Supplementary Materials for "Heteroscedastic Nested Error Regression Models with Variance Functions"

Shonosuke Sugasawa[1] and Tatsuya Kubokawa[2]

[1]*The Institute of Statistical Mathematics and* [2]*University of Tokyo*

In this supplementary file, we provide the additional figure in the data application and the proofs of the results in the paper.

S1 Proof of Theorem 1

Since $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are mutually independent, the consistency of $\widehat{\boldsymbol{\gamma}}$ follows from the standard argument, so that $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\beta}}$ are also consistent. In what follows, we derive the asymptotic expressions of the estimators.

First we consider the asymptotic approximation of $\widehat{\tau}^2 - \tau^2$. From (6), we obtain

$$
\begin{aligned}
\widehat{\tau}^2 - \tau^2 &= \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}_{\rm OLS})^2 - \widehat{\sigma}_{ij}^2\right\} - \tau^2 \\
&= \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2 - \sigma_{ij}^2\right\} - \tau^2 - \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\sigma_{ij(1)}^2\boldsymbol{z}_{ij}'(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\
&\quad - \frac{2}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})\boldsymbol{x}_{ij}'(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta}) + o_p(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta}) \\
&= \frac{1}{m}\sum_{i=1}^m u_{1i} - \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\sigma_{ij(1)}^2\boldsymbol{z}_{ij}'(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(m^{-1/2}) + o_p(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}), \qquad \text{(S1)}
\end{aligned}
$$

where $u_{1i} = mN^{-1}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2 - \sigma_{ij}^2\right\} - \tau^2$ and we used the fact that $\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta} = O_p(m^{-1/2})$ and $N^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})\boldsymbol{x}_{ij} = O_p(m^{-1/2})$ from the central limit theorem.

For the asymptotic expansion of $\widehat{\boldsymbol{\gamma}}$, remember that the estimator $\widehat{\boldsymbol{\gamma}}$ is given as the solution of the estimating equation

$$
\frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{\rm OLS}\right\}^2\boldsymbol{z}_{ij} - \sigma_{ij}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right] = \boldsymbol{0}
$$

Using Taylor expansions, we have

$$
\begin{aligned}
\boldsymbol{0} &= \frac{1}{m}\sum_{i=1}\boldsymbol{u}_{2i} - \frac{2}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}\boldsymbol{z}_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta}) \\
&\quad - \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(m^{-1/2}),
\end{aligned}
$$

where

$$\boldsymbol{u}_{2i} = mN^{-1}\sum_{j=1}^{n_i}\left[\left\{y_{ij}-\bar{y}_i-(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2\boldsymbol{z}_{ij}-\sigma_{ij}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\right].$$

From the central limit theorem, it follows that

$$\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{y_{ij}-\bar{y}_i-(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}\boldsymbol{z}_{ij}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)' = O_p(m^{-1/2}),$$

so that the second terms in the expansion formula is $o_p(m^{-1/2})$. Then we get

$$\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} = \frac{N}{m}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'\right)^{-1}\sum_{i=1}^{m}\boldsymbol{u}_{2i}+o_p(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(m^{-1/2}).$$

Under (A1)-(A5), we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}' = O(m).$$

From the independence of $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m$ and the fact $E(\boldsymbol{u}_{2i})=\boldsymbol{0}$, we can use the central limit theorem to show that the leading term in the expansion of $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ is $O_p(m^{-1/2})$. Thus,

$$\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} = \frac{N}{m}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'\right)^{-1}\sum_{i=1}^{m}\boldsymbol{u}_{2i}+o_p(m^{-1/2}).$$

Using the approximation of $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ and $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=O_p(m^{-1/2})$, we get the asymptotic expression of $\widehat{\tau}^2-\tau^2$ from (S1), which establishes the result for $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\gamma}}$.

Finally we consider the asymptotic expansion of $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}$. From the expression in (4), it follows that

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}+\sum_{s=1}^{q}\left(\frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}\right)'(\widehat{\gamma}_s-\gamma)+\left(\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}\right)'(\widehat{\tau}^2-\tau^2)+o_p(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(\widehat{\tau}^2-\tau^2).$$

Since

$$\frac{\partial}{\partial\tau^2}\boldsymbol{\Sigma}_i = \boldsymbol{J}_{n_i}, \quad \frac{\partial}{\partial\gamma_s}\boldsymbol{\Sigma}_i = \boldsymbol{W}_{i(s)}, \quad s=1,\ldots,q,$$

for $\boldsymbol{W}_{i(s)} = \mathrm{diag}(\sigma_{i1(1)}^2 z_{i1s},\ldots,\sigma_{in_i(1)}^2 z_{in_is})$, we have

$$\begin{aligned}\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{n_i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)\left(\widetilde{\boldsymbol{\beta}}_{\tau}^{*}-\widetilde{\boldsymbol{\beta}}\right),\\ \frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)\left(\widetilde{\boldsymbol{\beta}}_{\gamma_s}^{*}-\widetilde{\boldsymbol{\beta}}\right), \quad s=1\ldots,q,\end{aligned} \tag{S2}$$

where

$$\widetilde{\boldsymbol{\beta}}_{\tau}^{*}=\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{n_i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{n_i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i,$$

$$\widetilde{\boldsymbol{\beta}}_{\gamma_s}^{*}=\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i,\quad s=1,\ldots,q.$$

Under (A1)-(A5), we have $\widetilde{\boldsymbol{\beta}}_a^{*}-\boldsymbol{\beta}=O_p(m^{-1/2})$ for $a\in\{\tau,\gamma_1,\ldots,\gamma_q\}$, whereby $\widetilde{\boldsymbol{\beta}}^{*}-\widetilde{\boldsymbol{\beta}}=O_p(m^{-1/2})$. Since $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=O_p(m^{-1/2})$ and $\widehat{\tau}^2-\tau^2=O_p(m^{-1/2})$ as shown above, we get

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})+o_p(m^{-1/2}),$$

which completes the proof.

S2 Proof of Corollary 1

Let $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_{p+q+1})'=(\boldsymbol{\beta}',\boldsymbol{\gamma}',\tau^2)'$. Note that $\psi_i^{\theta_k},k=1,\ldots,p+q+1$ does not depend on $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{i-1},\boldsymbol{y}_{i+1},\ldots,\boldsymbol{y}_m$ and that $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m$ are mutually independent. Then,

$$\begin{aligned}\frac{1}{m^2}E\left[\left(\sum_{j=1}^{m}\psi_j^{\theta_k}\right)\left(\sum_{j=1}^{m}\psi_j^{\theta_l}\right)\middle|\boldsymbol{y}_i\right]&=\frac{1}{m^2}\sum_{j=1,j\neq i}^{m}E\left[\psi_j^{\theta_k}\psi_j^{\theta_l}\right]+\frac{1}{m^2}\psi_i^{\theta_k}\psi_i^{\theta_l}\\&=\boldsymbol{\Omega}_{kl}+\frac{1}{m^2}\left\{\psi_i^{\theta_k}\psi_i^{\theta_l}-E\left[\psi_i^{\theta_k}\psi_i^{\theta_l}\right]\right\},\end{aligned}$$

where $\boldsymbol{\Omega}_{kl}$ is the (k,l)-element of $\boldsymbol{\Omega}$ and we used the fact that $E[\psi_j^{\theta_k}|\boldsymbol{y}_i]=E[\psi_j^{\theta_k}]=0$ for $j\neq i$. Hence, we get the result from the asymptotic approximation of $\widehat{\boldsymbol{\theta}}$ given in Theorem 1.

S3 Proof of Theorem 2

We begin by deriving the conditional asymptotic bias of $\widehat{\boldsymbol{\gamma}}$. Let $\widetilde{\boldsymbol{\gamma}}$ be the solution of the equation

$$\boldsymbol{F}(\boldsymbol{\gamma};\boldsymbol{\beta})\equiv\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\left\{y_{ij}-\bar{y}_i-(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2\boldsymbol{z}_{ij}-\sigma_{ij}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\right]=\boldsymbol{0}$$

with $\sigma_{ij}^2=\sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$. For notational simplicity, we use $\boldsymbol{F}$ instead of $\boldsymbol{F}(\boldsymbol{\gamma};\boldsymbol{\beta})$ without any confusion and $F_r,r=1,\ldots,q$ denotes the r-th component of $\boldsymbol{F}$, namely $\boldsymbol{F}=(F_1,\ldots,F_q)'$. Define the derivatives $\boldsymbol{F}_{(\boldsymbol{a})}$ and $F_{h(\boldsymbol{ab})}$ by

$$\boldsymbol{F}_{(\boldsymbol{a})}=\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{a}'},\qquad F_{r(\boldsymbol{ab})}=\frac{\partial^2F_r}{\partial\boldsymbol{a}\partial\boldsymbol{b}'}.$$

It is noted that $F_{h(\boldsymbol{\beta\gamma})}=0$. Expanding $\boldsymbol{F}(\widehat{\boldsymbol{\gamma}};\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})=\boldsymbol{0}$, we obtain

$$\boldsymbol{0}=\boldsymbol{F}+\boldsymbol{F}_{(\boldsymbol{\gamma})}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+\boldsymbol{F}_{(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}-\boldsymbol{\beta})+\frac{1}{2}\boldsymbol{t}_1+\frac{1}{2}\boldsymbol{t}_2+o_p(m^{-1}),$$

where $\boldsymbol{t}_s = (t_{s1}, \dots, t_{sq}), s = 1, 2$ for

$$t_{1r} = (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' F_{r(\boldsymbol{\gamma}\boldsymbol{\gamma})} (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}), \qquad t_{2r} = (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})' F_{r(\boldsymbol{\beta}\boldsymbol{\beta})} (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}).$$

It is also noted that

$$\begin{aligned}
\boldsymbol{F}_{(\boldsymbol{\gamma})} &= -\frac{1}{m} \sum_{k=1}^{m} \sum_{j=1}^{n_k} \sigma^2_{kj(1)} (\boldsymbol{z}_{kj} - 2n_k^{-1} \boldsymbol{z}_{kj} + n_k^{-1} \bar{\boldsymbol{z}}_k) \boldsymbol{z}'_{kj} \\
\boldsymbol{F}_{(\boldsymbol{\beta})} &= -\frac{2}{N} \sum_{k=1}^{m} \sum_{j=1}^{n_k} \left\{ y_{kj} - \bar{y}_k - (\boldsymbol{x}_{kj} - \bar{\boldsymbol{x}}_k)' \boldsymbol{\beta} \right\} \boldsymbol{z}_{ij} (\boldsymbol{x}_{kj} - \bar{\boldsymbol{x}}_k)',
\end{aligned}$$

so that $\boldsymbol{F}_{(\boldsymbol{\gamma})}$ is non-stochastic. Thus we have

$$E[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} | \boldsymbol{y}_i] = -(\boldsymbol{F}_{(\boldsymbol{\gamma})})^{-1} \left\{ E[\boldsymbol{F}(\boldsymbol{\gamma}; \boldsymbol{\beta}) | \boldsymbol{y}_i] + E\left[\boldsymbol{F}_{(\boldsymbol{\beta})} (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \Big| \boldsymbol{y}_i \right] + \frac{1}{2} E[\boldsymbol{t}_1 | \boldsymbol{y}_i] + \frac{1}{2} E[\boldsymbol{t}_2 | \boldsymbol{y}_i] \right\} + o_p(m^{-1}).$$

In what follows, we shall evaluate the each term in the parenthesis in the above expression. For the first term, since $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ are mutually independent and $E(\boldsymbol{u}_{2i}) = \boldsymbol{0}$, we have

$$E[\boldsymbol{F}(\boldsymbol{\gamma}; \boldsymbol{\beta}) | \boldsymbol{y}_i] = \frac{1}{m} \boldsymbol{u}_{2i}.$$

For evaluation of the second term, we define $\boldsymbol{Z}_{kr} = \text{diag}(z_{k1r}, \dots, z_{kn_k r})$, where z_{kjr} denotes the r-th element of $\boldsymbol{z}_{kj}$. Then it follows that

$$\begin{aligned}
&E\left[\boldsymbol{F}_{r(\boldsymbol{\beta})} (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \Big| \boldsymbol{y}_i \right] = -\frac{2}{N} \sum_{k=1}^{m} E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})' \boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \Big| \boldsymbol{y}_i \right] \\
&= -\frac{2}{N} \sum_{k=1, k \neq i}^{m} E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})' \boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \Big| \boldsymbol{y}_i \right] - \frac{2}{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{E}_i \boldsymbol{Z}_{ir} \boldsymbol{E}_i \boldsymbol{X}_i E\left[\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta} \Big| \boldsymbol{y}_i \right].
\end{aligned}$$

Noting that it holds for $\ell = 1, \dots, m$ and $k \neq i$

$$E\left[(\boldsymbol{y}_\ell - \boldsymbol{X}_\ell \boldsymbol{\beta})(\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})' \Big| \boldsymbol{y}_i \right] = 1_{\{\ell = k\}} \boldsymbol{\Sigma}_k, \qquad E[\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta} | \boldsymbol{y}_i] = \left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{X}'_i (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}),$$

we have

$$\begin{aligned}
&E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})' \boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \Big| \boldsymbol{y}_i \right] \\
&\quad = \sum_{\ell=1}^{m} \text{tr} \left\{ \boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'_k E\left[(\boldsymbol{y}_\ell - \boldsymbol{X}_\ell \boldsymbol{\beta})(\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})' \Big| \boldsymbol{y}_i \right] \right\} \\
&\quad = \text{tr} \left\{ (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'_k \boldsymbol{\Sigma}_k \boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k \right\},
\end{aligned}$$

which is $O(m^{-1})$ and

$$\frac{1}{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k E\left[\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta} \Big| \boldsymbol{y}_i \right] = o_p(m^{-1}).$$

Thus, we get

$$E\left[\boldsymbol{F}_{r(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}}_{\text{OLS}}-\boldsymbol{\beta})\Big|\boldsymbol{y}_i\right]=-\frac{2}{m}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\text{tr}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\right\}+o_p(m^{-1}), \tag{S3}$$

where the leading term is $O(m^{-1})$. For the third and forth terms, note that

$$F_{r(\boldsymbol{\gamma}\boldsymbol{\gamma})}=-\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(2)}^2(\boldsymbol{z}_{kj}-2n_k^{-1}\boldsymbol{z}_{kj}+n_k^{-1}\bar{\boldsymbol{z}}_k)\boldsymbol{z}_{kj}'z_{kjr}\qquad F_{r(\boldsymbol{\beta}\boldsymbol{\beta})}=\frac{2}{N}\sum_{k=1}^{m}\boldsymbol{X}_k'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k,$$

which are non-stochastic. Then for $h=1,\ldots,q$,

$$E[t_{1r}|\boldsymbol{y}_i]=-\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}z_{kjr}\sigma_{kj(2)}^2(\boldsymbol{z}_{kj}-2n_k^{-1}\boldsymbol{z}_{kj}+n_k^{-1}\bar{\boldsymbol{z}}_k)'\boldsymbol{\Omega}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\boldsymbol{z}_{kj}+o_p(m^{-1}),$$
$$E[t_{2r}|\boldsymbol{y}_i]=\frac{2}{N}\sum_{k=1}^{m}\text{tr}\left(\boldsymbol{X}_k'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\boldsymbol{V}_{\text{OLS}}\right)+o_p(m^{-1}),$$

for $\boldsymbol{V}_{\text{OLS}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$, where we used Corollary 1 and

$$E\left[(\widehat{\boldsymbol{\beta}}_{\text{OLS}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{\text{OLS}}-\boldsymbol{\beta})'\big|\boldsymbol{y}_i\right]=\boldsymbol{V}_{\text{OLS}}+o_p(m^{-1}), \tag{S4}$$

which follows from the similar argument to the proof of Corollary 1. Thus we obtain

$$E[\boldsymbol{t}_1|\boldsymbol{y}_i]=-\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\boldsymbol{z}_{kj}\sigma_{kj(2)}^2(\boldsymbol{z}_{kj}-2n_k^{-1}\boldsymbol{z}_{kj}+n_k^{-1}\bar{\boldsymbol{z}}_k)'\boldsymbol{\Omega}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\boldsymbol{z}_{kj}+o_p(m^{-1}),$$
$$E[\boldsymbol{t}_2|\boldsymbol{y}_i]=\frac{2}{N}\sum_{k=1}^{m}\left\{\text{tr}\left(\boldsymbol{X}_k'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\boldsymbol{V}_{\text{OLS}}\right)\right\}_r+o_p(m^{-1}),$$

where $\{\boldsymbol{a}_r\}_r$ denotes the q-dimensional vector $(a_1,\ldots,a_q)$. Therefore, we have established the result for $\widehat{\boldsymbol{\gamma}}$ in (13).

We next derive the result for $\widehat{\tau}^2$. Let

$$\widetilde{\tau}^2=\frac{1}{N}\sum_{k=1}^{m}\left\{(\boldsymbol{y}_k-\boldsymbol{X}_k\boldsymbol{\beta})'(\boldsymbol{y}_k-\boldsymbol{X}_k\boldsymbol{\beta})-\sum_{j=1}^{n_k}\sigma_{kj}^2\right\}.$$

Using the Taylor series expansion, we have

$$\begin{aligned}\widehat{\tau}^2=\widetilde{\tau}^2&+\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}}(\widehat{\gamma}-\gamma)+\frac{1}{2}(\widehat{\gamma}-\gamma)'\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right)(\widehat{\gamma}-\gamma)\\&+\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_{\text{OLS}}-\boldsymbol{\beta})+\frac{1}{2}(\widehat{\boldsymbol{\beta}}_{\text{OLS}}-\boldsymbol{\beta})'\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)(\widehat{\boldsymbol{\beta}}_{\text{OLS}}-\boldsymbol{\beta})+o_p(m^{-1}),\end{aligned}$$

where we used the fact that $\partial^2\widetilde{\tau}^2/\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}' = 0$. The straight calculation shows that

$$\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}} = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma^2_{kj(1)}\boldsymbol{z}_{kj}, \qquad \frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma^2_{kj(2)}\boldsymbol{z}_{kj}\boldsymbol{z}'_{kj}, \qquad \frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = \frac{2}{N}\sum_{k=1}^{m}\boldsymbol{X}'_i\boldsymbol{X}_i,$$

which are non-stochastic. Thus we obtain

$$\begin{aligned}
E[\widehat{\tau}^2 - \tau^2|\boldsymbol{y}_i] &= E[\widetilde{\tau}^2 - \tau^2|\boldsymbol{y}_i] + \left(\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}}\right)' E\left[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}|\boldsymbol{y}_i\right] + \frac{1}{2}\mathrm{tr}\left\{\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right) E\left[(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'|\boldsymbol{y}_i\right]\right\} \\
&+ E\left[\left(\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\beta}}\right)'(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] + \frac{1}{2}\mathrm{tr}\left\{\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) E\left[(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})'|\boldsymbol{y}_i\right]\right\} + o_p(m^{-1}) \\
&\equiv B_{\tau 1}(\boldsymbol{y}_i) + B_{\tau 2}(\boldsymbol{y}_i) + B_{\tau 3}(\boldsymbol{y}_i) + B_{\tau 4}(\boldsymbol{y}_i) + B_{\tau 5}(\boldsymbol{y}_i) + o_p(m^{-1}).
\end{aligned}$$

From the expression of $\widetilde{\tau}^2$, it holds that

$$\begin{aligned}
B_{\tau 1}(\boldsymbol{y}_i) &= \frac{1}{N}\sum_{k=1, k\neq i}^{m} n_k\tau^2 + \frac{1}{N}\left\{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) - \sum_{j=1}^{n_i}\sigma^2_{ij}\right\} - \tau^2 \\
&= \left(1 - \frac{n_i}{N}\right)\tau^2 + \frac{1}{m}u_{1i} + \frac{n_i}{N}\tau^2 - \tau^2 = \frac{1}{m}u_{1i},
\end{aligned}$$

for u_{1i} defined in (8). Also, we immediately have

$$B_{\tau 2}(\boldsymbol{y}_i) = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma^2_{kj(1)}\boldsymbol{z}'_{kj}\boldsymbol{b}^{(i)}_{\gamma}(\boldsymbol{y}_i)$$

For evaluation of $B_{\tau 4}(\boldsymbol{y}_i)$, note that

$$\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\beta}} = -\frac{2}{N}\sum_{k=1}^{m}\boldsymbol{X}'_k(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}).$$

Similarly to (S3), we get

$$\begin{aligned}
B_{\tau 4}(\boldsymbol{y}_i) &= -\frac{2}{N}\sum_{k=1}^{m} E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\boldsymbol{X}_k(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] \\
&= -\frac{2}{N}\sum_{k=1}^{m}\mathrm{tr}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'_k\boldsymbol{\Sigma}_k\boldsymbol{X}_k\right\} + o_p(m^{-1}).
\end{aligned}$$

Moreover, Corollary 1 and (S4) enable us to obtain the expression of $B_{\tau 3}(\boldsymbol{y}_i)$ and $B_{\tau 5}(\boldsymbol{y}_i)$, whereby we get

$$b^{(i)}_{\tau}(\boldsymbol{y}_i) = m^{-1}u_{1i} - \frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma^2_{kj(1)}\boldsymbol{z}'_{kj}\left\{\boldsymbol{b}^{(i)}_{\gamma}(\boldsymbol{y}_i) - \boldsymbol{b}_{\gamma}\right\} + b_{\tau},$$

which completes the proof for $\widehat{\tau}^2$ in (13).

We finally derive the result for $\widehat{\boldsymbol{\beta}}$. By the Taylor series expansion,

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}+\sum_{s=1}^{q}\left(\frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\gamma}_s-\gamma)+\left(\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\tau}^2-\tau^2)+o_p(m^{-1}),$$

since

$$\left(\frac{\partial\widetilde{\boldsymbol{\beta}}}{\partial\boldsymbol{\phi}}\right)'(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})'\left(\frac{\partial\widetilde{\boldsymbol{\beta}}}{\partial\boldsymbol{\phi}}\right)=o_p(m^{-1}),$$

from $\partial\widetilde{\boldsymbol{\beta}}/\partial\boldsymbol{\phi}=O_p(m^{-1/2})$ as shown in the proof of Theorem 1. From (S2), we have

$$\begin{aligned}&\sum_{s=1}^{q}\left(\frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\gamma}_s-\gamma_s)\\&\quad=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\sum_{s=1}^{q}\left(\sum_{k=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)\left\{\left(\widetilde{\boldsymbol{\beta}}^*_{\gamma_s}-\boldsymbol{\beta}\right)(\widehat{\gamma}_s-\gamma_s)-(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\gamma}_s-\gamma_s)\right\},\end{aligned}$$

and

$$\left(\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\tau}^2-\tau^2)=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\sum_{k=1}^{m}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{J}_{n_k}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k\right)\left\{(\widetilde{\boldsymbol{\beta}}^*_{\tau}-\boldsymbol{\beta})(\widehat{\tau}^2-\tau^2)-(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\tau}^2-\tau^2)\right\}.$$

Let $\boldsymbol{\Omega}_{\beta^*\gamma_s}=E[(\widetilde{\boldsymbol{\beta}}^*_{\gamma_s}-\boldsymbol{\beta})(\widehat{\gamma}_s-\gamma_s)]$ and $\boldsymbol{\Omega}_{\beta^*\tau}=E[(\widetilde{\boldsymbol{\beta}}^*_{\tau}-\boldsymbol{\beta})(\widehat{\tau}-\tau)]$. Then it can be shown that

$$E[(\widetilde{\boldsymbol{\beta}}^*_{\tau}-\boldsymbol{\beta})(\widehat{\tau}-\tau)|\boldsymbol{y}_i]=\boldsymbol{\Omega}_{\beta^*\gamma_s}+o_p(m^{-1}),\quad E[(\widetilde{\boldsymbol{\beta}}^*_{\gamma_s}-\boldsymbol{\beta})(\widehat{\gamma}_s-\gamma_s)|\boldsymbol{y}_i]=\boldsymbol{\Omega}_{\beta^*\tau}+o_p(m^{-1}),$$

which can be proved by the same arguments as in Corollary 1. Thus from Corollary 1 and the fact that

$$E\left[\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}|\boldsymbol{y}_i\right]=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta}),$$

we obtain the result for $\widehat{\boldsymbol{\beta}}$ in (13).

S4 Proof of (18)

From the expansion of $\widehat{\mu}_i$, we have

$$E\left[(\widehat{\mu}_i-\widetilde{\mu}_i)^2\right]=E\left[\left\{\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right\}^2\right]+\frac{1}{2}U_1+\frac{1}{4}U_2,$$

where

$$\begin{aligned}U_1&=E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right]\\U_2&=E\left[\left\{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right\}^2\right].\end{aligned}$$

It is noted that

$$U_1 = \sum_{j=1}^{p+q+1}\sum_{k=1}^{p+q+1}\sum_{\ell=1}^{p+q+1} E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\theta_j}\right)\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\theta_k\partial\theta_\ell}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\widehat{\theta}_j-\theta_j)(\widehat{\theta}_k-\theta_k)(\widehat{\theta}_\ell-\theta_\ell)\right] \equiv \sum_{j=1}^{p+q+1}\sum_{k=1}^{p+q+1}\sum_{\ell=1}^{p+q+1} U_{1jk\ell},$$

and

$$\begin{aligned}
|U_{1jkl}| &\le E\left[\left|\left(\frac{\partial\widetilde{\mu}_i}{\partial\theta_j}\right)\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\theta_k\partial\theta_\ell}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)\right|\left|(\widehat{\theta}_j-\theta_j)(\widehat{\theta}_k-\theta_k)(\widehat{\theta}_\ell-\theta_\ell)\right|\right] \\
&\le E\left[\left|\left(\frac{\partial\widetilde{\mu}_i}{\partial\theta_j}\right)\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\theta_k\partial\theta_\ell}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)\right|^4\right]^{1/4} E\left[\left|(\widehat{\theta}_j-\theta_j)(\widehat{\theta}_k-\theta_k)(\widehat{\theta}_\ell-\theta_\ell)\right|^{4/3}\right]^{3/4} \qquad \text{(S5)}
\end{aligned}$$

using Holder's inequality. Since both $\partial\widetilde{\mu}_i/\partial\theta_j$ and $\partial^2\widetilde{\mu}_i/\partial\theta_k\partial\theta_\ell$ are linear functions of $\boldsymbol{y}_i$, the first term of (S5) is finite under (A4). Moreover, from Theorem 1, it follows $\sqrt{m}|\widehat{\theta}_j-\theta_j| \le C(\boldsymbol{y})$ for some quadratic function of $\boldsymbol{y}$, so that the second term in (S5) is also finite. Hence, we have $U_1 = o(m^{-1})$. Similarly, we also obtain $U_2 = o(m^{-1})$. Therefore, using Corollary 1, we have

$$\begin{aligned}
E\left[(\widehat{\mu}_i-\widetilde{\mu}_i)^2\right] &= E\left[\left\{\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right\}^2\right] + o(m^{-1}) \\
&= \mathrm{tr}\left\{E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)' E\left((\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\Big|\boldsymbol{y}_i\right)\right]\right\} + o(m^{-1}) \\
&= \mathrm{tr}\left\{E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'\boldsymbol{\Omega} + \left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)' c(\boldsymbol{y}_i)o(m^{-1})\right]\right\} + o(m^{-1}) \\
&= \mathrm{tr}\left\{E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'\right]\boldsymbol{\Omega}\right\} + o(m^{-1})
\end{aligned}$$

since $c(\boldsymbol{y}_i)$ is fourth-order function of $\boldsymbol{y}_i$ and $\partial\widetilde{\mu}_i/\partial\boldsymbol{\theta}$ is a linear function of $\boldsymbol{y}_i$, which completes the proof.

S5 Derivation of $R_{31i}(\boldsymbol{\phi}, \boldsymbol{\kappa})$

Since $\boldsymbol{y}_i$ given $v_i, \boldsymbol{\epsilon}_i$ is non-stochastic, we have

$$\begin{aligned}
&E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\right] \\
&= E\left[E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\Big|v_i,\boldsymbol{\epsilon}_i\right]\right] = E\left[E(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}|\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\right] \\
&= E\left[\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\beta}}\right)w_i\right] + E\left[\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\gamma}}\right)w_i\right] + E\left[b_{\tau}^{(i)}(\boldsymbol{y}_i)\left(\frac{\partial\widetilde{\mu}_i}{\partial\tau}\right)w_i\right] + o(m^{-1}) \\
&\equiv R_{31i}(\boldsymbol{\phi}) + o(m^{-1}).
\end{aligned}$$

It is noted that $E(w_i) = 0$ and

$$E\left[(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right] = E\left[(v_i+\varepsilon_{ij})w_i\right] = \left(\sum_{j=1}^{n_i}\lambda_{ij}-1\right)\tau^2 + \sum_{j=1}^{n_i}\lambda_{ij}\sigma_{ij}^2 = 0. \qquad \text{(S6)}$$

Using the expression (13) and (17), it follows that

$$
\begin{aligned}
E\left[\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\beta}}\right)w_i\right] &= \left(\boldsymbol{c}_i-\sum_{j=1}^{n_i}\lambda_{ij}\boldsymbol{x}_{ij}\right)'\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}E\big[(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})w_i\big]=0\\
E\left[\boldsymbol{b}_{\gamma}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\gamma}\right)w_i\right] &= \eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}'\left(\sum_{k=1}^{m}\sum_{h=1}^{n_k}\sigma_{kh(1)}^2\boldsymbol{z}_{kh}\boldsymbol{z}_{kh}'\right)^{-1}\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\\
E\left[b_{\tau}^{(i)}(\boldsymbol{y}_i)\left(\frac{\partial\widetilde{\mu}_i}{\partial\tau}\right)w_i\right] &= m^{-1}\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\Big\{M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa})-\boldsymbol{T}_1(\gamma)'\boldsymbol{T}_2(\gamma)\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\Big\},
\end{aligned}
$$

where

$$
\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})=E\left[\boldsymbol{u}_{2i}(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right],\qquad M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa})=E\left[u_{1i}(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right].
$$

To evaluate M_{1ij} and $\boldsymbol{M}_{2ij}$, we first prove the following result for fixed $j,k,\ell\in\{1,\ldots,n_i\}$.

$$
\begin{aligned}
E\big[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})w_i\big] &= \tau^2\eta_i^{-1}\Big[\tau^2(3-\kappa_v)+\kappa_\varepsilon\sigma_{ij}^2 1_{\{j=k=\ell\}}+\sigma_{ij}^2(1_{\{j=k\neq\ell\}}-1_{\{j=k\}})\\
&\quad+\sigma_{ij}^2(1_{\{j=\ell\neq k\}}-1_{\{j=\ell\}})+\sigma_{ik}^2(1_{\{k=\ell\neq j\}}-1_{\{k=\ell\}})\Big].
\end{aligned}
\tag{S7}
$$

To show (S7), we note that the left side can be rewritten as

$$
-\eta_i^{-1}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})v_i\right]+\sum_{h=1}^{n_i}\lambda_{ih}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})\varepsilon_{ih}\right] \tag{S8}
$$

from the definition of w_i. Using the fact that $\varepsilon_{i1},\ldots,\varepsilon_{in_i}$ and v_i are independent, the first term in (S8) is calculated as

$$
E\left[v_i^4+(\varepsilon_{ij}\varepsilon_{ik}+\varepsilon_{ij}\varepsilon_{i\ell}+\varepsilon_{ik}\varepsilon_{i\ell})v_i^2\right]=\kappa_v\tau^4+\tau^2\left(\sigma_{ij}^2 1_{\{j=k\}}+\sigma_{ij}^2 1_{\{j=\ell\}}+\sigma_{ik}^2 1_{\{k=\ell\}}\right).
$$

Moreover, we have

$$
\begin{aligned}
&E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})\varepsilon_{ih}\right]=E\left[\varepsilon_{ih}(\varepsilon_{ij}+\varepsilon_{i\ell}+\varepsilon_{ik})v_i^2+\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{i\ell}\varepsilon_{ih}\right]\\
&\quad=\tau^2\sigma_{ih}^2\left(1_{\{h=j\}}+1_{\{h=k\}}+1_{\{h=\ell\}}\right)+\kappa_\varepsilon\sigma_{ih}^4 1_{\{j=k=\ell=h\}}\\
&\qquad+\sigma_{ih}^2\left(\sigma_{ij}^2 1_{\{j=k\neq\ell=h\}}+\sigma_{ij}^2 1_{\{j=\ell\neq k=h\}}+\sigma_{ik}^2 1_{\{j=h\neq k=\ell\}}\right),
\end{aligned}
$$

whereby the second term in (S8) can be calculated as

$$
\tau^2\eta_i^{-1}\left[3\tau^2+\kappa_\varepsilon\sigma_{ij}^2 1_{\{j=k=\ell\}}+\sigma_{ij}^2 1_{\{j=k\neq\ell\}}+\sigma_{ij}^2 1_{\{j=\ell\neq k\}}+\sigma_{ik}^2 1_{\{k=\ell\neq j\}}\right],
$$

where we used the expression $\lambda_{ih}=\tau^2\eta_i^{-1}\sigma_{ih}^{-2}$. Then we established the result (S7). From (S7), we immediately have

$$
\begin{aligned}
\sum_{\ell=1}^{n_i}E\big[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})w_i\big] &= \tau^2\eta_i^{-1}\left[n_i\tau^2(3-\kappa_v)+\sigma_{ij}^2(\kappa_\varepsilon-3)1_{\{j=k\}}\right]\\
&= E\big[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})^2w_i\big].
\end{aligned}
$$

Now, we return to the evaluation of M_{1ij} and $\boldsymbol{M}_{2ij}$. It follows that

$$
\begin{aligned}
M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= \frac{m}{N}\sum_{h=1}^{n_i} E\left[(y_{ih}-\boldsymbol{x}_{ih}'\boldsymbol{\beta})^2(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right] \\
&= mN^{-1}\eta_i^{-1}\tau^2\Big\{n_i\tau^2(3-\kappa_v)+\sigma_{ij}^2(\kappa_\varepsilon-3)\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= \frac{m}{N}\sum_{h=1}^{n_i}\boldsymbol{z}_{ih}E\left[\{v_i+\varepsilon_{ih}-(v_i+\bar{\varepsilon}_i)\}^2(v_i+\varepsilon_{ij})w_i\right] \\
&= \frac{m}{N}\sum_{h=1}^{n_i}\boldsymbol{z}_{ih}\Bigg\{E\left[(v_i+\varepsilon_{ih})^2(v_i+\varepsilon_{ij})w_i\right]-2n_i^{-1}\sum_{k=1}^{n_i}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{ih})w_i\right] \\
&\quad + n_i^{-2}\sum_{k=1}^{n_i}\sum_{\ell=1}^{n_i}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})w_i\right]\Bigg\}.
\end{aligned}
$$

Using the identity given in (S7), we have

$$
\begin{aligned}
\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= mN^{-1}\tau^2\eta_i^{-1}\sum_{h=1}^{n_i}\boldsymbol{z}_{ih}\Big\{\sigma_{ij}^2(\kappa_\varepsilon-3)(1_{\{j=h\}}-2n_i^{-1}1_{\{j=h\}}+n_i^{-2})\Big\} \\
&= mN^{-1}\tau^2\eta_i^{-1}n_i^{-2}(n_i-1)^2(\kappa_\varepsilon-3)\sigma_{ij}^2\boldsymbol{z}_{ij},
\end{aligned}
$$

which completes the result in (20).

S6 Evaluation of $R_{32i}(\boldsymbol{\phi})$

Since $\boldsymbol{y}_i$ given v_i and $\boldsymbol{\epsilon}_i$ is non-stochastic, we have

$$
\begin{aligned}
R_{32i}(\boldsymbol{\phi}) &= \frac{1}{2}E\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\right] = \frac{1}{2}E\left[E\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\Big|v_i,\boldsymbol{\epsilon}_i\right]\right] \\
&= \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Omega}E\left[\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)w_i\right]\right\}+o(m^{-1})E\left[\mathrm{tr}\left\{c(\boldsymbol{y}_i)\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)\right\}w_i\right],
\end{aligned}
$$

where we used Corollary 1 in the last equation. Note that

$$
\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}+\sum_{k=1}^{p+q+1}(\theta_k^*-\theta_k)\left(\frac{\partial^3\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\partial\theta_k}\Big|_{\theta_k=\theta_k^{**}}\right), \tag{S9}
$$

where θ_k^{**} is an intermediate value between θ_k^* and θ_k. Further note that the third order partial derivatives of $\widetilde{\mu}_i$ is a linear function of $\boldsymbol{y}_i$, so that the second term of R_{32i} is $o(m^{-1})$. Similarly, it follows that

$$
E\left[\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)w_i\right] = E\left[\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)w_i\right]+o(1)=o(1),
$$

since the second order partial derivatives of $\widetilde{\mu}_i$ is a linear function of $y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta}$ and the identity (S6). Therefore, we finally get $R_{32i}(\boldsymbol{\phi})=o(m^{-1})$.

S7 Predicted values of μ_i in data analysis

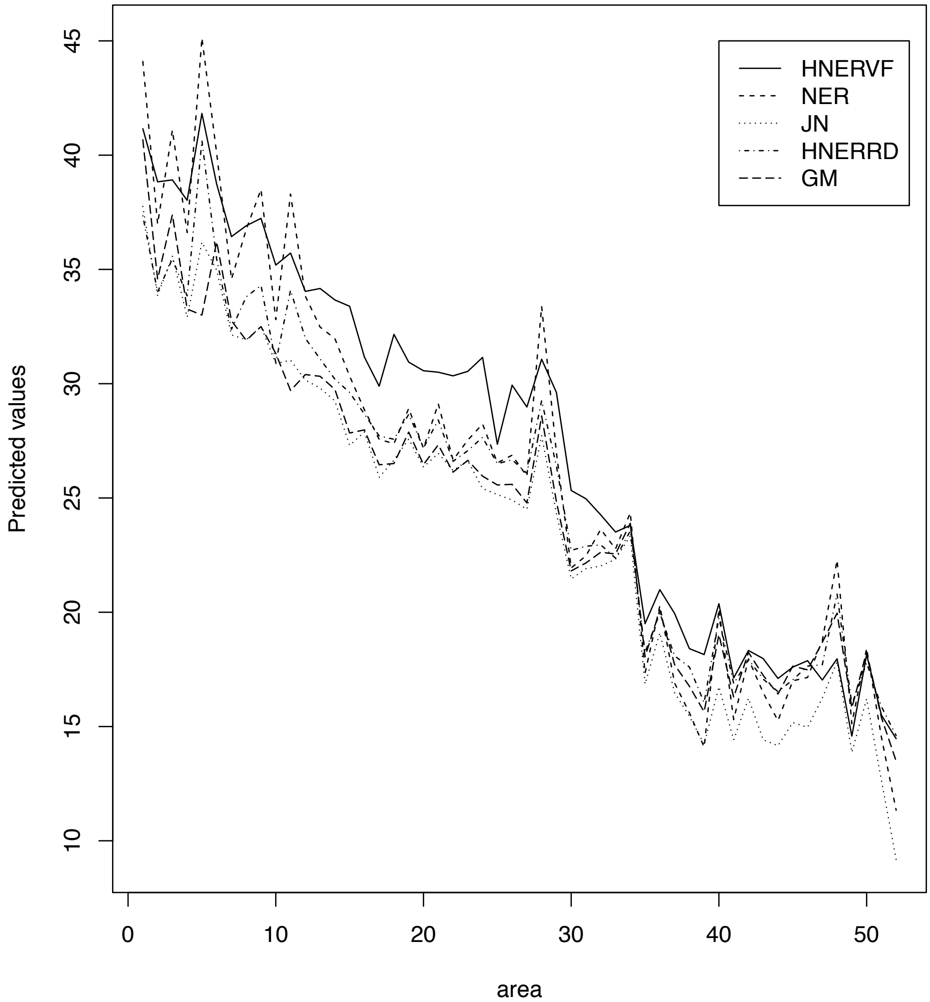

Figure 1: Predicted Values of μ_i from Each Model.