iP\mathcal{V}_kP\mathcal{V}_j + \mathcal{V}_iP\frac{\partial\mathcal{V}_j}{\partial\delta_k} - \frac{\partial^2\mathcal{V}_i}{\partial\delta_j\partial\delta_k})P \\ &\quad - P(\mathcal{V}_jP\mathcal{V}_i + \mathcal{V}_iP\mathcal{V}_j + \frac{\partial\mathcal{V}_i}{\partial\delta_j})P\mathcal{V}_kP\Big\}\epsilon - E(\epsilon^TR_{ijk}\epsilon) \\ &= -\epsilon^T(P\mathcal{V}_kP\mathcal{V}_jP\mathcal{V}_iP + P\mathcal{V}_kP\mathcal{V}_iP\mathcal{V}_jP + P\mathcal{V}_iP\mathcal{V}_kP\mathcal{V}_jP)\epsilon + 2^{-1}\epsilon^TP\frac{\partial^2\mathcal{V}_i}{\partial\delta_j\partial\delta_k}P\epsilon \\ &\quad - \epsilon^T(P\mathcal{V}_kP\frac{\partial\mathcal{V}_i}{\partial\delta_j}P + P\mathcal{V}_iP\frac{\partial\mathcal{V}_j}{\partial\delta_k}P + P\mathcal{V}_jP\frac{\partial\mathcal{V}_i}{\partial\delta_k}P)\epsilon + E\{\epsilon^TP(R_{ijk} + R^*_{ijk} - 2^{-1}\frac{\partial^2\mathcal{V}_i}{\partial\delta_j\partial\delta_k})P\epsilon\},\end{aligned}$$

where $R_{ijk} = \mathcal{V}_kP\mathcal{V}_jP\mathcal{V}_i + \mathcal{V}_kP\mathcal{V}_iP\mathcal{V}_j + \mathcal{V}_iP\mathcal{V}_kP\mathcal{V}_j$ and $R^*_{ijk} = \mathcal{V}_kP(\partial\mathcal{V}_i/\partial\delta_j) + \mathcal{V}_iP(\partial\mathcal{V}_j/\partial\delta_k) + \mathcal{V}_jP(\partial\mathcal{V}_i/\partial\delta_k)$. Consider the first term in the third derivatives. Denote $\tilde{\Sigma}$ for Σ evaluated at $\tilde{\delta}$ and similarly for $\tilde{\mathcal{V}}_j$. Then it can be shown that

$$(\mathcal{T}^T\tilde{\Sigma}\mathcal{T})^{-1} = (\mathcal{T}^T\Sigma\mathcal{T})^{-1} + (\mathcal{T}^T\tilde{\Sigma}\mathcal{T})^{-1}\mathcal{T}^T(\Sigma - \tilde{\Sigma})\mathcal{T}(\mathcal{T}^T\Sigma\mathcal{T})^{-1}$$

and $\mathcal{T}^T\tilde{\mathcal{V}}_j\mathcal{T} = \mathcal{T}^T\mathcal{V}_j\mathcal{T} + \mathcal{T}^T(\tilde{\mathcal{V}}_j - \mathcal{V}_j)\mathcal{T}$. For convenience, denote $H = (\mathcal{T}^T\Sigma\mathcal{T})^{-1}$ and $G_i = \mathcal{T}^T\mathcal{V}_i\mathcal{T}$. Further $\Delta_1 = \tilde{H}\mathcal{T}^T(\Sigma - \tilde{\Sigma})\mathcal{T}H$, $\Delta_{2j} = \mathcal{T}^T(\tilde{\mathcal{V}}_j - \mathcal{V}_j)\mathcal{T}$. It can be seen that

$$\tilde{H} \;=\; H + \sum_{i=1}^{2(p+1)}(\delta_i - \tilde{\delta}_i)HG_i\tilde{H} + \tilde{\psi}_0HT^T(\Delta A_m(\rho 0)\otimes I_n)\mathcal{T}\tilde{H} + \sum_{i=1}^{p}\tilde{\psi}_iH\mathcal{T}^TD\{\Delta A_m(\rho_i)\}\mathcal{T}\tilde{H},$$

where $D(\Delta A_m) = \text{diag}\{\text{diag}(X_{1i})\Delta A_m \text{diag}(X_{1i}), \cdots, \text{diag}(X_{ni})\Delta A_m \text{diag}(X_{ni})\}$, $\Delta A_m(\rho_i) = A_m(\rho_i) - A_m(\tilde{\rho}_i)$.For $1 \leq j \leq (p+1)$, $\tilde{\mathcal{V}}_j = \mathcal{V}_j$; if $j = p+2$, $\mathcal{V}_j - \tilde{\mathcal{V}}_j = \Delta A_m(\rho_0) \otimes I_n$; if $p+3 \leq j \leq 2(p+2)$, $\mathcal{V}_j - \tilde{\mathcal{V}}_j = D\{\Delta A_m(\rho_k)\}$; if $2(p+1)+1 \leq j \leq 2(p+1)+q$, $\mathcal{V}_j - \tilde{\mathcal{V}}_j = \Delta(\partial A_m(\rho_0)/\partial \rho_{0,j^T}) \otimes I_n$ with $j' = j - 2(p+1)$ and if $2(p+1)+(k-1)q+1 \leq j \leq 2(p+1)+kq$, $\mathcal{V}_j - \tilde{\mathcal{V}}_j = D\{\Delta(\partial A_m(\rho_k)/\partial \rho_{k,j'})\}$ with $2 \leq k \leq (p+1)$ and $j' = j - 2(p+1) + (k-1)q$. Since $H, A_m(\rho_i)$ and $A_m^T(\rho_i)$ are positive definite, if γ in S_γ is small enough such that, $(1/2)H \leq \tilde{H} \leq 2H$, $(1/2)A_m(\rho_k) \leq A_m(\tilde{\rho}_k) \leq 2A_m(\rho_k)$ and $(1/2)A_m^T(\rho_k) \leq A_m^T(\tilde{\rho}_k) \leq 2A_m^T(\rho_k)$. Then if $i \leq p+1$

$$\begin{aligned}
\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\| &= \|\tilde{H}^{1/2}G_i\tilde{H}\mathcal{T}^T\epsilon\| \leq \sqrt{2}\|H^{1/2}G_i\tilde{H}\mathcal{T}^T\epsilon\| \\
&\leq \sqrt{2}\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| + \sum_{j=1}^{2(p+1)} |\delta_j - \tilde{\delta}_j|\|H^{1/2}G_iHG_j\tilde{H}\mathcal{T}^T\epsilon\| \\
&\quad + |\tilde{\psi}_0|\|H^{1/2}G_iH\mathcal{T}^T(\Delta A_m(\rho_0) \otimes I_n)\mathcal{T}\tilde{H}\mathcal{T}^T\epsilon\| \\
&\quad + \sum_{i=1}^{p} |\tilde{\psi}_j|\|H^{1/2}G_iH\mathcal{T}^T D(\Delta A_m(\rho_j))\mathcal{T}\tilde{H}\mathcal{T}^T\epsilon\|.
\end{aligned}$$

It can be shown that there exists some constant $C(\gamma)$ such that

$$\begin{aligned}
&\|H^{1/2}G_iH\mathcal{T}^T(\Delta A_m(\rho_0) \otimes I_n)\mathcal{T}\tilde{H}\mathcal{T}^T\epsilon\| \\
\leq\ & C(\gamma)\|H^{1/2}G_iH\tilde{G}_{2p+3}\tilde{H}\mathcal{T}^T\epsilon\| \leq C(\gamma)\|H^{1/2}G_iH^{1/2}\|\|H^{1/2}\tilde{G}_{2p+3}\tilde{H}\mathcal{T}^T\epsilon\|,
\end{aligned}$$

and

$$\|H^{1/2}G_iH\mathcal{T}^T D(\Delta A_m(\rho_j))\mathcal{T}\tilde{H}\mathcal{T}^T\epsilon\| \leq C(\gamma)\|H^{1/2}G_iH^{1/2}\|\|H^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\| \text{ for } k = j+2(p+1)+1.$$

Therefore, for $i \leq p+1$,

$$\begin{aligned}
\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\| &\leq \sqrt{2}\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| + \sqrt{2}\sum_{j=1}^{2(p+1)} |\delta_j - \tilde{\delta}_j|\|H^{1/2}G_iH^{1/2}\|\|H^{1/2}G_j\tilde{H}\mathcal{T}^T\epsilon\| \\
&\quad + \sqrt{2}\sum_{k=2(p+1)+1}^{3(p+1)} C(\gamma)|\tilde{\psi}_{k-2p-3}|\|H^{1/2}G_iH^{1/2}\|\|H^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\| \\
&\leq \sqrt{2}\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| + 2\gamma d_*\|H^{1/2}G_iH^{1/2}\|\sum_{j=1}^{2(p+1)} d_j^{-1}\|\tilde{H}^{1/2}\tilde{G}_j\tilde{H}\mathcal{T}^T\epsilon\| \\
&\quad + 2C(\gamma)\|H^{1/2}G_iH^{1/2}\|\sum_{k=2(p+1)+1}^{3(p+1)} |\tilde{\psi}_{k-2p-3}|\|\tilde{H}^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\|,
\end{aligned}$$

where $\|H^{1/2}G_iH^{1/2}\| = \mathrm{tr}(P\mathcal{V}_iP\mathcal{V}_i)^{1/2}$. For $(q+2)(p+1) \geq i > p+1$,

$$\begin{aligned}\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\| &\leq \sqrt{2}\{1+C^*(\gamma)\}\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| \\ &\quad +2\gamma\{1+C^*(\gamma)\}d_*\|H^{1/2}G_iH^{1/2}\| \sum_{j=1}^{2(p+1)} d_j^{-1}\|\tilde{H}^{1/2}\tilde{G}_j\tilde{H}\mathcal{T}^T\epsilon\| \\ &\quad +2\{1+C^*(\gamma)\}C(\gamma)\|H^{1/2}G_iH^{1/2}\| \sum_{k=2(p+1)+1}^{3(p+1)} |\tilde{\psi}_{k-2p-3}|\|\tilde{H}^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\|.\end{aligned}$$

In summary, define

$$k_j = \begin{cases} 2\gamma\{1+C^*(\gamma)\}d_j^{-1}d_*\|H^{1/2}G_iH^{1/2}\| & \text{for } j \leq p+1 \\ 2C\{1+C^*(\gamma)\}C(\gamma)\|H^{1/2}G_iH^{1/2}\| & \text{for } p+1 \leq j \leq s, \end{cases}$$

then

$$\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\| \leq \sqrt{2}(1+C^*(\gamma))\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| + \sum_{j=1}^{s} k_j\|\tilde{H}^{1/2}\tilde{G}_j\tilde{H}\mathcal{T}^T\epsilon\|. \tag{S6}$$

It follows that

$$\sup_{\tilde{\delta}\in S_\gamma}\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\| \leq \sqrt{2}\{1+C^*(\gamma)\}\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| + \sum_{j=1}^{s} k_j \sup_{\tilde{\delta}\in S_\gamma}\|\tilde{H}^{1/2}\tilde{G}_j\tilde{H}\mathcal{T}^T\epsilon\|. \tag{S7}$$

If we take γ smaller enough such that $\sum_{j=1}^{s} k_j < 1$, then

$$\begin{aligned}\sup_{\tilde{\delta}\in S_\gamma}\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\| &\leq \sqrt{2}\{1+C^*(\gamma)\}\|H^{1/2}G_iH\mathcal{T}^T\epsilon\| \\ &\quad +\sqrt{2}\{1+C^*(\gamma)\}(1-\sum_{j=1}^{s} k_j)^{-1}\sum_{j=1}^{s} k_j\|H^{1/2}G_jH\mathcal{T}^T\epsilon\|. \end{aligned}\tag{S8}$$

For any $g > 4$ and some constant C,

$$E\|H^{1/2}G_jH\mathcal{T}^T\epsilon\|^g = E|\epsilon^T\mathcal{T}HG_jHG_jH\mathcal{T}^T\epsilon|^{g/2} = E|\epsilon^TP\mathcal{V}_jP\mathcal{V}_jP\epsilon|^{g/2} \leq C\mathrm{tr}^{g/2}(P\mathcal{V}_jP\mathcal{V}_j).$$

Hence the first term in $\partial^3\ell(\delta)/(\partial\delta_i\partial\delta_j\partial\delta_k)$ can be bounded by

$$\begin{aligned} |\epsilon^T\tilde{P}\tilde{\mathcal{V}}_k\tilde{P}\tilde{\mathcal{V}}_j\tilde{P}\tilde{\mathcal{V}}_i\tilde{P}\epsilon| &= \epsilon^T\mathcal{T}\tilde{H}\tilde{G}_k\tilde{H}\tilde{G}_j\tilde{H}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon \\ &\leq \lambda_{\max}(\tilde{H}^{1/2}\tilde{G}_j\tilde{H}^{1/2})\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\|\|\tilde{H}^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\| \\ &\leq C_1(\gamma)\lambda_{\max}(H^{1/2}G_jH^{1/2})\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\|\|\tilde{H}^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\|. \end{aligned}$$

Combining (S8) and the above two inequality, it can be seen that

$$\begin{aligned} & E\Big(\frac{d_*}{d_id_jd_k}\sup_{\tilde{\delta}\in S_\gamma}|\epsilon^T\tilde{P}\tilde{\mathcal{V}}_k\tilde{P}\tilde{\mathcal{V}}_j\tilde{P}\tilde{\mathcal{V}}_i\tilde{P}\epsilon|\Big)^g \\ \leq\ & C_1^g(\gamma)\lambda_{\max}^g(H^{1/2}G_jH^{1/2})E(\frac{1}{d_id_k}\sup_{\tilde{\delta}\in S_\gamma}\|\tilde{H}^{1/2}\tilde{G}_i\tilde{H}\mathcal{T}^T\epsilon\|\sup_{\tilde{\delta}\in S_\gamma}\|\tilde{H}^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\|)^g \\ \leq\ & C_1^g(\gamma)\lambda_{\max}^g(H^{1/2}G_jH^{1/2}). \end{aligned}$$

We choose γ small enough such that $C_1(\gamma)\lambda_{\max}(H^{1/2}G_jH^{1/2}) < \infty$. The other terms in $\partial^3\ell(\delta)/(\partial\delta_i\partial\delta_j\partial\delta_k)$ can be bounded similarly. For example,

$$|\epsilon^T\tilde{P}\tilde{\mathcal{V}}_k\tilde{P}\frac{\partial\tilde{\mathcal{V}}_i}{\partial\delta_j}\tilde{P}\epsilon| \;\leq\; \|\tilde{H}^{1/2}\tilde{G}_k\tilde{H}\mathcal{T}^T\epsilon\|\|\tilde{H}^{1/2}\frac{\partial\tilde{G}_k}{\partial\delta_j}\tilde{H}\mathcal{T}^T\epsilon\|,$$

where the bound for the right hand side can be obtained similarly as of (S8). Therefore, condition (iv) in Theorem 2.1 of Das et al. (2004) holds. Notice that from (S5),

$$\begin{aligned}(A)_{ij} = E(\frac{\partial^2\ell(\delta)}{\partial\delta_i\partial\delta_j}) &= -2^{-1}\Big\{\mathrm{tr}(Q_{ij}\Sigma) - \mathrm{tr}(P\mathcal{V}_jP\mathcal{V}_i) - \mathrm{tr}(P\frac{\partial\mathcal{V}_i}{\partial\delta_j})\Big\}\\ &= -2^{-1}\mathrm{tr}(P\mathcal{V}_iP\mathcal{V}_j).\end{aligned}$$

Then the (i,j)th component of $D_1^{-1}AD_1^{-1}$, where $D_1 = \mathrm{Diag}(d_1,\cdots,d_n)$, is $\mathrm{tr}(P\mathcal{V}_iP\mathcal{V}_j)/(d_id_j)$. Condition (iii) in Das et al. (2004) is equivalent to require that the smallest eigenvalue of $-(D^{-1}AD^{-1})$ must be bounded away from 0 and ∞. Suppose the smallest eigenvalue of $-(D^{-1}AD^{-1})$ is $\lambda^*_{\min}$. Since

$$\lambda^*_{\min} = \inf_{x\neq 0}\frac{x^T(-D^{-1})(-A)(-D^{-1})x}{x^Tx} \leq \lambda_{\max}(-A)\inf_{x\neq 0}\frac{x^TD^{-2}x}{x^Tx} \leq \lambda_{\max}(-A)(\min_i(d_i))^{-2} < \infty,$$

we require that

$$\lambda_{\max}(-A) = O(\min_i d_i^2). \tag{S9}$$

Under condition (S9), condition (iii) of Das et al. (2004) holds. Therefore, conditions (i)-(iv) in Theorem 2.1 of Das et al. (2004) hold and g can be any integer greater than 4. This finishes the proof of Lemma 3.

Lemma 4 *Define $t(\delta) = \tilde{l}^T\widehat{b}_F + \tilde{m}^T\widehat{\nu}$ as the BLUP estimator of $\overline{Y}_{i_0}(t_m;\delta)$ for some specific i_0 and $\widehat{\delta}$ be the REML estimator of δ. If conditions (a)-(d) hold, then*

$$E\{t(\widehat{\delta}) - t(\delta)\}^2 = E\Big\{\frac{\partial t(\delta)}{\partial\delta}(\widehat{\delta}-\delta)\Big\}^2 + o(n^{-1}).$$

Proof: For convenience, let us define $\tilde{u} := (\tilde{u}_1^T,\cdots,\tilde{u}_n^T)^T = Y - Z_F\widehat{b}_F$, $u = Y - Z_Fb_F$ and $\zeta^T(\delta) := \tilde{m}^TGW^T\Sigma^{-1} = (\zeta_1^T(\delta),\cdots,\zeta_n^T(\delta))$ where

$$\zeta_k^T(\delta) = \begin{cases} Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_k}^T\Sigma_k^{-1} & \text{if } k\neq i_0\\ Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_{i_0}}^T\Sigma_{i_0}^{-1} & \\ \quad +\Sigma_{u0}^{(m)}\Sigma_{i_0}^{-1} + \sum_{q=1}^p X_{i_0k}(t_m)[\Sigma_{uq}\mathrm{Diag}(X_{qi_0})]^{(m)}\Sigma_{i_0}^{-1} & \text{if } k = i_0.\end{cases}$$

where i_0 is the area we are interested in predicting (in the main text, we used i instead of i_0. In this supplemental, we used i_0), $\Sigma_{u0}^{(m)}$ is the mth row of Σ_{u0} and $[\Sigma_{uq}\mathrm{Diag}(X_{qi_0})]^{(m)}$ is the

mth row of $\Sigma_{uq}\text{Diag}(X_{qi_0})$. Let C_1 and C_2 be constants which may take different values in each appearance. By the Taylor expansion of $t(\widehat{\delta})$ around δ, we have

$$t(\widehat{\delta}) - t(\delta) \;=\; \frac{\partial t(\delta)}{\partial \delta}(\widehat{\delta}-\delta) + \frac{1}{2}(\widehat{\delta}-\delta)^T \frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}}(\widehat{\delta}-\delta)$$

where $\|\delta^* - \delta\| \leq \|\widehat{\delta}-\delta\|$. Then

$$\begin{aligned} E\{t(\widehat{\delta}) - t(\delta)\}^2 &= E\{\frac{\partial t(\delta)}{\partial \delta}(\widehat{\delta}-\delta)\}^2 + E\Big\{\frac{\partial t(\delta)}{\partial \delta}(\widehat{\delta}-\delta)(\widehat{\delta}-\delta)^T \frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}}(\widehat{\delta}-\delta)\Big\} \\ &\quad +(1/4)E\Big\{(\widehat{\delta}-\delta)^T \frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}}(\widehat{\delta}-\delta)\Big\}^2 := E\Big\{\frac{\partial t(\delta)}{\partial \delta}(\widehat{\delta}-\delta)\Big\}^2 + R_1 + R_2. \end{aligned}$$

First, we would like to show $R_2 = o(n^{-1})$. Notice that

$$\begin{aligned} E\Big\{(\widehat{\delta}-\delta)^T \frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}}(\widehat{\delta}-\delta)\Big\}^2 &= E\Big[\text{tr}^2\Big\{\frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}}(\widehat{\delta}-\delta)(\widehat{\delta}-\delta)^T\Big\}\Big] \\ &\leq E\Big(\text{tr}\Big\{(\frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}})^2\Big\}\text{tr}\Big[\{(\widehat{\delta}-\delta)(\widehat{\delta}-\delta)^T\}^2\Big]\Big) \\ &= E\Big(\text{tr}\Big\{(\frac{\partial^2 t(\delta^*)}{\partial \delta^{*2}})^2\Big\}\Big\{(\widehat{\delta}-\delta)^T(\widehat{\delta}-\delta)\Big\}^2\Big) \\ &= \sum_{i=1}^{s}\sum_{j=1}^{s} E\Big[(\frac{\partial^2 t(\delta^*)}{\partial \delta_i^*\partial \delta_j^*})^2\Big\{(\widehat{\delta}-\delta)^T(\widehat{\delta}-\delta)\Big\}^2\Big]. \end{aligned}$$

Because s is a fixed number, we only need to show that

$$E\Big[(\frac{\partial^2 t(\delta^*)}{\partial \delta_i^*\partial \delta_j^*})^2\Big\{(\widehat{\delta}-\delta)^T(\widehat{\delta}-\delta)\Big\}^2\Big] = o(n^{-1}). \tag{S10}$$

The first derivative of $t(\delta)$ is

$$\frac{\partial t(\delta)}{\partial \delta_i} \;=\; \tilde{l}^T \frac{\partial \widehat{b}_F}{\partial \delta_i} + \frac{\partial b^T(\delta)}{\partial \delta_i}\tilde{u} - b^T(\delta) Z_F \frac{\partial \widehat{b}_F}{\partial \delta_i},$$

where $\partial \widehat{b}_F/\partial \delta_i = (Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}(\partial\Sigma/\partial\delta_i)\Sigma^{-1}\tilde{u}$ and the second derivatives of $t(\delta)$ is

$$\begin{aligned} \frac{\partial^2 t(\delta)}{\partial \delta_i \partial \delta_j} &= \tilde{l}^T \frac{\partial^2 \widehat{b}_F}{\partial \delta_i \partial \delta_j} + \frac{\partial^2 \zeta^T(\delta)}{\partial \delta_i \partial \delta_j}\tilde{u} - \frac{\partial \zeta^T(\delta)}{\partial \delta_i} Z_F \frac{\partial \widehat{b}_F}{\partial \delta_j} - \frac{\partial \zeta^T(\delta)}{\partial \delta_j} Z_F \frac{\partial \widehat{b}_F}{\partial \delta_i} - \zeta^T(\delta) Z_F \frac{\partial^2 \widehat{b}_F}{\partial \delta_i \partial \delta_j} \\ &:= J_1(\delta) + J_2(\delta) + J_3(\delta) + J_4(\delta) + J_5(\delta), \end{aligned}$$

where

$$\begin{aligned}
\frac{\partial^2 \widehat{b}_F}{\partial\delta_i\partial\delta_j} &= -(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}Z_F\tilde{u} \\
&\quad -(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}Z_F\tilde{u} \\
&\quad +(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u} \\
&\quad +(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\tilde{u} \\
&\quad -(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\delta_i\partial\delta_j}\Sigma^{-1}\tilde{u} \\
&:= I_1(\delta)+I_2(\delta)+I_3(\delta)+I_4(\delta)+I_5(\delta).
\end{aligned}$$

Let us look at $J_1(\delta)$. We can write $J_1(\delta) = \tilde{l}^T\{I_1(\delta)+I_2(\delta)+I_3(\delta)+I_4(\delta)+I_5(\delta)\}$. Since $\tilde{l}^T I_1(\delta^*)$ and $\tilde{l}^T I_2(\delta^*)$ are similar, we only show that $|\tilde{l}^T I_1(\delta^*)|$ is bounded by $|\tilde{l}^T I_1(\delta^*)| \le C_1 n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$, where C_1 and C_2 are some constants. By the Cauchy-Schwarz inequality and $\Sigma^{-1/2}Z_F^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F\Sigma^{-1/2}$ being an idempotent matrix, we have

$$\begin{aligned}
|\tilde{l}^T I_1(\delta)| &= |\tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}Z_F\tilde{u}| \\
&\le |\tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l}|^{1/2}|\tilde{u}^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}|^{1/2}.
\end{aligned}$$

Denote $d_i^T = \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_{F_i}^T\Sigma_i^{-1/2}$ and

$$P_1(\delta) = |\tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l}|.$$

Then we can write

$$\begin{aligned}
P_1(\delta) &= \sum_{k=1}^n d_k^T\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}d_k \\
&\le \sum_{k=1}^n [\mathrm{tr}(d_kd_k^Td_kd_k^T)]^{1/2}[\mathrm{tr}(\Sigma_k^{-1/2}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1/2})]^{1/2} \\
&= \sum_{k=1}^n (d_k^Td_k)[\mathrm{tr}(\Sigma_k^{-1/2}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1/2})]^{1/2} \\
&\le \sum_{k=1}^n (d_k^Td_k)\mathrm{tr}^2(\Sigma_k^{-1/2}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1/2}) \\
&= \sum_{k=1}^n (d_k^Td_k)\mathrm{tr}^2(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}).
\end{aligned}$$

Similar to the proof of Lemma 2, we have $\mathrm{tr}^2\{\Sigma_k^{*-1}(\partial\Sigma_k^*/\partial\delta_j^*)\} \le C_1\mathrm{tr}^2\{\Sigma_k^{-1}(\partial\Sigma_k/\partial\delta_j)\}(1+C_2\|\delta^*-\delta\|)$ and assuming that $\max_{1\le k\le n}\mathrm{tr}^2\{\Sigma_k^{-1}(\partial\Sigma_k/\partial\delta_j)\}<\infty$, we have

$$\begin{aligned}
P_1(\delta^*) &\le C_1\max_{1\le k\le n}\mathrm{tr}^2(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j})(1+C_2\|\delta^*-\delta\|)\sum_{k=1}^n d_k^{*T}d_k^* \\
&= C_1\max_{1\le k\le n}\mathrm{tr}^2(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j})(1+C_2\|\delta^*-\delta\|)\sum_{k=1}^n \tilde{l}^T(Z_F^T\Sigma^{*-1}Z_F)^{-1}Z_{F_k}^T\Sigma_k^{*-1}Z_{F_k}(Z_F^T\Sigma^{*-1}Z_F)^{-1}\tilde{l} \\
&= C_1\max_{1\le k\le n}\mathrm{tr}^2(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j})(1+C_2\|\delta^*-\delta\|)\tilde{l}^T(Z_F^T\Sigma^{*-1}Z_F)^{-1}\tilde{l}.
\end{aligned}$$

It can be shown that $|\tilde{l}^T(Z_F^T\Sigma^{*-1}Z_F)^{-1}\tilde{l}| = C_1\tilde{l}^T(Z_F^T\Sigma^{*-1}Z_F)^{-1}\tilde{l}(1+C_2\|\delta^*-\delta\|)$. Since $Z_F^T\Sigma^{-1}Z_F = O(n^{-1})$ and $\max_{1\le k\le n}\mathrm{tr}^2\{\Sigma_k^{-1}(\partial\Sigma_k/\partial\delta_j\}<\infty$, $|P_1(\delta^*)|\le C_1n^{-1}(1+C_2\|\delta^*-\delta\|)$. Next,

$$\begin{aligned}
|\tilde{u}^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}| &= \sum_{k=1}^n[\mathrm{tr}(\Sigma_k^{-1/2}\tilde{u}_k\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k\tilde{u}_k^T\Sigma_k^{-1/2})]^{1/2}[\mathrm{tr}(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i})^2]^{1/2} \\
&\le \sum_{k=1}^n\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k\mathrm{tr}(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}) \\
&\le \sum_{k=1}^n\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k\mathrm{tr}^2(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}).
\end{aligned}$$

Hence

$$\begin{aligned}
|\tilde{u}^T\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_i^*}\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_i^*}\Sigma^{*-1}\tilde{u}| &\le C_1\max_{1\le k\le n}\mathrm{tr}^2(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i})(1+C_2\|\delta^*-\delta\|)|\tilde{u}^T\Sigma^{*-1}\tilde{u}| \\
&\le C_1|u^T\Sigma^{-1}u|(1+C_2\|\delta^*-\delta\|). \qquad \text{(S11)}
\end{aligned}$$

Note that we used the fact that $|\tilde{u}^T\Sigma^{-1}\tilde{u}|\le|u^T\Sigma^{-1}u|$. It follows that $|\tilde{l}^TI_1|\le C_1n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$. Similarly, $|\tilde{l}^TI_2|\le C_1n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$. The third term in J_1 is

$$\begin{aligned}
|\tilde{l}^TI_3| &= |\tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}| \\
&\le |\tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l}|^{1/2}|\tilde{u}^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}|^{1/2}.
\end{aligned}$$

Applying the inequality $\mathrm{tr}(A^2)\le\mathrm{tr}^2(A)$ for any nonnegative matrix A, we have

$$\begin{aligned}
&|\tilde{u}^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}| = \sum_{k=1}^n\tilde{u}_k^T\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1}\tilde{u}_k \\
\le &\sum_{k=1}^n[\mathrm{tr}(\Sigma_k^{-1/2}\tilde{u}_k\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k\tilde{u}_k^T\Sigma_k^{-1/2})]^{1/2}[\mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1})^2]^{1/2} \\
\le &\sum_{k=1}^n\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k\mathrm{tr}^2(\Sigma_k^{-1})^2\mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_j})^2\mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_i})^2.
\end{aligned}$$

Hence,

$$|\tilde{u}^T\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_i^*}\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_j^*}\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_j^*}\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_i^*}\Sigma^{*-1}\tilde{u}| \le \sum_{k=1}^n \tilde{u}_k^T{\Sigma^*}_k^{-1}\tilde{u}_k\mathrm{tr}^2({\Sigma^*}_k^{-1})^2\mathrm{tr}(\frac{\partial\Sigma_k^*}{\partial\delta_j^*})^2\mathrm{tr}(\frac{\partial\Sigma_k^*}{\partial\delta_i^*})^2$$

$$\begin{aligned}
&\le C_1\max_k \mathrm{tr}^2(\Sigma_k^{-1})^2\max_k \mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_j})^2\max_k \mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_i})^2(1+C_2\|\delta^*-\delta\|)\sum_{k=1}^n \tilde{u}_k^T{\Sigma^*}_k^{-1}\tilde{u}_k\\
&\le C_1\max_k \mathrm{tr}^2(\Sigma_k^{-1})^2\max_k \mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_j})^2\max_k \mathrm{tr}(\frac{\partial\Sigma_k}{\partial\delta_i})^2(1+C_2\|\delta^*-\delta\|)\sum_{k=1}^n u_k^T\Sigma_k^{-1}u_k,
\end{aligned}$$

and it is easy to see that $|\tilde{l}^T(Z_F^T\Sigma^{*-1}Z_F)^{-1}\tilde{l}| \le C_1|\tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l}|(1+C_2\|\delta^*-\delta\|)$. Therefore, $|\tilde{l}^TI_3(\delta^*)| \le C_1n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$. Similarly, we can bound $|\tilde{l}^TI_3(\delta^*)|$ and $|\tilde{l}^TI_5(\delta^*)|$. So, in summary, $|J_1(\delta^*)| \le C_1n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$.

Next, for J_3, we have

$$\begin{aligned}
|J_3| &= |\frac{\partial\zeta^T(\delta)}{\partial\delta_i}Z_F\frac{\partial\widehat{b}_F}{\partial\delta_j}| = |\frac{\partial\zeta^T(\delta)}{\partial\delta_i}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}|\\
&\le |\frac{\partial\zeta^T(\delta)}{\partial\delta_i}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\frac{\partial\zeta(\delta)}{\partial\delta_i}|^{1/2}|\tilde{u}^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\tilde{u}|^{1/2}.
\end{aligned}$$

Noting that for $k \ne i_0$

$$\frac{\partial\zeta_k^T(\delta)}{\partial\delta_j} = Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_k}^T\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1} + Z_{R_{i_0}}^T(t_m)\frac{\partial\mathrm{cov}(b_R)}{\partial\delta_j}Z_{R_k}^T\Sigma_k^{-1} = O(n^{-1/2}),$$

where $\partial\mathrm{cov}(b_R)/\partial\delta_j = 0$ if $\delta_j \ne \sigma_{b_k}^2$ and $\partial\mathrm{cov}(b_R)/\partial\delta_j = \mathrm{Diag}(0,\cdots,\mathrm{Diag}(\lambda_{L_1}^{-1}),\ldots,0)$ if $\delta_j = \sigma_{b_k}^2$, and $\partial\zeta_{i_0}^T(\delta)/\partial\delta_j = O(1)$ for all δ_j. Hence, $\{\partial\zeta^T(\delta)/\partial\delta_i\}Z_F$ is of order O(1) for each component. It follows that $\{\partial\zeta^T(\delta)/\partial\delta_i\}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\{\partial\zeta(\delta)/\partial\delta_i\} = O(n^{-1})$. We have already shown in (S11) that

$$|\tilde{u}^T\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_i^*}\Sigma^{*-1}\frac{\partial\Sigma^*}{\partial\delta_i^*}\Sigma^{*-1}\tilde{u}| \le C_1|u^T\Sigma^{-1}u|(1+C_2\|\delta^*-\delta\|).$$

Therefore, $|J_3(\delta^*)| \le C_1n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$. Similarly, we can show the same bound for $|J_4(\delta^*)|$.

Now let us check J_5, the proof is almost the same as J_1, where we replace $\tilde{l}$ by $\zeta^T(\delta)Z_F$. Notice that each component of $\zeta^T(\delta)Z_F$ is O(1). Then $|\zeta^T(\delta)Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\zeta(\delta)| = O(n^{-1})$. Hence, as we have shown for J_1, it can also be shown that $|J_5(\delta^*)| \le C_1n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1+C_2\|\delta^*-\delta\|)$. It remains to show $J_2 = O_p(1)$. Notice that for $k \ne i_0$,

$$\begin{aligned}
\frac{\partial\zeta_k^T(\delta)}{\partial\delta_i\partial\delta_j} &= Z_{R_{i_0}}^T(t_m)\frac{\partial\mathrm{cov}(b_R)}{\partial\delta_i}Z_{R_1}^T\Sigma_k^{-1}\frac{\partial\Sigma_1}{\partial\delta_j}\Sigma_k^{-1} + Z_{R_{i_0}}^T(t_m)\frac{\partial\mathrm{cov}(b_R)}{\partial\delta_j}Z_{R_1}^T\Sigma_k^{-1}\frac{\partial\Sigma_1}{\partial\delta_i}\Sigma_k^{-1}\\
&+ Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_k}^T\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1} + Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_1}^T\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\\
&+ Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_k}^T\Sigma_k^{-1}\frac{\partial^2\Sigma_k}{\partial\delta_i\partial\delta_j}\Sigma_k^{-k} = O(n^{-1/2}),
\end{aligned}$$

and $\{\partial\zeta_{i_0}^T(\delta)/\partial\delta_i\partial\delta_j\} = O(1)$. It follows that

$$|J_2| = |\frac{\partial\zeta^T(\delta)}{\partial\delta_i\partial\delta_j}\tilde{u}| = |\sum_{k=1}^n \frac{\partial\zeta_k^T(\delta)}{\partial\delta_i\partial\delta_j}\tilde{u}| \le \sum_{k=1}^n |\frac{\partial\zeta_k^T(\delta)}{\partial\delta_i\partial\delta_j}\tilde{u}_k| \le \sum_{i=1}^n (\frac{\partial\zeta_k^T(\delta)}{\partial\delta_i\partial\delta_j}\Sigma_k\frac{\partial\zeta_k^T(\delta)}{\partial\delta_i\partial\delta_j})^{1/2}(\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k)^{1/2},$$

where $\tilde{u}_k = (\tilde{u}_{k1}, \cdots, \tilde{u}_{km})^T$. It is easy to see that $\tilde{u}_k^T\Sigma_k^{-1}\tilde{u}_k = O_p(1)$. Hence $|J_2(\delta^*)| \le C_1 n^{-1/2}|u^T\Sigma^{-1}u|^{1/2}(1 + C_2\|\delta^* - \delta\|)$. In summary, from $J_1(\delta^*) - J_5(\delta^*)$,

$$\Big|\frac{\partial^2 t(\delta^*)}{\partial\delta_i^*\partial\delta_j^*}\Big| \le C_1 n^{-1/2}(u^T\Sigma^{-1}u)^{1/2}(1 + C_2\|\widehat{\delta} - \delta\|),$$

where C is some constant. Applying the Cauchy-Schwarz inequality,

$$\begin{aligned}
R_2 = E((\frac{\partial^2 t(\delta^*)}{\partial\delta_i^*\partial\delta_j^*})^2\Big[(\widehat{\delta} - \delta)^T(\widehat{\delta} - \delta)\Big]^2) &\le 2Cn^{-1}\{E(u^T\Sigma^{-1}u\|\widehat{\delta} - \delta\|^4) + E(u^T\Sigma^{-1}u\|\widehat{\delta} - \delta\|^6)\} \\
&\le 2Cn^{-1}\Big[\{E(u^T\Sigma^{-1}u)^2\}^{1/2}\{E(\|\widehat{\delta} - \delta\|^8)\}^{1/2} \\
&\quad + \{E(u^T\Sigma^{-1}u)^2\}^{1/2}\{E(\|\widehat{\delta} - \delta\|^{12})\}^{1/2}\Big].
\end{aligned}$$

Because $E(u^T\Sigma^{-1}u)^2 = O(n^2)$ and $E(\|\widehat{\delta} - \delta\|^8) = O(n^{-4})$, we have $R_2 = o(n^{-1})$. To show the order of R_1, we would like to know the order of

$$E\{\frac{\partial t(\delta)}{\partial\delta}(\widehat{\delta} - \delta)\}^2 \le C\sum_{i=1}^s \{E(\frac{\partial t(\delta)}{\partial\delta_i})^4\}^{1/2}\{E(\widehat{\delta}_i - \delta_i)^4)\}^{1/2}.$$

Now we can rewrite $\partial t(\delta)/\partial\delta_j$ in the following form

$$\frac{\partial t(\delta)}{\partial\delta_j} = \left(f_j(\delta) - \zeta(\delta)Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\frac{\partial\Sigma}{\partial\delta_j}D + \frac{\partial b(\delta)}{\partial\delta_j}D\right)\epsilon := h_j(\delta)^T\epsilon, \tag{S12}$$

where $f_j(\delta) = \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T(\partial\Sigma/\partial\delta_j)D$, $D = I - Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1}$. Define $h_j^{(2)}(\delta)^T = \zeta^T(\delta)Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T(\partial\Sigma/\partial\delta_j)D$ and $h_j^{(3)}(\delta) = \{\partial\zeta^T(\delta)/\partial\delta_j\}D$. Since $\epsilon \sim N(0, \Sigma)$,

$$E(\frac{\partial t(\delta)}{\partial\delta_j})^4 = 3(h_j(\delta)^T\Sigma h_j(\delta))^2 \le 6\{(f_j^T(\delta)\Sigma f_j(\delta))^2 + (h_j^{(2)}(\delta)^T\Sigma h_j^{(2)}(\delta))^2 + (h_j^{(3)}(\delta)^T\Sigma h_j^{(3)}(\delta))^2\}.$$

Define $B = Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T$. Notice that $\Sigma^{-1/2}B\Sigma^{-1/2}$ is an idempotent matrix. Then the first term on the right hand side of (S12) is

$$\begin{aligned}
f_j^T(\delta)\Sigma f_j(\delta) &= \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\frac{\partial\Sigma}{\partial\delta_j}D\Sigma D\frac{\partial\Sigma}{\partial\delta_j}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l} \\
&= \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\frac{\partial\Sigma}{\partial\delta_j}(\Sigma - B)\frac{\partial\Sigma}{\partial\delta_j}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l} \\
&\le \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\frac{\partial\Sigma}{\partial\delta_j}\Sigma\frac{\partial\Sigma}{\partial\delta_j}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l} \\
&= \sum_{i=1}^n \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_{F_i}^T\frac{\partial\Sigma_i}{\partial\delta_j}\Sigma_i\frac{\partial\Sigma_i}{\partial\delta_j}Z_{F_i}(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l} \\
&\le \lambda_{\max}(\frac{\partial\Sigma_i}{\partial\delta_j}\Sigma_i\frac{\partial\Sigma_i}{\partial\delta_j})\sum_{i=1}^n \tilde{l}^T(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_{F_i}^TZ_{F_i}(Z_F^T\Sigma^{-1}Z_F)^{-1}\tilde{l},
\end{aligned}$$

which is of order $O(n^{-1})$. Similarly, the second term of the right hand side of (S12) is

$$\begin{aligned} h_j^{(2)}(\delta)^T\Sigma h_j^{(2)}(\delta) &= \zeta(\delta)B\frac{\partial\Sigma}{\partial\delta_j}(\Sigma-B)\frac{\partial\Sigma}{\partial\delta_j}B\zeta^T(\delta) \\ &\le \zeta^T(\delta)B\frac{\partial\Sigma}{\partial\delta_j}\Sigma\frac{\partial\Sigma}{\partial\delta_j}B\zeta(\delta) \le \zeta^T(\delta)\Sigma\frac{\partial\Sigma}{\partial\delta_j}\Sigma\frac{\partial\Sigma}{\partial\delta_j}\Sigma\zeta(\delta)^T \\ &\le \sum_{i=1}^n \zeta_i^T(\delta)\Sigma_i\frac{\partial\Sigma_i}{\partial\delta_j}\Sigma_i\frac{\partial\Sigma_i}{\partial\delta_j}\Sigma_i\zeta_i(\delta). \end{aligned}$$

If $\lambda_{L_1}^{-1} = O(n^{-1/2})$ and $\lambda_{\max}\{(\partial\Sigma_i/\partial\delta_j)\Sigma_i(\partial\Sigma_i/\partial\delta_j)\Sigma_i\} < \infty$. Then $h_j^{(2)}(\delta)^T\Sigma h_j^{(2)}(\delta) = O(1)$.

Then the third term on the right hand side of (S12) is

$$\begin{aligned} h_j^{(3)}(\delta)^T\Sigma h_j^{(3)}(\delta) &= \frac{\partial\zeta^T(\delta)}{\partial\delta_j}D\Sigma D^T\frac{\partial\zeta(\delta)}{\partial\delta_j} = \frac{\partial\zeta(\delta)}{\partial\delta_j}\{\Sigma - Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\}\frac{\partial\zeta(\delta)^T}{\partial\delta_j} \\ &= \frac{\partial\zeta(\delta)}{\partial\delta_j}\Sigma^{1/2}\{I-\Sigma^{-1/2}Z_F(Z_F^T\Sigma^{-1}Z_F)^{-1}Z_F^T\Sigma^{-1/2}\}\Sigma^{1/2}\frac{\partial\zeta(\delta)^T}{\partial\delta_j} \le \frac{\partial\zeta(\delta)}{\partial\delta_j}\Sigma\frac{\partial\zeta(\delta)^T}{\partial\delta_j} \\ &= \sum_{i=1}^n\frac{\partial\zeta_i(\delta)}{\partial\delta_j}\Sigma_i\frac{\partial\zeta_i(\delta)^T}{\partial\delta_j} \le \lambda_{\max}(\Sigma_i)\sum_{i=1}^n\frac{\partial\zeta_i(\delta)}{\partial\delta_j}\frac{\partial\zeta_i(\delta)^T}{\partial\delta_j}. \end{aligned}$$

If $\lambda_{\max}(\Sigma_i) < \infty$ and $\lambda_{L_1}^{-1} = O(n^{-1/2})$, then $\partial\zeta_i(\delta)/\partial\delta_j = O(n^{-1/2})$ and hence $h_j^{(3)}(\delta)^T\Sigma h_j^{(3)}(\delta) = O(1)$. Hence, $E\{(\partial t(\delta)/\partial\delta_i)^4\} = O(1)$. It follows that $E\{(\partial t(\delta)/\partial\delta)(\widehat{\delta}-\delta)\}^2 = O(n^{-1})$. Again by the Cauchy-Schwarz inequality, it is easy to see that $R_1 = o(n^{-1})$. Therefore, we have

$$E\{t(\widehat{\delta}) - t(\delta)\}^2 = E\{\frac{\partial t(\delta)}{\partial\delta}(\widehat{\delta}-\delta)\}^2 + o(n^{-1}).$$

This completes the proof of Lemma 4.

Some additional details in the proof of Theorem 2:

For K_1, because $\tilde{m}^T G \tilde{m} = Z_{R_{i_0}}^T(t_m)\mathrm{cov}(b_R)Z_{R_{i_0}} + e_{im,0}^T(I_n \otimes \Sigma_{u0})e_{im,0} + \sum_{k=1} e_{ik,x_k}^T(I_n \otimes \Sigma_{uk})e_{ik,x_k}$, $\tilde{m}^T(\partial^2 G/\partial\delta_i\partial\delta_j)\tilde{m}$ is a summation of fixed number functions of variance components δ. Therefore, it can be shown that $|K_1^* - K_1| \le C\|\delta^* - \delta\|$. For K_2, notice that

$$\frac{\partial^2\gamma^T(\delta)}{\partial\delta_i\partial\delta_j}\Sigma^{-1}\gamma(\delta) = \begin{cases} 0 & \text{both } \delta_i \text{ and } \delta_j \text{ are } \sigma_{b_k}^2\text{s;} \\ \frac{\partial^2\gamma_{i_0}^T(\delta)}{\partial\delta_i\partial\delta_j}\Sigma_k^{-1}\gamma_{i_0}(\delta) & \text{if one of } \delta_i \text{ and } \delta_j \text{ is not } \sigma_{b_k}^2\text{s,} \end{cases}$$

where $\gamma_{i_0}(\delta)$ is the i_0th m-dimensional subvector of $\gamma^T(\delta) = (\gamma_1^T(\delta), \cdots, \gamma_m^T(\delta))$. Therefore,

$$\begin{aligned} |K_2^* - K_2| &\le |\mathrm{tr}\{\gamma_{i_0}(\delta^*)\frac{\partial^2\gamma_{i_0}^T(\delta^*)}{\partial\delta_i\partial\delta_j}\Sigma_k^{*-1}\} - \mathrm{tr}\{\gamma_{i_0}(\delta)\frac{\partial^2\gamma_{i_0}^T(\delta)}{\partial\delta_i\partial\delta_j}\Sigma_k{}^{-1}\}| \\ &\le |\mathrm{tr}\{(\gamma_{i_0}(\delta^*)\frac{\partial^2\gamma_{i_0}^T(\delta^*)}{\partial\delta_i\partial\delta_j} - \gamma_{i_0}(\delta)\frac{\partial^2\gamma_{i_0}^T(\delta)}{\partial\delta_i\partial\delta_j}\}\Sigma_k^{-1})| \\ &+ |\mathrm{tr}\{\gamma_{i_0}(\delta)\frac{\partial^2\gamma_{i_0}^T(\delta)}{\partial\delta_i\partial\delta_j}(\Sigma_k^{*-1} - \Sigma_k^{-1})\}| \\ &+ |\mathrm{tr}\{(\gamma_{i_0}(\delta^*)\frac{\partial^2\gamma_{i_0}^T(\delta^*)}{\partial\delta_i\partial\delta_j} - \gamma_{i_0}(\delta)\frac{\partial^2\gamma_{i_0}^T(\delta)}{\partial\delta_i\partial\delta_j})(\Sigma_k^{*-1} - \Sigma_k^{-1})\}| := K_2^{(1)} + K_2^{(2)} + K_2^{(3)}. \end{aligned}$$

From Lemma 2 we know $\mathrm{tr}(\Sigma_k^{*-1} - \Sigma_k^{-1}) \le C\|\widehat{\delta} - \delta\|$, hence to show that $|K_2^* - K_2| \le C\|\widehat{\delta} - \delta\|$, it is enough to show that $|(\partial^2\gamma_{i_0}^{(l)}(\delta^*)/\partial\delta_i\partial\delta_j)\gamma_{i_0}^{(k)}(\delta^*) - (\partial^2\gamma_{i_0}^{(l)}(\delta)/\partial\delta_i\partial\delta_j)\gamma_{i_0}^{(k)}(\delta)| \le C\|\widehat{\delta} - \delta\|$ where subscript (l) denotes the lth component. Notice that

$$\begin{aligned} |\frac{\partial^2\gamma_{i_0}^{(l)}(\delta^*)}{\partial\delta_i\partial\delta_j}\gamma_{i_0}^{(k)}(\delta^*) - \frac{\partial^2\gamma_{i_0}^{(l)}(\delta)}{\partial\delta_i\partial\delta_j}\gamma_{i_0}^{(k)}(\delta)| &\le C|\frac{\partial^2\gamma_{i_0}^{(l)}(\delta^*)}{\partial\delta_i\partial\delta_j} - \frac{\partial^2\gamma_{i_0}^{(l)}(\delta)}{\partial\delta_i\partial\delta_j}| + C|\gamma_{i_0}^{(k)}(\delta^*) - \gamma_{i_0}^{(k)}(\delta)| \\ &+ C|\frac{\partial^2\gamma_{i_0}^{(l)}(\delta^*)}{\partial\delta_i\partial\delta_j} - \frac{\partial^2\gamma_{i_0}^{(l)}(\delta)}{\partial\delta_i\partial\delta_j}||\gamma_{i_0}^{(k)}(\delta^*) - \gamma_{i_0}^{(k)}(\delta)|. \end{aligned}$$

Clearly, $|\gamma_{i_0}^{(k)}(\delta^*) - \gamma_{i_0}^{(k)}(\delta)| \le C\|\widehat{\delta} - \delta\|$ from the expression of $\gamma_{i_0}^{(k)}(\delta)$ and it also easy to show that $|\partial^2\gamma_{i_0}^{(l)}(\delta^*)/\partial\delta_i\partial\delta_j - \partial^2\gamma_{i_0}^{(l)}(\delta)/\partial\delta_i\partial\delta_j| \le C\|\widehat{\delta} - \delta\|$. It follows that $|K_2^* - K_2| \le C\|\widehat{\delta} - \delta\|$.

The derivation of K_3 to K_7 are similar, here we only give the details for K_4. We first write

$$\begin{aligned} |K_4^* - K_4| &\le C\sum_{k=1}^n |\frac{\partial\gamma_k^T(\delta^*)}{\partial\delta_i}\Sigma_k^{*-1}\frac{\partial\Sigma_k^*}{\partial\delta_j}\Sigma^*{}_k^{-1}\gamma_k(\delta^*) - \frac{\partial\gamma_k^T(\delta)}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\gamma_k(\delta)| \\ &\le C\sum_{k=1}^n (K_{41}^{(k)} + K_{42}^{(k)} + K_{41}^{(k)}K_{42}^{(k)}), \end{aligned}$$

where $K_{41}^{(k)} = |\mathrm{tr}[\{\gamma_k(\delta^*)(\partial\gamma_k^T(\delta^*)/\partial\delta_i) - \gamma_k(\delta)(\partial\gamma_k^T(\delta)/\partial\delta_i)\}\Sigma_k^{-1}(\partial\Sigma_k/\partial\delta_j)\Sigma_k^{-1}]|$ and $K_{42}^{(k)} = |\mathrm{tr}[\{\Sigma^*{}_k^{-1}(\partial\Sigma_k^*/\partial\delta_j^*)\Sigma^*{}_k^{-1} - \Sigma_k^{-1}(\partial\Sigma_k/\partial\delta_j)\Sigma_k^{-1}\}\gamma_k(\delta)(\partial\gamma_k^T(\delta)/\partial\delta_i)]|$. It can be seen that

$$\begin{aligned} K_{41}^{(k)} &= |\mathrm{tr}\{(\gamma_k(\delta^*) - \gamma_k(\delta))\frac{\partial\gamma_k^T(\delta)}{\partial\delta_i}\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\}| + |\mathrm{tr}\{\gamma_k(\delta)(\frac{\partial\gamma_k^T(\delta^*)}{\partial\delta_i} - \frac{\partial\gamma_k^T(\delta)}{\partial\delta_i})\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\}| \\ &+ |\mathrm{tr}\{(\gamma_k(\delta^*) - \gamma_k(\delta))(\frac{\partial\gamma_k^T(\delta^*)}{\partial\delta_i} - \frac{\partial\gamma_k^T(\delta)}{\partial\delta_i})\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\delta_j}\Sigma_k^{-1}\}|. \end{aligned}$$

For $k \neq i_0$, $|\gamma_k^{(l)}(\delta^*) - \gamma_k^{(l)}(\delta)| \leq C\|\lambda_{L_1}\|^{-1}\sum_{k=0}^p |\widehat{\sigma}_{b_k}^2 - \sigma_{b_k}^2|$ and each element of $\partial\gamma_k^T(\delta)/\partial\delta_i$ is of the same order of $\|\lambda_{L_1}\|^{-1}$. Hence, $|\text{tr}\{(\gamma_k(\delta^*) - \gamma_k(\delta))(\partial\gamma_k^T(\delta)/\partial\delta_i)\Sigma_k^{-1}(\partial\Sigma_k/\partial\delta_j)\Sigma_k^{-1}\}| \leq C\|\lambda_{L_1}\|^{-2}\|\widehat{\delta}-\delta\|$. Similarly, we can show the other terms are also bounded by $C\|\lambda_{L_1}\|^{-2}\|\widehat{\delta}-\delta\|$. It follows that $K_{41}^{(k)} \leq C\|\lambda_{L_1}\|^{-2}\|\widehat{\delta}-\delta\|$ if $k \neq i_0$. By noting that $\text{tr}(\Sigma^{*-1}_k - \Sigma_k^{-1})^2 \leq C\|\widehat{\delta}-\delta\|$, $\text{tr}\{(\partial\Sigma_k^*/\partial\delta_j^*) - (\partial\Sigma_k/\partial\delta_j)\}^2 \leq C\|\widehat{\delta}-\delta\|$, $\gamma_k(\delta)$ and $\partial\gamma_k^T(\delta)/\partial\delta_i$ are both $O(\|\lambda_{L_1}\|^{-1})$ for $k \neq i_0$, it can be shown that $K_{42}^{(k)} \leq C\|\lambda_{L_1}\|^{-2}\|\widehat{\delta}-\delta\|$ for $k \neq i_0$. For $k = i_0$, $K_{41}^{(i_0)} \leq C\|\widehat{\delta}-\delta\|$ and $K_{42}^{(i_0)} \leq C\|\widehat{\delta}-\delta\|$. In summary, using the assumption $\|\lambda_{L_1}\| = O(n^{-1/2})$, we have

$$|K_4^* - K_4| \leq C(\sum_{k\neq i_0}\|\lambda_{L_1}\|^{-2} + 1)\|\widehat{\delta}-\delta\| \leq C\|\widehat{\delta}-\delta\|.$$

Here we show that $\partial g_4(\delta)/\partial\delta_i = o(n^{-1/2})$. Observe that

$$\begin{aligned}
\frac{\partial g_4^{jl}(\delta)}{\partial\delta_k} &= \frac{\partial\eta_j^T}{\partial\delta_k}\Sigma P\mathcal{V}_jP\mathcal{V}_lP\Sigma\eta_l + \eta_j^T\frac{\partial\Sigma}{\partial\delta_k}P\mathcal{V}_jP\mathcal{V}_lP\Sigma\eta_l + \eta_j^T\Sigma P\mathcal{V}_kPP\mathcal{V}_jP\mathcal{V}_lP\Sigma\eta_l \\
&\quad +\eta_j^T\Sigma P\frac{\partial\mathcal{V}_j}{\partial\delta_k}P\mathcal{V}_lP\Sigma\eta_l + \eta_j^T\Sigma P\mathcal{V}_jP\mathcal{V}_kP\mathcal{V}_lP\Sigma\eta_l + \eta_j^T\Sigma P\mathcal{V}_jP\frac{\partial\mathcal{V}_l}{\partial\delta_k}P\Sigma\eta_l \\
&\quad +\eta_j^T\Sigma P\mathcal{V}_jP\mathcal{V}_lP\mathcal{V}_kP\Sigma\eta_l + \eta_j^T\Sigma P\mathcal{V}_jP\mathcal{V}_lP\frac{\partial\Sigma}{\partial\delta_k}\eta_l + \eta_j\Sigma P\mathcal{V}_jP\mathcal{V}_lP\Sigma\frac{\partial\eta_l^T}{\partial\delta_k}.
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$|\eta_j^T\Sigma P\frac{\partial\mathcal{V}_j}{\partial\delta_k}P\mathcal{V}_lP\Sigma\eta_l| \leq (\eta_j^T\Sigma P\frac{\partial\mathcal{V}_j}{\partial\delta_k}P\frac{\partial\mathcal{V}_j}{\partial\delta_k}P\Sigma\eta_j)^{1/2}(\eta_l^T\Sigma P\mathcal{V}_lP\mathcal{V}_lP\Sigma\eta_l)^{1/2}$$

By the definition of η and $h(\delta)^T\Sigma h(\delta) = o(n^{-3/2})$, we can see that $|\eta_j^T\Sigma P(\partial\mathcal{V}_j/\partial\delta_k)P\mathcal{V}_lP\Sigma\eta_l| = o(n^{-1/2})$. The order of the other terms of $\partial g_4(\delta)/\partial\delta_i$ can be derived similarly.