Likelihood Based Inference for Skew–Normal Independent Linear Mixed Models

Supplemental Material

V. H. Lachos, Pulak Ghosh, and R. B. Arellano-Vall

Universidade Estadual de Campinas, Brazil, Indian Institute of Management, Bangalore, India, and Pontificia Universidad Católica de Chile, Chile

This note contains the proofs of Lemma 1, 2, proposition 1. In appendix b, we have the restricted estimation, appendix C, has the observed information matrix, and appendix D contains the additional simulation study results.

Appendix A: Some Lemmas and Proofs

Lemma 1. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *and* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 $(p_1 + p_2 = p)$*, respectively; let* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *and*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

be the corresponding partitions of $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$*. Then,* **marginally** $\mathbf{Y}_1 \sim SNI_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}}; H)$**, where**

$$\widetilde{\boldsymbol{\upsilon}} = \frac{\boldsymbol{\upsilon}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}},$$

with $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$*,* $\boldsymbol{\upsilon} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda} = (\boldsymbol{\upsilon}_1^\top, \boldsymbol{\upsilon}_2^\top)^\top$.

Proof. See Proposition 5.4 and 5.5 in Branco and Dey (2001). □

Lemma 2. *Let* $\mathbf{Y} \in \mathbb{R}^p$ *a random vector with the following extended skew-normal pdf*

$$f(\mathbf{y}) = \frac{1}{\Phi\left(c/\sqrt{1 + \mathbf{a}^\top \boldsymbol{\Sigma}\mathbf{a}}\right)}\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(c + \mathbf{a}^\top(\mathbf{y} - \boldsymbol{\mu})),$$

where c is A fixed real constant and $\mathbf{a}$ *is a fixed p-dimensional vector. Then,*

$$E\{\mathbf{Y}\} = \boldsymbol{\mu} + \boldsymbol{\delta} W_{\Phi_1}(\bar{c}),$$

where $\boldsymbol{\delta} = \boldsymbol{\Sigma}\mathbf{a}/\sqrt{1+\mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}}$ *and* $\bar{c} = c/\sqrt{1+\mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}}$*.*

Proof. It follows by using the well known stochastic representation $\mathbf{Y} \overset{d}{=} \mu + \boldsymbol{\delta} Z_{(-\bar{c},\infty)} + \mathbf{Z}$, where $Z_{(-\bar{c},\infty)} \overset{d}{=} (Z_0|Z_0 + \bar{c} > 0)$, with $Z_0 \sim N(0,1)$ and being independent of $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)$ (see e.g. Arellano-Valle and Azzalini, 2006), and the fact that $\mathrm{E}\{Z_{(-\bar{c},\infty)}\} = W_{\Phi_1}(\bar{c})$ (see Johnson et al., 1994, Section 10.1). □

Proof of Proposition 1: Consider the case where U is (absolutely) continues. Let $h(u;\nu)$ be the pdf of U, $h(u|\mathbf{y}) = f(\mathbf{y}|u)h(u;\nu)/f(\mathbf{y})$ be the conditional pdf of U given $\mathbf{Y} = \mathbf{y}$, and $h_0(u|\mathbf{y}_0)$ be the conditional pdf of U given $\mathbf{Y}_0 = \mathbf{y}$, where $\mathbf{Y}_0|U = u \sim N_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})$. Since $h(u|\mathbf{y}) = f(\mathbf{y}|u)h(u;\nu)/f(\mathbf{y})$ and $h_0(u|\mathbf{y}) = \phi_p(\mathbf{y};\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})h(u;\nu)/f_0(\mathbf{y})$, where f_0 is the pdf of $\mathbf{Y}_0$, we have for any integrable function $g(u)$ that

$$\begin{aligned}
\mathrm{E}[g(U)|\mathbf{Y}=\mathbf{y}] &= (1/f(\mathbf{y}))\int_0^\infty g(u)f(\mathbf{y}|u)h(u;\nu)du \\
&= (2/f(\mathbf{y}))\int_0^\infty g(u)\phi_p(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})h(u;\nu)\Phi(u^{1/2}\mathrm{A})du \\
&= (2/f(\mathbf{y}))\int_0^\infty g(u)f_0(\mathbf{y}|u)h(u;\nu)\Phi(u^{1/2}\mathrm{A})du \\
&= (2f_0(\mathbf{y})/f(\mathbf{y}))\int_0^\infty g(u)\Phi(u^{1/2}\mathrm{A})h_0(u|\mathbf{y}_0)du.
\end{aligned}$$

Thus, the proof follows by considering the special cases with $g(u) = u^r$ and $g(u) = u^{r/2}W_\Phi(u^{1/2}\mathrm{A})$. The proof when U is discrete is analogous if we replace $\int$ by $\sum$. □

Appendix B:

- **Restricted estimation**

As in Lachos et al. (2007), suppose that our interest centers in estimating the parameter vector $\boldsymbol{\beta}$ under k linearly independent restrictions defined as $\mathbf{C}_j^\top\boldsymbol{\beta} - c_j = 0$, where the $\mathbf{C}_j,\ j = 1,\ldots,k$, are $p\times 1$ vectors and $c_j,\ j = 1,\ldots,k$, are scalars which are assumed to be both known. The interest is to maximize the complete log-likelihood function $\ell_c(\boldsymbol{\theta};\mathbf{y}_c)$ subject to the linear constraints $\mathbf{C}\boldsymbol{\beta}-\mathbf{c}=\mathbf{0}$, where $\mathbf{C}=(\mathbf{C}_1,\ldots,\mathbf{C}_k)^\top$ and $\mathbf{c}=(c_1,\ldots,c_k)^\top$. Following Lachos et al. (2007), one can show that the equality restricted estimate of $\boldsymbol{\beta}$ is given by

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_c^{(k+1)} &= \widehat{\boldsymbol{\beta}}^{(k)}+(\sum_{i=1}^{n}\widehat{u}_i^{(k)}\mathbf{X}_i^\top[\widehat{\boldsymbol{\Sigma}_i}^{(k)}]^{-1}\mathbf{X}_i)^{-1}\\ &\quad\times\mathbf{C}^\top[\mathbf{C}(\sum_{i=1}^{n}\widehat{u}_i^{(k)}\mathbf{X}_i^\top[\widehat{\boldsymbol{\Sigma}_i}^{(k)}]^{-1}\mathbf{X}_i)^{-1}\mathbf{C}^\top]^{-1}(\mathbf{c}-\mathbf{C}\widehat{\boldsymbol{\beta}}^{(k)}),\end{aligned}$$

for $k=0,1,\ldots,$ where $(\widehat{u}_i^{(k)},\widehat{(u\mathbf{b})}_i^{(k)})$, $i=1,\ldots,n$, and $\widehat{\boldsymbol{\beta}}^{(k)}$ are obtained from (**??**) and (**??**), respectively. The EM-algorithm for estimating the parameters of the model (**??**) and (**??**) under the restriction $\mathbf{C}\boldsymbol{\beta}=\mathbf{d}$, denoted by $\widetilde{\boldsymbol{\theta}}_0$, follows the same procedures of the EM-algorithm, replacing $\widehat{\boldsymbol{\beta}}^{(k)}$ by $\widetilde{\boldsymbol{\beta}}_c$ in the M-step.

- **Generating of truncated gamma distribution**

Suppose that $X\sim Gamma(\alpha,\beta)$. To draw X such that $a<X<b$

- Generate $U\sim U[0,1]$
- Let $V=F(a)+[F(b)-F(a)]U$
- return $X=F^{-1}(V)$,

where F is the cdf of the $Gamma(\alpha,\beta)$ distribution. In R software to generate of $0<X<1$, we use the following code:

$$X=qgamma((pgamma(1,\alpha,\beta))*runif(1),\alpha,\beta)$$

- **Derivation of the conditional expectation in the EM algorithm**

The notation used is that of Section 3. From (**??**) we have that

$$\begin{aligned}f(\mathbf{y}_i,\mathbf{b}_i,t_i,u_i|\boldsymbol{\theta}) &= f(\mathbf{y}_i|\mathbf{b}_i,u_i,\boldsymbol{\theta})f(\mathbf{b}_i|t_i,u_i,\boldsymbol{\theta})f(t_i|\boldsymbol{\theta})f(u_i|\boldsymbol{\theta})\\ &= f(\mathbf{y}_i|\boldsymbol{\theta})f(u_i|\mathbf{y}_i,\boldsymbol{\theta})f(t_i|u_i,\mathbf{y}_i,\boldsymbol{\theta})f(\mathbf{b}_i|t_i,u_i,\mathbf{y}_i,\boldsymbol{\theta}),\end{aligned}$$

by using successively Lemma 3 in Arellano–Valle et al. (2005), we can see that

$$\begin{aligned}\mathbf{b}_i|T_i = t_i, U_i = u_i, \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta} &\sim N_q(u_i^{-1/2}\mathbf{s}_i t_i + \mathbf{B}_i\mathbf{Z}_i\mathbf{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), u_i^{-1}\mathbf{B}_i),\\ T_i|U_i = u_i, \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta} &\sim N_1(u_i^{1/2}\mu_i, M_i^2)\mathbb{I}_{(0,\infty)}(t_i),\end{aligned}$$

and since $\mathbf{Y}_i|\boldsymbol{\theta} \sim SNI_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{\Psi}_i, \bar{\boldsymbol{\lambda}}_{b_i}, H)$, then from Section 2 the conditional expectations of $U_i|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}$ are known for the particular SNI distributions considered. Then, all the necessary conditional expectations can be computed by noting that for any integrable function g

$$\mathrm{E}\{U_i^r T_i^s g(\mathbf{b}_i)|\mathbf{Y}_i, \boldsymbol{\theta}\} = \mathrm{E}\{U_i^r \mathrm{E}\{T_i^s \mathrm{E}\{g(\mathbf{b}_i)|T_i, U_i, \mathbf{Y}_i\}|U_i, \mathbf{Y}_i\}|\mathbf{Y}_i, \boldsymbol{\theta}\},$$

$i = 1, \ldots, n$.

Appendix C: The observed information matrix

First we reparameterize $\mathbf{D} = \mathbf{F}^2$ for ease of computation and theoretical derivation, where $\mathbf{F}$ is the square root of $\mathbf{D}$ containing $q(q+1)/2$ distinct elements $\boldsymbol{\alpha}_b = (\alpha_1, \ldots, \alpha_{q(q+1)/2})^\top$. Given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \boldsymbol{\nu}^\top)^\top$, with $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\alpha}_b^\top, \boldsymbol{\lambda}^\top)^\top$ is given by $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$, where

$$\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{n_i}{2} log 2\pi - \frac{1}{2}\log|\mathbf{\Psi}_i| + \log K_i \tag{1}$$

with $K_i(\boldsymbol{\theta}) = \int_0^\infty u_i^{n_i/2}\exp\{-\frac{1}{2}u_i \mathrm{d_i}\}\Phi(\mathrm{u_i^{1/2}A_i})\mathrm{dH}(\mathrm{u_i};\boldsymbol{\nu})$, where $\mathrm{d_i} = (\mathbf{y}_\mathrm{i} - \mathbf{X}_\mathrm{i}\boldsymbol{\beta})^\top\mathbf{\Psi}_\mathrm{i}^{-1}(\mathbf{y}_\mathrm{i} - \mathbf{X}_\mathrm{i}\boldsymbol{\beta})$ and $\mathrm{A}_i = \dfrac{\boldsymbol{\lambda}^\top\mathbf{F}\mathbf{Z}_i\mathbf{\Psi}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{(1+\boldsymbol{\lambda}^\top\mathbf{F}^{-1}\mathbf{\Lambda}_i\mathbf{F}^{-1}\boldsymbol{\lambda})^{1/2}}$. Thus, we have after some algebraic manipulations that the score vector is given by

$$\mathbf{s} = \sum_{i=1}^n \frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\frac{n}{2}\frac{\partial\log|\mathbf{\Psi}_i|}{\partial\boldsymbol{\theta}} - \sum_{i=1}^n \frac{1}{K_i}\frac{\partial K_i}{\partial\boldsymbol{\theta}}, \tag{2}$$

where $\dfrac{\partial K_i}{\partial\boldsymbol{\theta}} = I_i^\phi(\dfrac{n_i+1}{2})\dfrac{\partial\mathrm{A}_i}{\partial\boldsymbol{\theta}} - \dfrac{1}{2}I_i^\Phi(\dfrac{n_i+2}{2})\dfrac{\partial\mathrm{d_i}}{\partial\boldsymbol{\theta}}$, with

$$I_i^\Phi(w) = \int_0^\infty u_i^w\exp\{-\frac{1}{2}u_i\mathrm{d_i}\}\Phi(\mathrm{u_i^{1/2}A_i})\mathrm{dH}(\mathrm{u_i};\boldsymbol{\nu}),$$

$$I_i^\phi(w) = \int_0^\infty u_i^w\exp\{-\frac{1}{2}u_i\mathrm{d_i}\}\phi_1(\sqrt{\mathrm{u_i^{1/2}}}\mathrm{A_i})\mathrm{dH}(\mathrm{u_i};\boldsymbol{\nu}).$$

and $K_i(\boldsymbol{\theta}) = I_i^{\Phi}(\frac{n_i}{2})$. Direct substitution of H in the integrals above yields immediately the following results for each distribution considered, namely

- *Skew–t:*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{2^w \nu^{\nu/2}\Gamma(w+\nu/2)}{\Gamma(\nu/2)(\nu+\mathrm{d_i})^{\nu/2+\mathrm{w}}} T\left(\sqrt{\frac{\nu+2w}{\mathrm{d_i}+\nu}}\,\mathrm{A}_i;\nu+2w\right) \quad \text{and} \\ I_i^{\phi}(w) &= \frac{2^w \nu^{\nu/2}\Gamma(\frac{\nu+2w}{2})}{\sqrt{2\pi}\Gamma(\nu/2)(\mathrm{d_i}+\mathrm{A_i^2}+\nu)^{\frac{\nu+2\mathrm{w}}{2}}}. \end{aligned}$$

- *Skew–slash:*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{2^{2+w}\Gamma(\nu+w)}{\mathrm{d_i^{\nu+w}}} P_1\left(\nu+w,\frac{\mathrm{d_i}}{2}\right) E\{\Phi(S_i^{1/2}\mathrm{A}_i)\} \quad \text{and} \\ I_i^{\phi}(w) &= \frac{\nu 2^{\nu+w}\Gamma(\nu+w)}{\sqrt{2\pi}(\mathrm{d_i}+\mathrm{A_i^2})^{\nu+\mathrm{w}}} P_1\left(\nu+w,\frac{\mathrm{d_i}+\mathrm{A_i^2}}{2}\right), \end{aligned}$$

  where $S_i \sim Gamma(\nu+w,\frac{\mathrm{d_i}}{2})\mathbb{I}_{(0,1)}$.

- *Contaminated skew–normal:*

$$\begin{aligned} I_i^{\Phi}(w) &= \sqrt{2\pi}\{\nu_1\nu_2^{w-1/2}\phi_1(\mathrm{d_i};0,\frac{1}{\nu_2})\Phi(\nu_2^{1/2}\mathrm{A_i}) + (1-\nu_1)\phi_1(\sqrt{\mathrm{d_i}})\Phi(\mathrm{A_i})\} \quad \text{and} \\ I_i^{\phi}(w) &= \nu_1\nu_2^{w-1/2}\phi_1(\sqrt{\mathrm{d_i}+\mathrm{A_i^2}}|0,\nu_2^{-1}) + (1-\nu_1)\phi_1(\sqrt{\mathrm{d_i}+\mathrm{A_i^2}}). \end{aligned}$$

the derivatives of $\log \boldsymbol{\Psi}_i$, d_i, A_i and $K_i(\boldsymbol{\theta})$ involves tedious but not complicated algebraic manipulations and are given next. We consider the common situation in which $\boldsymbol{\Sigma}_i = \sigma_e^2\mathbf{R}_i$

- **For** $\log \boldsymbol{\Psi}_i$**:**

$$\begin{aligned} \frac{\partial \log|\boldsymbol{\Psi}_i|}{\partial \boldsymbol{\tau}} &= \mathbf{0}, \text{ for } \boldsymbol{\tau}=\boldsymbol{\beta},\boldsymbol{\lambda} \text{ and } \boldsymbol{\nu} \\ \frac{\partial \log|\boldsymbol{\Psi}_i|}{\partial \sigma_e^2} &= \mathrm{tr}(\boldsymbol{\Psi}^{-1}\mathbf{R}_i),\ \frac{\partial \log|\boldsymbol{\Psi}_i|}{\partial \alpha_r} = \mathrm{tr}(\boldsymbol{\Psi}^{-1}\mathbf{Z}_i(\dot{\mathbf{F}}_r\mathbf{F}+\mathbf{F}\dot{\mathbf{F}}_r)\mathbf{Z}_i^{\top}), \end{aligned}$$

• **For** A_i**:**

$$
\begin{aligned}
\frac{\partial A_i}{\partial \boldsymbol{\beta}} &= -\frac{1}{a_i}\mathbf{X}_i^\top \boldsymbol{\Psi}_i^{-1}\mathbf{Z}_i\mathbf{F}\boldsymbol{\lambda},\\
\frac{\partial A_i}{\partial \boldsymbol{\lambda}} &= \frac{1}{a_i}\mathbf{F}\mathbf{Z}_i^\top \boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}) - \frac{1}{a_i^2}A_i\mathbf{F}^{-1}\boldsymbol{\Lambda}_i\mathbf{F}^{-1}\boldsymbol{\lambda},\\
\frac{\partial A_i}{\partial \sigma_e^2} &= -\frac{1}{a_i}\boldsymbol{\lambda}^\top\mathbf{F}\mathbf{Z}_i^\top \boldsymbol{\Psi}_i^{-1}\mathbf{R}_i\boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}) - \frac{1}{2\sigma_e^4 a_i^2}A_i\boldsymbol{\lambda}^\top\mathbf{F}^{-1}\boldsymbol{\Lambda}_i\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i\boldsymbol{\Lambda}_i\mathbf{F}^{-1}\boldsymbol{\lambda},\\
\frac{\partial A_i}{\partial \alpha_r} &= \frac{1}{a_i}\boldsymbol{\lambda}^\top\dot{\mathbf{F}}_r\mathbf{Z}_i^\top \boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}) - \frac{1}{a_i^2}\boldsymbol{\lambda}^\top\mathbf{F}\mathbf{Z}_i^\top\boldsymbol{\Psi}_i^{-1}\mathbf{Z}_i(\dot{\mathbf{F}}_r\mathbf{F}+\mathbf{F}\dot{\mathbf{F}}_r)\mathbf{Z}_i^\top\boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta})\\
&- \frac{1}{2a_i^2}A_i\boldsymbol{\lambda}^\top\mathbf{F}^{-1}(-\dot{\mathbf{F}}_r\mathbf{F}^{-1}\boldsymbol{\Lambda}_i - \boldsymbol{\Lambda}_i\mathbf{F}^{-1}\dot{\mathbf{F}}_r - \boldsymbol{\Lambda}_i\mathbf{F}^{-1}(-\dot{\mathbf{F}}_r\mathbf{F}^{-1}-\mathbf{F}^{-1}\dot{\mathbf{F}}_r)\mathbf{F}^{-1}\boldsymbol{\Lambda}_i)\mathbf{F}^{-1}\boldsymbol{\lambda},
\end{aligned}
$$

where $a_i = (1+\boldsymbol{\lambda}^\top\mathbf{F}^{-1}\boldsymbol{\Lambda}_i\mathbf{F}^{-1}\boldsymbol{\lambda})^{1/2}$.

• **For** d_i**:**

$$
\begin{aligned}
\frac{\partial d_i}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}_i^\top\boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}), \quad \frac{\partial d_i}{\partial \boldsymbol{\lambda}} = \mathbf{0},\\
\frac{\partial d_i}{\partial \sigma_e^2} &= -(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\Psi}_i^{-1}\mathbf{R}_i\boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}),\\
\frac{\partial d_i}{\partial \alpha_r} &= -(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\boldsymbol{\Psi}_i^{-1}\mathbf{Z}_i(\dot{\mathbf{F}}_r\mathbf{F}+\mathbf{F}\dot{\mathbf{F}}_r)\mathbf{Z}_i^\top\boldsymbol{\Psi}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}),
\end{aligned}
$$

where $\dot{\mathbf{F}}_r = \dfrac{\partial \mathbf{F}}{\partial \alpha_r}$ and $r = 1,\ldots,q(q+1)/2$.

• **For** $K_i(\boldsymbol{\theta})$**:**

- skew-t

$$
\begin{aligned}
\frac{\partial K_i(\boldsymbol{\theta})}{\partial \nu} &= \frac{1}{2}\{\log\frac{\nu}{2}+1-\Psi^*\left(\frac{\nu}{2}\right)\}I_i^\Phi(\frac{n_i}{2}) - \frac{1}{2}I_i^\Phi(\frac{n_i}{2}+1)\\
&+ \frac{1}{2}\int_0^\infty u_i^{n_i/2}\log u_i \exp\{-\frac{1}{2}u_i\mathrm{d_i}\}\Phi(\mathrm{u_i^{1/2}A_i})\mathrm{h(u_i},\nu)\mathrm{du_i},
\end{aligned}
$$

where $\Psi^* = \frac{\partial \log(x)}{\partial x}$ is digamma function.

- *skew-slash*

$$
\frac{\partial K_i(\boldsymbol{\theta})}{\partial \nu} = I_i^\Phi(\frac{n_i}{2}+\nu-1) + \nu\int_0^1 u_i^{n_i/2+\nu-1}\log u_i\exp\{-\frac{1}{2}u_i\mathrm{d_i}\}\Phi(\mathrm{u_i^{1/2}A_i})\mathrm{du_i},
$$

- *skew-contaminated normal*

$$\begin{aligned}
\frac{\partial K_i(\boldsymbol{\theta})}{\partial \nu_1} &= \nu_2^{n_i/2} \exp\{-\frac{1}{2}\nu_2 d_i\}\Phi_1(\nu_2^{1/2}A_i) - \exp\{-\frac{1}{2}d_i\}\Phi_1(A_i), \\
\frac{\partial K_i(\boldsymbol{\theta})}{\partial \nu_2} &= \nu_1 \frac{n_i}{2}\nu_2^{n_i/2-1} \exp\{-\frac{1}{2}\nu_2 d_i\}\Phi_1(\nu_2^{1/2}A_i) - \frac{1}{2}\nu_2^{n_i/2} d_i \exp\{-\frac{1}{2}\nu_2 d_i\}\Phi_1(\nu_2^{1/2}A_i) \\
&+ \frac{1}{2}\nu_2^{(n_i-1)/2}A_i \exp\{-\frac{1}{2}\nu_2 d_i\}\phi_1(\sqrt{\nu_2^{1/2}A_i})
\end{aligned}$$

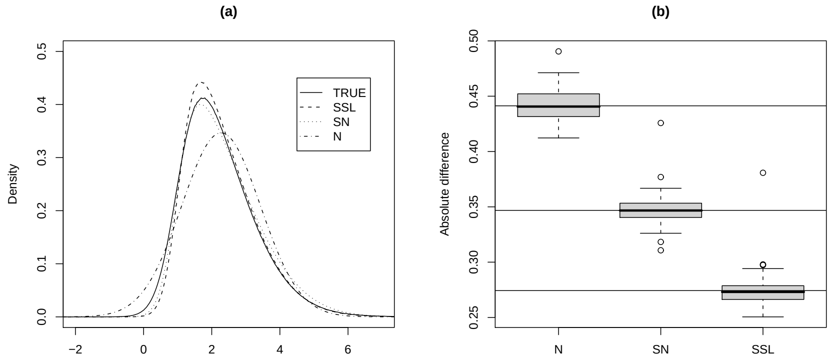
Appendix D: Additional simulation study results

Simulation results for SSL-LMM

Table 1: Monte Carlo results based on 500 data sets, true $SSL_1(0, 2, 3, 3)$ distribution for the random effects and $SL_1(0, 0.25, 3)$ for the random errors. MEAN and SD are average and standard deviation of the estimates, AVE SE is average of estimated standard errors. True values of parameters are in parentheses.

Parameter	MEAN	SD	AVE SE	MEAN	SD	AVE SE	MEAN	SD	AVE SE
	(i) SSL–LMM			(ii) SN–LMM			(iii) N–LMM		
β_0 (1)	1.0404	0.2855	0.2371	0.9651	0.2621	0.2560	2.2735	0.1728	0.1722
β_1 (2)	2.0000	0.0183	0.0191	2.0001	0.0191	0.0193	2.0001	0.0191	0.0198
β_2 (1)	1.0236	0.1995	0.1994	1.0344	0.2035	0.2082	1.0300	0.2294	0.2448
σ_e^2 (0.25)	0.2607	0.0375	0.0364	0.3715	0.0335	0.0256	0.3716	0.0334	0.0268
σ_b^2 (2)	2.0537	0.6060	0.5461	3.1105	0.8718	0.6604	1.3255	0.3248	0.2048
λ (3)	4.1008	3.7138	3.3410	4.3293	2.5387	2.5377	-	-	-
ν (4)	4.9442	5.5248	9.3553	-	-	-	-	-	-

Figure 1: Simulation study based on 500 data sets of SSL–LMM. (a) box-plot of the mean absolute difference of the estimated and simulated random effects for the 100 individuals. (b) True density of the random effects (solid line) and Monte Carlo average estimated densities for 500 data set: using N–LMM (dashed-dotted), SN–LMM (dotted) and SSL–LMM (dotted-line) fitted. The solid lines are the respective mean.



Simulation results for SCN-LMM

Table 2: Monte Carlo results based on 500 data sets, true $SCN_1(0,2,3,0.3,0.3)$ distribution for the random effects and $CN_1(0,0.25,0.3,0.3)$ for the random errors. MEAN and SD are average and standard deviation of the estimates, AVE SE is average of estimated standard errors. True values of parameters are in parentheses.

Parameter	MEAN	SD	AVE SE	MEAN	SD	AVE SE	MEAN	SD	AVE SE
	(i) SCN–LMM			(ii) SN–LMM			(iii) N–LMM		
β_0 (1)	1.0375	0.2416	0.2195	0.8750	0.2126	0.1996	2.3463	0.1916	0.1878
β_1 (2)	1.9989	0.0179	0.0194	1.9990	0.0196	0.0207	1.9990	0.0196	0.0210
β_2 (1)	0.9991	0.1965	0.2034	1.0145	0.2050	0.2164	1.0199	0.2664	0.2662
σ_e^2 (0.25)	0.2438	0.0397	0.0349	0.4235	0.0438	0.0295	0.4238	0.0440	0.0304
σ_b^2 (2)	1.9514	0.5864	0.5381	3.8002	0.9099	0.7080	1.5867	0.3504	0.2430
λ (3)	4.4608	4.1936	3.9464	5.4969	2.7730	3.0077	-	-	-
ν (0.3)	0.3208	0.1220	0.1079	-	-	-	-	-	-
γ (0.3)	0.2961	0.0515	0.0536	-	-	-	-	-	-

Figure 2: Simulation study based on 500 data sets of SCN–LMM. (a) box-plot of the mean absolute difference of the estimated and simulated random effects for the 100 individuals. (b) True density of the random effects (solid line) and Monte Carlo average estimated densities for 500 data set: using N–LMM (dashed-dotted), SN–LMM (dotted) and SCN–LMM (dotted-line) fitted. The solid lines are the respective mean.