Statistica Sinica: Supplement

GENERALIZED S-ESTIMATORS FOR LINEAR MIXED EFFECTS MODELS

Inna Chervoneva and Mark Vishnyakov

Thomas Jefferson University

Supplementary Material

This supplementary material provides proofs of Theorems 1-3, derivation of the fixed point equations for $\widehat{\boldsymbol{\theta}}_M$, and Supplementary Figures 1-6.

1 Proof of Theorem 1.

Using Lagrange multipliers to solve the constrained minimization problem, one minimizes

$$L = D - \lambda g\left(\mathbf{y}, \boldsymbol{\xi}\right) = \sum\nolimits_{i=1}^{M} v_i \ln\left[\det(\mathbf{V}_i)\right] - \lambda \left[\sum\nolimits_{i=1}^{M} \rho\left(d_i\right) - b_M\right].$$

By taking derivatives one obtains the system of $q + l + 1$ estimating equations,

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\beta}} &= \frac{\partial g}{\partial \boldsymbol{\beta}} = \mathbf{0} \\ \frac{\partial L}{\partial \boldsymbol{\theta}} &= \frac{\partial D}{\partial \boldsymbol{\theta}} - \lambda \frac{\partial g}{\partial \boldsymbol{\theta}} = \mathbf{0} \\ \frac{\partial L}{\partial \lambda} &= g\left(\mathbf{y}, \boldsymbol{\xi}\right) = \mathbf{0} \end{aligned} \tag{S1}$$

where

$$\frac{\partial g}{\partial \boldsymbol{\beta}} = \frac{1}{2} \sum_{i=1}^{M} u(d_i) \mathbf{A}_i' \mathbf{V}_i^{-1} \left(\mathbf{y}_i - \mathbf{A}_i \boldsymbol{\beta}\right) \tag{S2}$$

$$\frac{\partial D}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{l} e_j^l \otimes \sum_{i=1}^{M} v_i tr \left[\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_j}\right] \tag{S3}$$

$$\frac{\partial g}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{l} e_j^l \otimes \sum_{i=1}^{M} \left[-\frac{1}{2} \sum_{i=1}^{M} u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \frac{\partial \mathbf{V}_i}{\partial \theta_j} \mathbf{V}_i^{-1/2} \mathbf{r}_i\right], \tag{S4}$$

e_j^l is j^{th} canonical $l \times 1$ vector and $u_{m_i}(d) = \frac{1}{d}\frac{\partial}{\partial d}\rho_{m_i}(d)$ is the weighting function corresponding to the $\rho_{m_i}-$function. System (S1) implies that λ must satisfy the equation $\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial \boldsymbol{\theta}} - \lambda \mathbf{1}'_{l\times 1}\frac{\partial g}{\partial \boldsymbol{\theta}} = 0$, where $\mathbf{1}_{l\times 1}$ is $l \times 1$ vector of ones, and, therefore,

$$\lambda = \left(\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial \boldsymbol{\theta}}\right)\left(\mathbf{1}'_{l\times 1}\frac{\partial g}{\partial \boldsymbol{\theta}}\right)^{-1}.$$

Then, for any j, equation $\frac{\partial \mathbf{L}}{\partial \theta_j} = 0$ in (S1) may be written as

$$\left(\mathbf{1}'_{l\times 1}\frac{\partial g}{\partial \boldsymbol{\theta}}\right)\frac{\partial D}{\partial \theta_j} - \left(\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial \boldsymbol{\theta}}\right)\frac{\partial g}{\partial \theta_j} = 0. \tag{S5}$$

Denoting $\mathbf{r}_i = \mathbf{V}_i^{-\frac{1}{2}}(\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})$ and using (S4), we can write (S5) as

$$\sum_{i=1}^{M} u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i = 0, \text{ where } \mathbf{F}_{ji} = \left(\sum_{k=1}^{l}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\right)\frac{\partial D}{\partial \theta_j} - \left(\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial \boldsymbol{\theta}}\right)\frac{\partial \mathbf{V}_i}{\partial \theta_j}, \tag{S6}$$

and $\mathbf{F}_{ji}$ is the same as in Theorem 1. Hence, system (S1) may be re-written as system (2.10)-(2.12) of $q + l + 1$ equations in $q + l$ unknowns. System (2.10)-(2.12) is conditionally unbiased under the true model parameters $\left[\boldsymbol{\beta}_0', \boldsymbol{\theta}_0'\right]'$ and normal distribution. This is trivial for (2.12). For (2.10), this easily follows from each summation term being an odd function of $\mathbf{r}_i$ and each $\mathbf{r}_i$ having a symmetric distribution around zero under the true model (e.g. Lemma 1 in Kackar and Harville, 1981). Finally, equations (2.11) are conditionally unbiased if $v_i = m_i^{-1}E_{d^2\sim\chi^2_{m_i}}\left[d^2u(d)\right]$. It is enough to show that under the true normal model, for any $j = 1, ...l$,

$$\frac{\partial D}{\partial \theta_j}\sum_{i=1}^{M} E_{\mathbf{r}_i\sim\mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\left(\sum_{k=1}^{l}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\right)\mathbf{V}_i^{-1/2}\mathbf{r}_i\right] \tag{S7}$$
$$= \left(\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial \boldsymbol{\theta}}\right)\sum_{k=1}^{M} E_{\mathbf{r}_i\sim\mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right]$$

To shorten the notation, let us denote $\mathbf{x}_i = [u(d_i)]^{1/2}\,\mathbf{r}_i$ and $\mathbf{W}_i = \sum_{k=1}^{l}\frac{\partial \mathbf{V}_i}{\partial \theta_k}$.

Then, $\sum_{i=1}^{M} E_{\mathbf{r}_i\sim\mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\left(\sum_{k=1}^{l}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\right)\mathbf{V}_i^{-1/2}\mathbf{r}_i\right] =$

$$\sum_{i=1}^{M} E_{\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[\mathbf{x}_i'\mathbf{V}_i^{-1/2}\mathbf{W}_i\mathbf{V}_i^{-1/2}\mathbf{x}_i\right] = \sum_{i=1}^{M} tr\left[\mathbf{V}_i^{-1/2}\mathbf{W}_i\mathbf{V}_i^{-1/2}\mathbf{Cov}\left(\mathbf{x}_i\right)\right].$$

For symmetric $\rho(d)$ with symmetric $u(d)$, under the true model, $\mathbf{x}_i$ has a symmetric distribution and $E_{\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[\mathbf{x}_i\right] = \mathbf{0}$. Then $\mathbf{Cov}\left(\mathbf{x}_i\right) = E_{\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[u(d_i)\mathbf{r}_i\mathbf{r}_i'\right]$ is a diagonal matrix with equal diagonal elements $E\left[z_1^2 u(z_1^2 + ... + z_{m_i}^2)\right]$, where $z_1, ..., z_{m_i}$ are i.i.d. $N(0,1)$, because off diagonal elements of matrix $u(d_i)\mathbf{r}_i\mathbf{r}_i'$ are asymmetric functions in each component of vector $\mathbf{r}_i$. Hence,

$$\sum_{i=1}^{M} E_{\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\mathbf{W}_i\mathbf{V}_i^{-1/2}\mathbf{r}_i\right] = \sum_{i=1}^{M} m_i^{-1} E_{d^2 \sim \chi^2_{m_i}}\left[d^2 u(d)\right] tr\left[\mathbf{V}_i^{-1}\mathbf{W}_i\right],$$

$$\sum_{i=1}^{M} E_{\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right] = \sum_{i=1}^{M} m_i^{-1} E_{d^2 \sim \chi^2_{m_i}}\left[d^2 u(d)\right] tr\left[\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\right],$$

and $v_i = E_{d^2 \sim \chi^2_{m_i}}\left[d^2 u(d)\right] m_i^{-1}$ implies that both sides of (S7) are equal to

$$\left(\sum_{i=1}^{M} v_i tr\left[\mathbf{V}_i^{-1}\mathbf{W}_i\right]\right)\sum_{i=1}^{M} v_i tr\left[\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\right]. \square$$

2 Proof of Theorem 2.

The GS-estimator may be viewed as a GEE estimator with estimating equations (3.1)-(3.2). The following regularity conditions (Yuan and Jennrich, 1998) imply existence of a consistent sequence of GS-estimators:

C1 With probability 1, $\mathbf{\Psi_M}(\mathbf{y}, \boldsymbol{\xi}_\mathbf{0}) \to \mathbf{0}$, where $\boldsymbol{\xi}_0$ is the true value of $\boldsymbol{\xi}$.

C2 There is a neighborhood of $\boldsymbol{\xi}_0$, $N(\boldsymbol{\xi}_0)$, such that with probability 1, all $\mathbf{\Psi_M}(\mathbf{y}, \boldsymbol{\xi})$ are continuously differentiable and Jacobians $\frac{\partial}{\partial \boldsymbol{\xi}}\mathbf{\Psi_M}(\mathbf{y}, \boldsymbol{\xi})$ converge uniformly to a nonstochastic limit, which is nonsingular at $\boldsymbol{\xi}_0$.

C1 follows from Lemma 5.3 in Shao (2003) under conditions

C3 For each $h_i(\mathbf{y}_i) = \sup_{\boldsymbol{\xi}\in\Theta}\left\|\mathbf{\Psi}_i(\mathbf{y}_i, \boldsymbol{\xi})\right\|$, $i = 1, ..., M$, there exists $\delta > 0$ such that $\sup_i E\left|h_i(\mathbf{y}_i)\right|^{1+\delta} < \infty$ and $\sup_i E\left\|\mathbf{y}_i\right\|^{\delta} < \infty$.

C4 for any $c > 0$ and any sequence $\{\mathbf{x}_i\} \in \mathcal{R}^{m_i}$ such that $\|\mathbf{x}_i\| \leq c$, the sequence of functions $\{g_i(\boldsymbol{\xi}) = \boldsymbol{\Psi}_i(\mathbf{x}_i, \boldsymbol{\xi})\}$ is equicontinuos on any open subset of Θ.

Under assumptions **R1-R4**, conditions **C3** and **C4** are satisfied. To show that **C2** holds, we compute the Jacobian $\frac{\partial}{\partial \boldsymbol{\xi}}\boldsymbol{\Psi}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi})$, which is partitioned as

$$\frac{\partial}{\partial \boldsymbol{\xi}}\boldsymbol{\Psi}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) = \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}'}\boldsymbol{\Psi}^{\boldsymbol{\beta}}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) & \frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{\Psi}^{\boldsymbol{\beta}}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) \\ \frac{\partial}{\partial \boldsymbol{\beta}'}\boldsymbol{\Psi}^{\boldsymbol{\theta}}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) & \frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{\Psi}^{\boldsymbol{\theta}}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) \end{bmatrix}$$

where $\boldsymbol{\Psi}^{\boldsymbol{\beta}}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{\Psi}_{\boldsymbol{\beta}\mathbf{i}}(\mathbf{y}_i, \boldsymbol{\xi})$, $\boldsymbol{\Psi}^{\boldsymbol{\theta}}_{\mathbf{M}}(\mathbf{y}, \boldsymbol{\xi}) = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{\Psi}_{\boldsymbol{\theta}\mathbf{i}}(\mathbf{y}_i, \boldsymbol{\xi})$ and

$$\frac{\partial}{\partial \boldsymbol{\beta}'}\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\beta}}(\mathbf{y_i}, \boldsymbol{\xi}) = -u(d_i)\mathbf{A}_i'\mathbf{V}_i^{-1}\mathbf{A}_i - u'(d_i)d_i^{-1}\mathbf{A}_i'\mathbf{V}_i^{-1/2}\mathbf{r}_i\mathbf{r}_i'\mathbf{V}_i^{-1/2}\mathbf{A}_i$$

$$\begin{aligned} \frac{\partial}{\partial \theta_j}\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\beta}}(\mathbf{y_i}, \boldsymbol{\xi}) &= u(d_i)\mathbf{A}_i'\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i \\ &\quad + u'(d_i)d_i^{-1}\left(\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right)\mathbf{A}_i'\mathbf{V}_i^{-1/2}\mathbf{r}_i \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}'}\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}_\mathbf{j}}(\mathbf{y_i}, \boldsymbol{\xi}) &= -2u'(d_i)d_i^{-1}\mathbf{A}_i'\mathbf{V}_i^{-1/2}\mathbf{r}_i\left[\mathbf{r}_i'\mathbf{V}_i^{-1/2}\mathbf{F}_{ji}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right] \\ &\quad -2u(d_i)\mathbf{A}_i'\mathbf{V}_i^{-1}\mathbf{F}_{ji}\mathbf{V}_i^{-1/2}\mathbf{r}_i - \frac{c_j}{2}u(d_i)\mathbf{A}_i'\mathbf{V}_i^{-1/2}\mathbf{r}_i \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{\Psi}_{\boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\xi}) &= \left(\mathbf{1}_{l\times 1}'\frac{\partial g}{\partial \boldsymbol{\theta}}\right)\frac{\partial^2 D}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} + \frac{\partial D}{\partial \boldsymbol{\theta}}\left(\mathbf{1}_{l\times 1}'\frac{\partial^2 g}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) \\ &\quad -\left(\mathbf{1}_{l\times 1}'\frac{\partial D}{\partial \boldsymbol{\theta}}\right)\frac{\partial^2 g}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - \frac{\partial g}{\partial \boldsymbol{\theta}}\left(\mathbf{1}_{l\times 1}'\frac{\partial^2 D}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) + \mathbf{c}\frac{\partial g}{\partial \boldsymbol{\theta}'}, \end{aligned}$$

$$\frac{\partial^2 D}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \sum_{j=1}^{l}\sum_{k=1}^{l} e_j^l \otimes (e_k^l)' \otimes \sum_{i=1}^{M} v_i\left\{tr\left(\mathbf{V}_i^{-1}\frac{\partial^2 \mathbf{V}_i}{\partial \theta_j \partial \theta_k} - \mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\right)\right\},$$

$$\begin{aligned} \frac{\partial^2 g}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} &= \sum_{j=1}^{l}\sum_{k=1}^{l} e_j^l \otimes (e_k^l)' \otimes \\ &\quad \sum_{i=1}^{M}\left\{-u'(d_i)d_i^{-1}\left(\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right. \\ &\quad \left. -u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\left(\frac{\partial \mathbf{V}_i}{\partial \theta_k}\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j} - \frac{\partial^2 \mathbf{V}_i}{\partial \theta_j \partial \theta_k} + \frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\right)\mathbf{V}_i^{-1/2}\mathbf{r}_i'\right\} \end{aligned}$$

Thus, under assumption **R5,** $\mathbf{\Psi_M}(\mathbf{y},\boldsymbol{\xi})$ is continuously differentiable. Assumptions **R1-R6** and above arguments imply that conditions 6-8 in Yuan and Jennrich (1998) are satisfied, and the main Theorem in Yuan (1997) implies that convergence of $\frac{\partial}{\partial\boldsymbol{\xi}}\mathbf{\Psi_M}(\mathbf{y},\boldsymbol{\xi})$ to $\Phi(\boldsymbol{\xi})$ is uniform. Hence condition **C2** is satisfied as well. Then existence of a consistent sequence of GS-estimators follows from the main result of Yuan and Jennrich (1998). □

3 Proof of Theorem 3.

Under assumptions **R1-R7**, the multivariate Lindenberg-Feller Centeral Limit Theorem (e.g. Mukhopadhyay, 2009) gives that $M^{1/2}\mathbf{\Psi}_M(\mathbf{y},\boldsymbol{\xi}_\mathbf{0}) \to \mathbf{N}\left(0,\mathbf{\Omega}(\boldsymbol{\xi}_0)\right)$ in distribution if the Lindenberg's condition is satisfied. The proof of Theorem 5.14 in Shao (2003) implies that the Lindenberg's condition follows if $\sup_i E\left\|\mathbf{\Psi_i}(\mathbf{y}_i,\boldsymbol{\xi}_\mathbf{0})\right\|^{2+\delta} < \infty$ for some $\delta > 0$. The latter condition is satisfied under the assumptions **R1-R4**. Thus, assumptions 1-3 in Yuan and Jennrich (1998) are satisfied, and Theorem 4 in Yuan and Jennrich (1998) implies that $M^{-1/2}\left(\widehat{\boldsymbol{\xi}}_M \to \boldsymbol{\xi}_0\right) \to \mathbf{N}\left(0,\mathbf{\Sigma}\right)$ where $\mathbf{\Sigma} = \left[\Phi(\boldsymbol{\xi}_0)\right]^{-1}\mathbf{\Omega}(\boldsymbol{\xi}_0)\left[\Phi(\boldsymbol{\xi}_0)'\right]^{-1}$,

$\mathbf{\Omega}\left(\boldsymbol{\xi}_\mathbf{0}\right) = \lim_{M\to\infty}\left\{\mathbf{\Omega}_M(\boldsymbol{\xi}_0)\right\}$, $\mathbf{\Omega}_M(\boldsymbol{\xi}_0) = \frac{1}{M}\sum_{i=1}^M E\big[\mathbf{\Psi_i}(\mathbf{y_i},\boldsymbol{\xi}_\mathbf{0})\mathbf{\Psi_i}(\mathbf{y_i},\boldsymbol{\xi}_\mathbf{0})'\big]$,

$\Phi(\boldsymbol{\xi}_0) = \lim_{M\to\infty}\left\{\Phi_M(\boldsymbol{\xi}_0)\right\}$, $\Phi_M(\boldsymbol{\xi}_0) = E\left[\frac{\partial}{\partial\boldsymbol{\xi}}\mathbf{\Psi_M}(\mathbf{y},\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right]$.

First, we compute $\Phi_M(\boldsymbol{\xi}_0) = E\left[\frac{\partial}{\partial\boldsymbol{\xi}}\mathbf{\Psi_M}(\mathbf{y},\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right]$. Under the true model parameters $\boldsymbol{\xi}_0 = \left[\boldsymbol{\beta}_0',\boldsymbol{\theta}_0'\right]'$, vectors $\mathbf{r}_i = \mathbf{V}_i^{-1/2}\left(\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta}\right)$ have multivarite standard normal distribution, and Lemma 5.1 in Lopuhaa (1989) implies that $d_i = \left(\mathbf{r}_i'\mathbf{r}_i\right)^{1/2}$ and $d_i^{-1}\mathbf{r}_i$ are independent, $E_{\mathbf{r}_i\sim\mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[d_i^{-1}\mathbf{r}_i\right] = 0$, and $Cov\left(d_i^{-1}\mathbf{r}_i\right) = E_{\mathbf{r}_i\sim\mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})}\left[d_i^{-2}\mathbf{r}_i\mathbf{r}_i'\right] = m_i^{-1}\mathbf{I}$. Hence,

$$E\left[\frac{\partial}{\partial\boldsymbol{\beta}'}\mathbf{\Psi}_{i\boldsymbol{\beta}}(\mathbf{y_i},\boldsymbol{\xi})\right] = -E\left[u(d_i)\right]\mathbf{A}_i'\mathbf{V}_i^{-1}\mathbf{A}_i - E\left[u'(d_i)d_i\right]\mathbf{A}_i'\mathbf{V}_i^{-1/2}\left(m_i^{-1}\mathbf{I}\right)\mathbf{V}_i^{-1/2}\mathbf{A}_i,$$

so that $E\left[\frac{\partial}{\partial\boldsymbol{\beta}'}\mathbf{\Psi}_\mathbf{M}^{\boldsymbol{\beta}}(\mathbf{y},\boldsymbol{\xi})\right] = -\frac{1}{M}\sum\limits_{i=1}^{M} E\left[u(d_i) + m_i^{-1}u'(d_i)d_i\right]\mathbf{A}_i'\mathbf{V}_i^{-1}\mathbf{A}_i$.

Further, $E\left[\frac{\partial}{\partial\theta_j}\mathbf{\Psi}_{\mathbf{i}\boldsymbol{\beta}}(\mathbf{y_i},\boldsymbol{\xi})\right] = \mathbf{0}$ since $\frac{\partial}{\partial\theta_j}\mathbf{\Psi}_{\mathbf{i}\boldsymbol{\beta}}(\mathbf{y_i},\boldsymbol{\xi})$ is asymmetric function in each component of $\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})$. Similarly, $E\left[\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}_\mathbf{j}}(\mathbf{y_i},\boldsymbol{\xi})\right] = \mathbf{0}$ since $\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}_\mathbf{j}}(\mathbf{r_i},\mathbf{A_i},\boldsymbol{\xi})$ is asymmetric function in each component of $\mathbf{r}_i \sim \mathbf{N}_{m_i}(\mathbf{0},\mathbf{I})$.

Finally, $E\left[\frac{\partial}{\partial\boldsymbol{\theta}'}M\boldsymbol{\Psi}_{\boldsymbol{\theta}}(\mathbf{y},\boldsymbol{\xi})\right] = \left(\mathbf{1}'_{l\times 1}E\left[\frac{\partial g}{\partial\boldsymbol{\theta}}\right]\right)\frac{\partial^2 D}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial D}{\partial\boldsymbol{\theta}}\left(\mathbf{1}'_{l\times 1}E\left[\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right)$

$-\left(\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial\boldsymbol{\theta}}\right)E\left[\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] - E\left[\frac{\partial g}{\partial\boldsymbol{\theta}}\right]\left(\mathbf{1}'_{l\times 1}\frac{\partial^2 D}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) + \mathbf{c}E\left[\frac{\partial g}{\partial\boldsymbol{\theta}'}\right]$,

where $E\left[\frac{\partial g}{\partial\boldsymbol{\theta}}\right] = \frac{\partial D}{\partial\boldsymbol{\theta}}$ because $v_i = E_{d^2\sim\chi^2_{m_i}}\left[d^2u(d)\right]m_i^{-1}$. Hence,

$E\left[\frac{\partial}{\partial\boldsymbol{\theta}'}M\boldsymbol{\Psi}_{\boldsymbol{\theta}}(\mathbf{y},\boldsymbol{\xi})\right] = -\left(\mathbf{1}'_{l\times 1}\frac{\partial D}{\partial\boldsymbol{\theta}}\mathbf{I} - \frac{\partial D}{\partial\boldsymbol{\theta}}\mathbf{1}'_{l\times 1}\right)\left(E\left[\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] - \frac{\partial^2 D}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) + \mathbf{c}\frac{\partial D}{\partial\boldsymbol{\theta}'}$,

where $E\left[\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] =$

$E\left[\sum_{j=1}^{l}\sum_{k=1}^{l}e_j^l\otimes(e_k^l)'\otimes\sum_{i=1}^{M}\left\{-u'(d_i)d_i^{-1}\left(\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right.\right.$

$\left.\left.-u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\left(\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_j} - \frac{\partial^2\mathbf{V}_i}{\partial\theta_j\partial\theta_k} + \frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\right)\mathbf{V}_i^{-1/2}\mathbf{r}_i'\right\}\right]$.

Again using Lemma 5.1 in Lopuhaa (1989),

$E\left[u'(d_i)d_i^{-1}\left(\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-1/2}\mathbf{r}_i\right] =$

$\frac{E\left[u'(d_i)d_i^3\right]}{m_i(m_i+2)}tr\left(\left[\left(\mathbf{V}_i^{-\frac{1}{2}}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-\frac{1}{2}}\right)'\otimes\left(\mathbf{V}_i^{-\frac{1}{2}}\frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-\frac{1}{2}}\right)\right]\mathbf{H}_{m_i}\right)$,

where $\mathbf{H}_{m_i} = \left[\mathbf{I} + \mathbf{K}_{m_i,m_i} + vec(\mathbf{I})\,vec(\mathbf{I})'\right]$, $\mathbf{K}_{m_i,m_i}$ is a $\left(m_i^2\times m_i^2\right)$-block matrix with (i,j)-th block $K(i,j)$ of size $(m_i\times m_i)$,which has 1 at entry (i,j) and 0 otherwise ($(K(i,j))_{ij} = \delta_{ij}$). Also, with the arguments similar to the above,

$E\left[u(d_i)\mathbf{r}_i'\mathbf{V}_i^{-1/2}\left(\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_j} - \frac{\partial^2\mathbf{V}_i}{\partial\theta_j\partial\theta_k} + \frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\right)\mathbf{V}_i^{-1/2}\mathbf{r}_i'\right] =$

$m_i^{-1}E\left[u(d_i)d_i^2\right]tr\left[\mathbf{V}_i^{-1}\left(\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_j} - \frac{\partial^2\mathbf{V}_i}{\partial\theta_j\partial\theta_k} + \frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\right)\right]$, so that,

$$\begin{aligned}
E\left[\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] &= \sum_{j=1}^{l}\sum_{k=1}^{l}e_j^l\otimes(e_k^l)'\otimes\sum_{i=1}^{M}\left\{-\frac{E\left[u'(d_i)d_i^3\right]}{m_i(m_i+2)}\times\right. \\
&\left. tr\left(\left[\left(\mathbf{V}_i^{-\frac{1}{2}}\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-\frac{1}{2}}\right)'\otimes\left(\mathbf{V}_i^{-\frac{1}{2}}\frac{\partial\mathbf{V}_i}{\partial\theta_j}\mathbf{V}_i^{-\frac{1}{2}}\right)\right]\mathbf{H}_{m_i}\right)\right\} \\
&+v_i tr\left[\mathbf{V}_i^{-1}\left(-2\frac{\partial\mathbf{V}_i}{\partial\theta_k}\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_j} + \frac{\partial^2\mathbf{V}_i}{\partial\theta_j\partial\theta_k}\right)\right]
\end{aligned}$$

Hence, $E\left[\frac{\partial}{\partial \boldsymbol{\xi}} \boldsymbol{\Psi}_{\mathbf{M}}(\mathbf{y},\boldsymbol{\xi_0})|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right] = Diag\left(\Phi_{\boldsymbol{\beta\beta}},\Phi_{\boldsymbol{\theta\theta}}\right)$, where

$$\Phi_{\boldsymbol{\beta\beta}} = -\frac{1}{M}\sum_{i=1}^{M} \mathbf{A}_i'\mathbf{V}_i^{-1}\mathbf{A}_i E\left[u(d_i) + m_i^{-1}u'(d_i)d_i\right]$$

$$\Phi_{\boldsymbol{\theta\theta}} = -\frac{1}{M}\left(\mathbf{1}_{l\times 1}'\left\{\sum_{j=1}^{l} e_j^l \otimes \sum_{i=1}^{M} v_i tr\left[\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\right]\right\}\mathbf{I} - \frac{\partial D}{\partial \boldsymbol{\theta}}\mathbf{1}_{l\times 1}'\right)\times$$

$$\sum_{j=1}^{l}\sum_{k=1}^{l} e_j^l \otimes (e_k^l)' \otimes \sum_{i=1}^{M}\left\{\frac{E\left[u'(d_i)d_i^3\right]}{m_i\left(m_i+2\right)} tr\left(\left[\left(\mathbf{V}_i^{-\frac{1}{2}}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\mathbf{V}_i^{-\frac{1}{2}}\right)' \otimes \left(\mathbf{V}_i^{-\frac{1}{2}}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-\frac{1}{2}}\right)\right]\mathbf{H}_{m_i}\right)\right.$$

$$\left. + v_i tr\left(\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_k}\right)\right\} + \frac{1}{M}\mathbf{c}\sum_{j=1}^{l}\left(e_j^l\right)' \otimes \sum_{i=1}^{M} v_i tr\left[\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \theta_j}\right].$$

Further, $\boldsymbol{\Omega}_M(\boldsymbol{\xi}_0) = \frac{1}{M}\sum_{i=1}^{M} E\left[\boldsymbol{\Psi}_{\mathbf{i}}(\mathbf{y_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}}(\mathbf{y_i},\boldsymbol{\xi_0})'\right]$ may be written as

$$\boldsymbol{\Omega}_M(\boldsymbol{\xi}_0) = \left[\begin{array}{cc} \frac{1}{M}\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] & \frac{1}{M}\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] \\ \frac{1}{M}\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] & \frac{1}{M}\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] \end{array}\right],$$

where $\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] = \sum_{\mathbf{i=1}}^{\mathbf{M}} m_i^{-1}E\left[u^2(d_i)d_i^2\right]\mathbf{A}_i'\mathbf{V}_i^{-1}\mathbf{A}_i$ and $\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] = \mathbf{0}$ because $\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\mathbf{A_i},\boldsymbol{\xi})$ is asymmetric and $\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\mathbf{A_i},\boldsymbol{\xi})'$ is symmetric function in each component of $\mathbf{r}_i$.

Similarly, $\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\beta}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})'\right] = \mathbf{0}$.

Finally, $\boldsymbol{\Omega}_M^{\boldsymbol{\theta}}(\boldsymbol{\xi}_0) = \sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})\boldsymbol{\Psi}_{\mathbf{i}\boldsymbol{\theta}}(\mathbf{r_i},\boldsymbol{\xi_0})'\right]$ has elements

$$\left[\boldsymbol{\Omega}_M^{\boldsymbol{\theta}}(\boldsymbol{\xi}_0)\right]_{jk} = E\left[\sum_{\mathbf{i=1}}^{\mathbf{M}} u^2(d_i)\left(\mathbf{r}_i'\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\left(\mathbf{r}_i'\mathbf{V}_{\mathbf{i}}^{-\frac{1}{2}}\mathbf{F}_{ki}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\right]$$

$$+E\left[c_j\sum_{\mathbf{i=1}}^{\mathbf{M}} u(d_i)\left[\rho\left(d_i\right) - b_{m_i}\right]\left(\mathbf{r}_i'\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\right] +$$

$$+E\left[c_k\sum_{\mathbf{i=1}}^{\mathbf{M}} u(d_i)\left[\rho\left(d_i\right) - b_{m_i}\right]\left(\mathbf{r}_i'\mathbf{V}_{\mathbf{i}}^{-\frac{1}{2}}\mathbf{F}_{ki}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\right] +$$

$$+E\left[c_jc_k\sum_{\mathbf{i=1}}^{\mathbf{M}}\left[\rho\left(d_i\right) - b_{m_i}\right]^2\right],$$ where $\mathbf{F}_{ji} = \mathbf{F}_{ji}\left(\boldsymbol{\theta}\right)$ is given in (2.13).

Using results from Neudecker (1969) and Lemma 5.1 in Lopuhaa (1989)),

$E\left[\sum_{\mathbf{i=1}}^{\mathbf{M}} u^2(d_i)\left(\mathbf{r}_i'\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\left(\mathbf{r}_i'\mathbf{V_i}^{-\frac{1}{2}}\mathbf{F}_{ki}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\right] =$

$\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[u^2(d_i)d_i^4\right] E\left[vec\left(\frac{\mathbf{r}_i\mathbf{r}_i'}{d_i^2}\right)'\left\{\left(\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ki}\mathbf{V}_i^{-\frac{1}{2}}\right)'\otimes\left(\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\right)\right\} vec\left(\frac{\mathbf{r}_i\mathbf{r}_i'}{d_i^2}\right)\right] =$

$\sum_{\mathbf{i=1}}^{\mathbf{M}} \frac{E\left[u^2(d_i)d_i^4\right]}{m_i(m_i+2)} tr\left\{\left[\left(\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ki}\mathbf{V}_i^{-\frac{1}{2}}\right)'\otimes\left(\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\right)\right]\left[\mathbf{I}+\mathbf{K}_{m_i}+vec\left(\mathbf{I}\right)vec\left(\mathbf{I}\right)'\right]\right\},$

and $E\left[\sum_{\mathbf{i=1}}^{\mathbf{M}} u(d_i)\left[\rho\left(d_i\right)-b_{m_i}\right]\left(\mathbf{r}_i'\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\mathbf{r}_i\right)\right] =$

$\sum_{\mathbf{i=1}}^{\mathbf{M}} m_i^{-1}E\left[u(d_i)d_i^2\left(\rho\left(d_i\right)-b_{m_i}\right)\right] tr\left[\mathbf{V}_i^{-1}\mathbf{F}_{ji}\right]$, so that

$$\begin{aligned}\left[\mathbf{\Omega}_M(\boldsymbol{\xi}_0)\right]_{jk} &= c_jc_k\sum_{\mathbf{i=1}}^{\mathbf{M}} E\left[\left(\rho\left(d_i\right)-b_{m_i}\right)^2\right]\\ &+\sum_{\mathbf{i=1}}^{\mathbf{M}} m_i^{-1}E\left[u(d_i)d_i^2\left(\rho\left(d_i\right)-b_{m_i}\right)\right] tr\left[\mathbf{V}_i^{-1}\left(c_j\mathbf{F}_{ji}+c_k\mathbf{F}_{ki}\right)\right]\\ &+\sum_{\mathbf{i=1}}^{\mathbf{M}} \frac{E\left[u^2(d_i)d_i^4\right]}{m_i\left(m_i+2\right)} tr\left\{\left[\left(\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ki}\mathbf{V}_i^{-\frac{1}{2}}\right)'\otimes\left(\mathbf{V}_i^{-\frac{1}{2}}\mathbf{F}_{ji}\mathbf{V}_i^{-\frac{1}{2}}\right)\right]\mathbf{H}_{m_i}\right\}.\square\end{aligned}$$

4 Derivation of fixed point equations for $\widehat{\boldsymbol{\theta}}_M$.

Linear covariance structure of $\mathbf{V}_i$ ($\mathbf{V}_i = \sum_{g=1}^{l}\theta_g\mathbf{L}_{ig}$, $i = 1, ..., M$) implies that $\forall i$, $\frac{\partial\mathbf{V}_i}{\partial\theta_j} = \mathbf{L}_{ij}$ and $\sum_{j=1}^{l}\theta_j\frac{\partial\mathbf{V}_i}{\partial\theta_j} = \mathbf{V}_i$ so that

$$tr\left(\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_j}\right) = tr\left(\mathbf{V}_i^{-1}\mathbf{L}_{ij}\mathbf{V}_i^{-1}\mathbf{V}_i\right) = \sum_{s=1}^{l}\theta_s tr\left[\mathbf{V}_i^{-1}\mathbf{L}_{ij}\mathbf{V}_i^{-1}\mathbf{L}_{is}\right]$$

Hence, $tr\left(\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\boldsymbol{\theta}}\right) = \mathbf{Q}_i(\boldsymbol{\theta})\boldsymbol{\theta}$, where $\mathbf{Q}_i(\boldsymbol{\theta})$ the $l\times l$ matrix with entries $\left[\mathbf{Q}_i(\boldsymbol{\theta})\right]_{jk} = tr\left[\mathbf{V}_i^{-1}\mathbf{L}_{ij}\mathbf{V}_i^{-1}\mathbf{L}_{ik}\right]$. Respectively, $tr\left(\mathbf{V}_i^{-1}\frac{\partial\mathbf{V}_i}{\partial\theta_j}\right) = \left[\mathbf{Q}_i(\boldsymbol{\theta})\right]_{j.}\boldsymbol{\theta}$, where $\left[\mathbf{Q}_i(\boldsymbol{\theta})\right]_{j.}$ stands for j^{th} row of matrix $\mathbf{Q}_i(\boldsymbol{\theta})$.

Under the assumption of linear covariance structure, $\mathbf{V}_i = \sum_{g=1}^{l}\theta_g\mathbf{L}_{ig}$, $i = 1, ..., M$, each matrix $\mathbf{F}_{ji}$ in (2.13) simplifies to

$\mathbf{F}_{ji} = \left(\sum_{s=1}^{l}\mathbf{L}_{is}\right)\sum_{k=1}^{M} v_k tr\left(\mathbf{V}_k^{-1}\frac{\partial\mathbf{V}_k}{\partial\theta_j}\right) - \mathbf{L}_{ij}\sum_{k=1}^{M} v_k tr\left(\mathbf{V}_k^{-1}\sum_{s=1}^{l}\mathbf{L}_{is}\right)$, where

$\sum_{k=1}^{M} v_k tr \left(\mathbf{V}_k^{-1} \sum_{s=1}^{l} \mathbf{L}_{is} \right)$ is a constant further denoted by S_{Ml},

$$S_{Ml} = \sum_{k=1}^{M} v_k tr \left(\mathbf{V}_k^{-1} \sum_{s=1}^{l} \frac{\partial \mathbf{V}_k}{\partial \theta_s} \right) = \sum_{k=1}^{M} v_k tr \left(\mathbf{V}_k^{-1} \sum_{s=1}^{l} \mathbf{L}_{is} \right).$$

Hence, $\mathbf{F}_{ji}$ may be expressed as

$$\mathbf{F}_{ji} = \left(\sum_{s=1}^{l} \mathbf{L}_{is} \right) \sum_{k=1}^{M} v_k \left[\mathbf{Q}_k(\boldsymbol{\theta}) \right]_{j.} \boldsymbol{\theta} - S_{Ml} \mathbf{L}_{ij}.$$

Let us further denote by $\mathbf{Q}(\boldsymbol{\theta})$ the $l \times l$ matrix with entries

$$\left[\mathbf{Q}(\boldsymbol{\theta}) \right]_{jk} = \sum_{i=1}^{M} v_i tr \left[\mathbf{V}_i^{-1} \mathbf{L}_{ij} \mathbf{V}_i^{-1} \mathbf{L}_{ik} \right]$$

Then $\mathbf{F}_{ji} = \left(\sum_{s=1}^{l} \mathbf{L}_{is} \right) \left[\mathbf{Q}(\boldsymbol{\theta}) \right]_{j.} \boldsymbol{\theta} - S_{Ml} \mathbf{L}_{ij}$, where $\left[\mathbf{Q}(\boldsymbol{\theta}) \right]_{j.}$ stands for j^{th} row of matrix $\mathbf{Q}(\boldsymbol{\theta})$, and system (3.2) is written in the matrix form as

$$\frac{1}{M} \sum_{j=1}^{l} e_j^l \otimes \sum_{i=1}^{M} \left[u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \left(\left(\sum_{s=1}^{l} \mathbf{L}_{is} \right) \left[\mathbf{Q}(\boldsymbol{\theta}) \right]_{j.} \boldsymbol{\theta} - S_{Ml} \mathbf{L}_{ij} \right) \mathbf{V}_i^{-1/2} \mathbf{r}_i + c_j \left(\rho \left(d_i \right) - b_{m_i} \right) \right] =$$

$$\frac{1}{M} \sum_{j=1}^{l} e_j^l \otimes \sum_{i=1}^{M} \left[u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \left(\left(\sum_{s=1}^{l} \mathbf{L}_{is} \right) \left[\mathbf{Q}(\boldsymbol{\theta}) \right]_{j.} \boldsymbol{\theta} \right) \mathbf{V}_i^{-1/2} \mathbf{r}_i \right] +$$

$$\frac{S_{Ml}}{M} \sum_{j=1}^{l} e_j^l \otimes \sum_{i=1}^{M} \left[u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \mathbf{L}_{ij} \mathbf{V}_i^{-1/2} \mathbf{r}_i \right] + \frac{1}{M} \sum_{j=1}^{l} e_j^l \otimes c_j \sum_{i=1}^{M} \left[\left(\rho \left(d_i \right) - b_{m_i} \right) \right] = \mathbf{0},$$

where $\sum_{i=1}^{M} \left[\left(\rho \left(d_i \right) - b_{m_i} \right) \right] = \sum_{i=1}^{M} \rho_{m_i}^{\epsilon} \left(d_i \right) - b_M = g \left(\mathbf{y}, \boldsymbol{\xi} \right),$

the third term is $l \times 1$ vector $M^{-1} g \left(\mathbf{y}, \boldsymbol{\xi} \right) \mathbf{c}$,

the second term is $l \times 1$ vector $\mathbf{U}_M$ with entries

$$U_{Mj} = M^{-1} S_{Ml} \sum_{i=1}^{M} \left[u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \mathbf{L}_{ij} \mathbf{V}_i^{-1/2} \mathbf{r}_i \right],$$

and the first term is written as

$$\frac{1}{M} \sum_{j=1}^{l} e_j^l \otimes \left[\mathbf{Q}(\boldsymbol{\theta}) \right]_{j.} \boldsymbol{\theta} \sum_{i=1}^{M} \left[u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \left(\left(\sum_{s=1}^{l} \mathbf{L}_{is} \right) \right) \mathbf{V}_i^{-1/2} \mathbf{r}_i \right].$$

Let us further denote

$$P_M = M^{-1} \sum_{i=1}^{M} \left[u(d_i)\mathbf{r}_i' \mathbf{V}_i^{-1/2} \left(\sum_{s=1}^{l} \mathbf{L}_{is} \right) \mathbf{V}_i^{-1/2} \mathbf{r}_i \right]$$

then the first term is $l \times 1$ vector $P_M \mathbf{Q}(\boldsymbol{\theta})\boldsymbol{\theta}$ and system (3.2) is re-written as

$$P_M \mathbf{Q}(\boldsymbol{\theta})\boldsymbol{\theta} + \mathbf{U}_M + M^{-1} g\left(\mathbf{y}, \boldsymbol{\xi}\right) \mathbf{c} = \mathbf{0}.$$

References

Kackar, R. N. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Comm. Statist. Theory Methods* **10,** 1249–1261.

Mukhopadhyay, P. (2009) *Multivariate statistical analysis.* New Jersey: World Scientific.

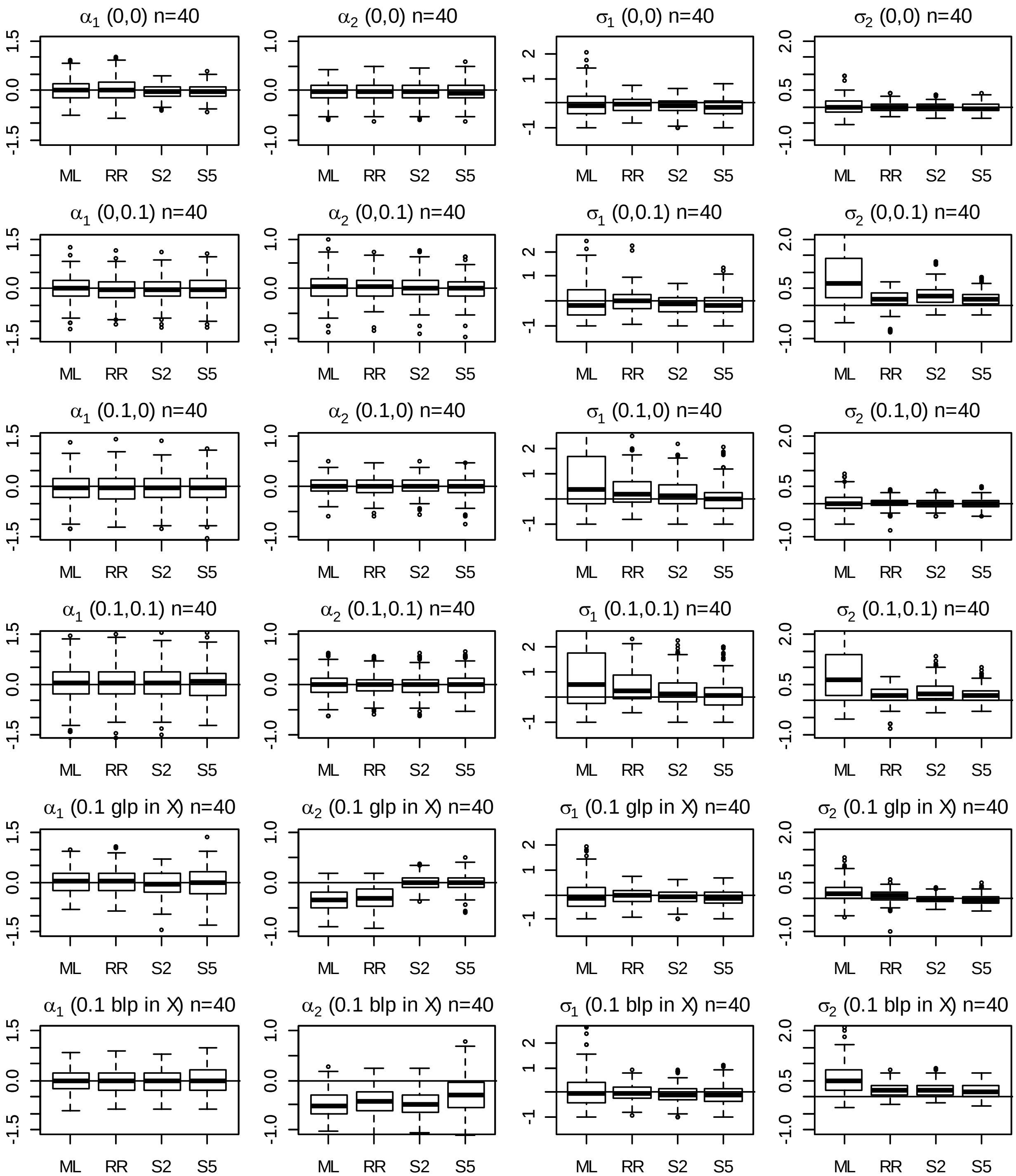
Neudecker, H. (1969) Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association* **64**, 953-963.

Shao, J (2003) *Mathematical Statistics.* New York: Springer.

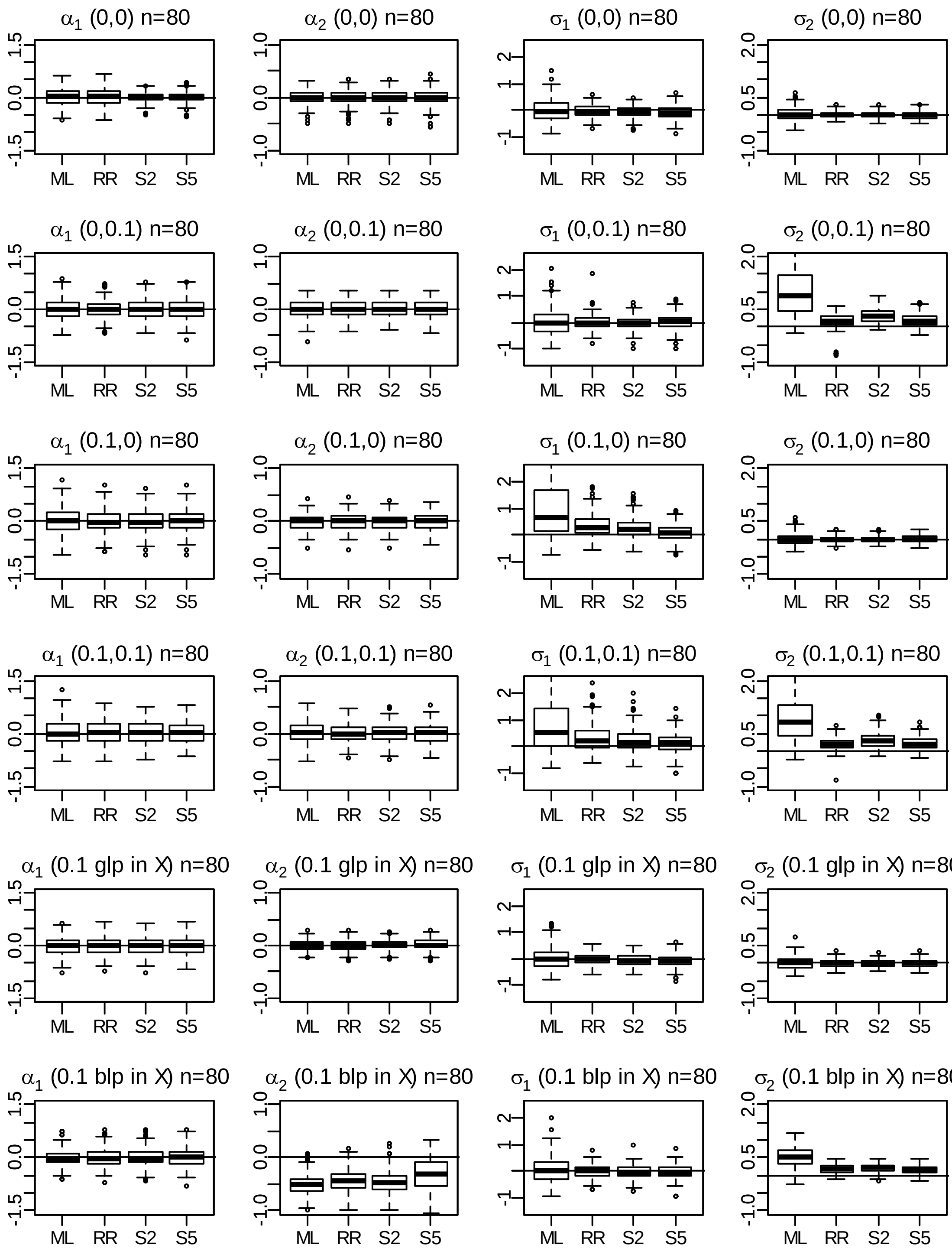
Varadhan, R. and Gilbert, P. (2009). BB: An R Package for Solving a Large System of Nonlinear Equations and for Optimizing a High-Dimensional Nonlinear Objective Function. *Journal of Statistical Software* **32(4)**, 1-26.

Yuan, K.-H. (1997) A theorem of uniform convergence of stochastic functions with applications, *Journal of Multivariate Analysis* **62**, 100-109.

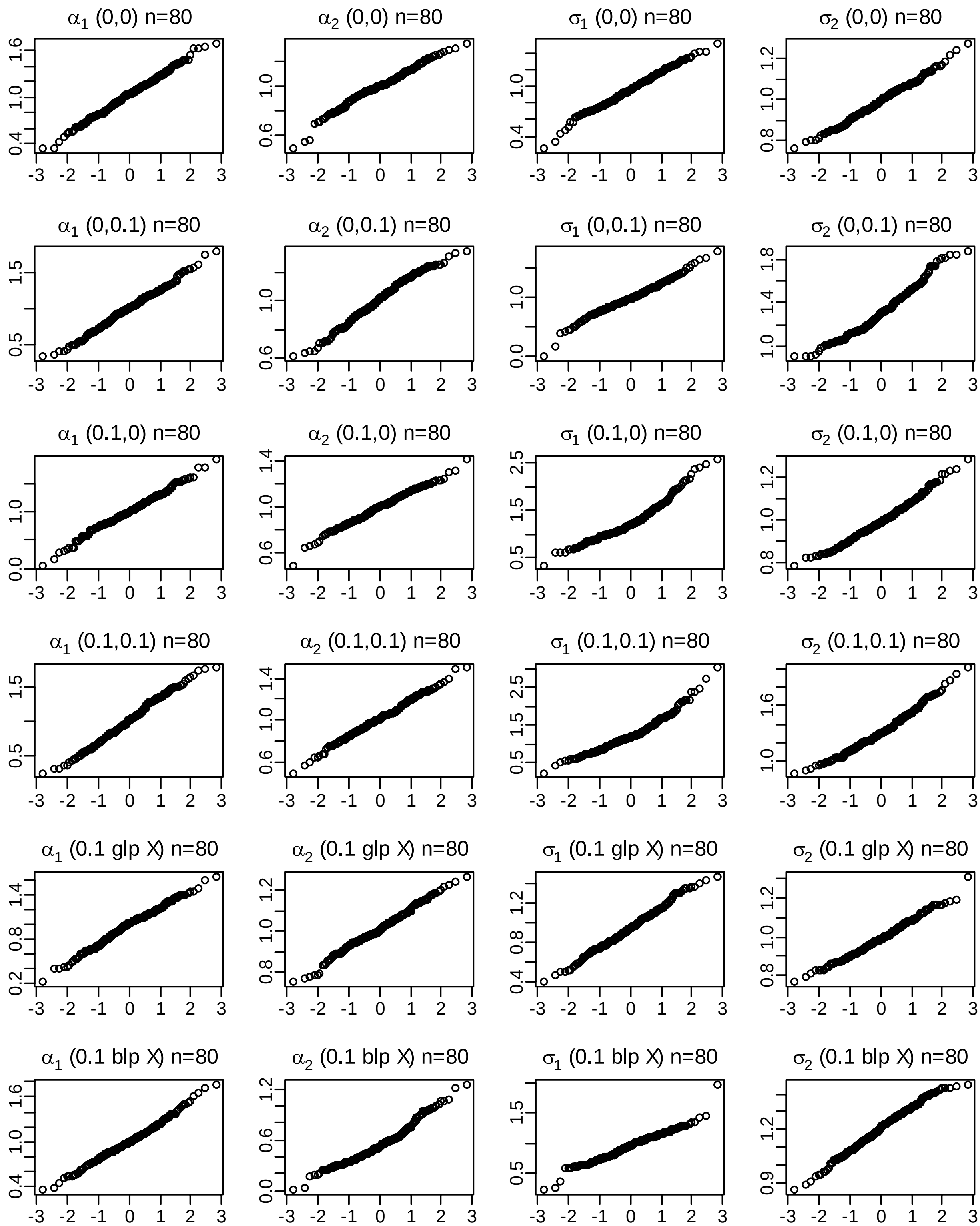
Yuan, K.-H. and Jennrich, R.I. (1998) Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis* **65**, 245-260.



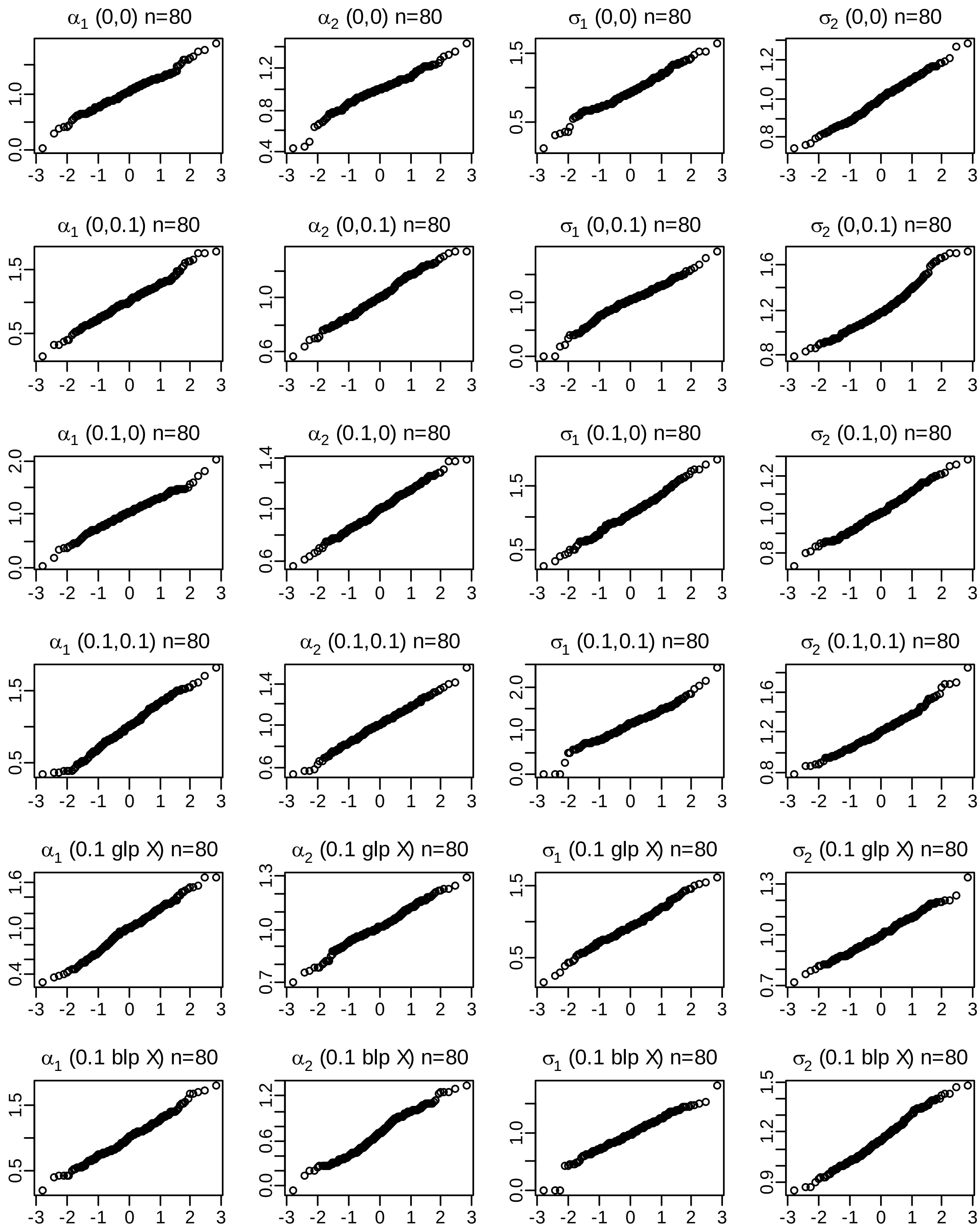
Supplementary Figure 1. Error distribution of parameter estimates for simulated data sets with n=40 without contamination (0,0), with 10% contamination in error terms (0,0.1), with 10% contamination in random effects (0.1,0), with 10% contamination in both error terms and random effects (0.1,0.1), with 10% of "good" (0.1 glp in X) and "bad" (0.1 glp in X) leverage points. Four estimation methods used are REML (ML), Robust REML II (RR), GS-estimator with breakdown parameter 0.2 (S2), GS-estimator with breakdown parameter 0.5 (S5).



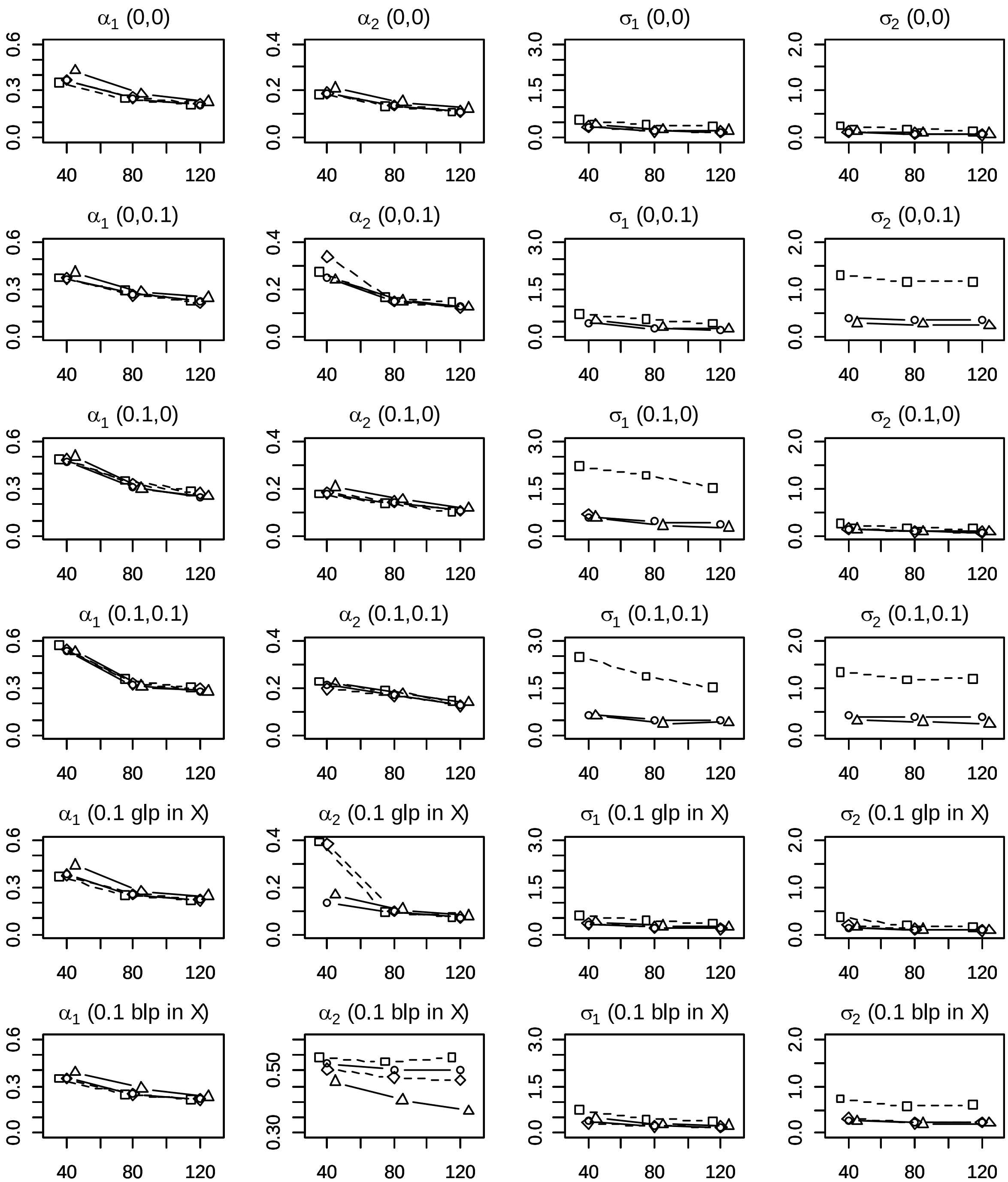
Supplementary Figure 2. Error distribution of parameter estimates for simulated data sets with n=80 without contamination (0,0), with 10% contamination in error terms (0,0.1), with 10% contamination in random effects (0.1,0), with 10% contamination in both error terms and random effects (0.1,0.1), and with 10% of "good" (0.1 glp in X) and "bad" (0.1 glp in X) leverage points. Four estimation methods used are REML (ML), Robust REML II (RR), GS-estimator with breakdown parameter 0.2 (S2), GS-estimator with breakdown parameter 0.5 (S5).



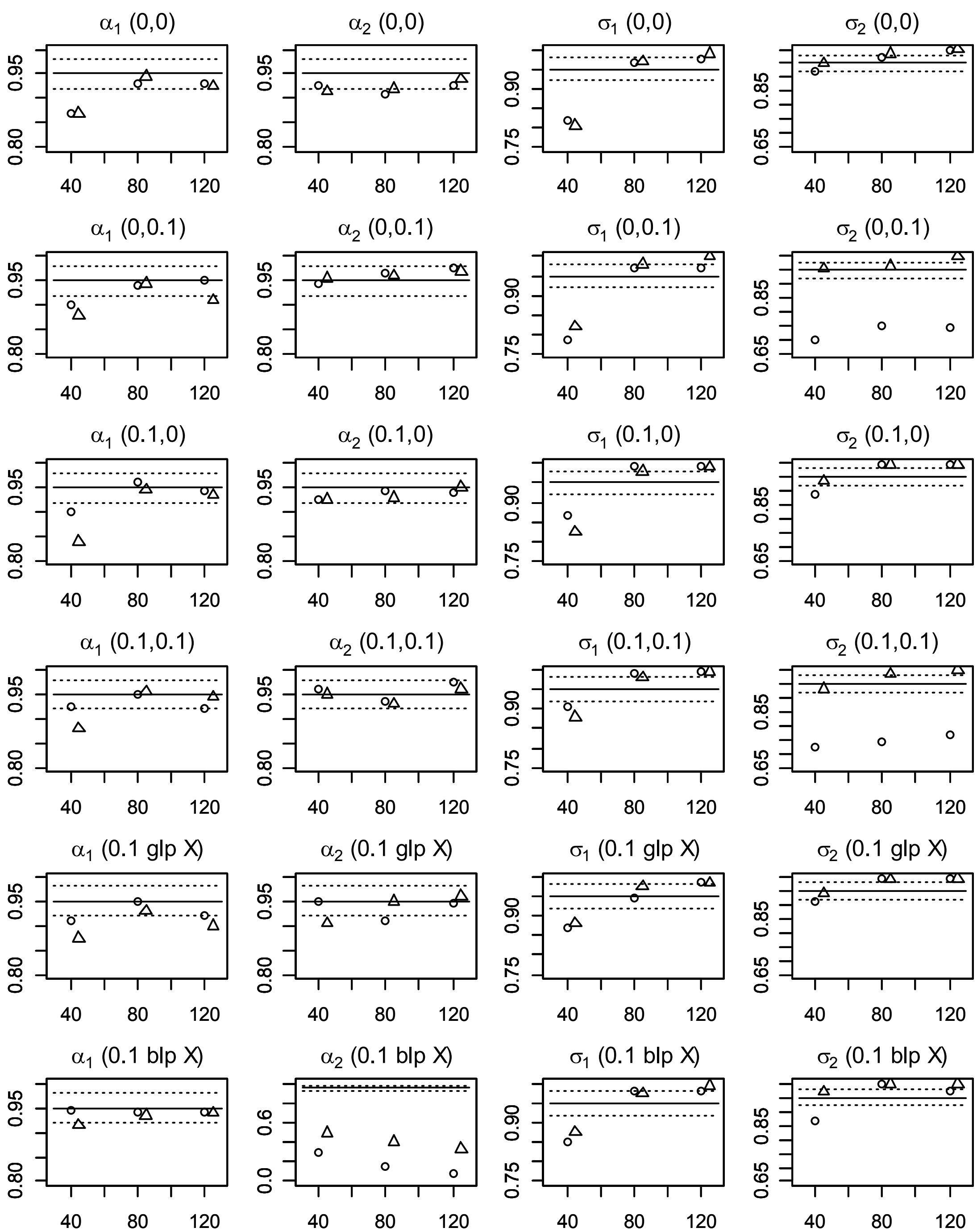
Supplementary Figure 3. QQplots of GS parameter estimates with breakdown parameter 0.2, n=80.



Supplementary Figure 4. QQplots of GS parameter estimates with breakdown parameter 0.5, n=80.



Supplementary Figure 5. Root mean squared errors (RMSE) in simulated data without contamination (0,0), with 10% contamination in error terms (0,0.1), 10% contamination in random effects (0.1,0), 10% contamination in both error terms and random effects (0.1,0.1), and 10% of "good" (0.1 glp in X) and "bad" (0.1 glp in X) leverage points. Estimation methods used are REML (squares, dashed lines), Robust REML II (diamonds, dashed lines), GS-estimator with breakdown parameter 0.2 (circles, solid lines), GS-estimator with breakdown parameter 0.5 (triangles, solid lines). Missing points for Robust REML II correspond to the mean squared errors outside the plotting range.



Supplementary Figure 6. Coverage of asymptotic 95% confidence intervals for sample sizes 40, 80 and 120 in simulated data sets without contamination (0,0), with 10% contamination in error terms (0,0.1), 10% contamination in random effect of group (0.1,0), 10% contamination in both error terms and random effect of group (0.1,0.1), and 10% of "good" (0.1 glp in X) and "bad" (0.1 glp in X) leverage points. Results for GS-estimator with breakdown parameter 0.2 plotted with circles and for GS-estimator with breakdown parameter 0.5 plotted with triangles. Solid lines show nominal 95%, and the dotted lines limit the approximate acceptance region [0.92,0.98] for the null hypothesis H_0: p =0.95.