Article

Inference and Local Influence Assessment in a Multifactor Skew-Normal Linear Mixed Model

Zeinolabedin Najafi [1], Karim Zare [1,*,†], Mohammad Reza Mahmoudi [2,†], Soheil Shokri [3] and Amir Mosavi [4,5,6,7]

[1] Department of Statistics, Marvdasht Branch, Islamic Azad University, Marvdasht 73711-13119, Iran
[2] Department of Statistics, Faculty of Science, Fasa University, Fasa 74616-86131, Iran
[3] Department of Statistics, Lahijan Branch, Islamic Azad University, Lahijan 44169-39515, Iran
[4] Faculty of Civil Engineering, Technische Universität Dresden, 01069 Dresden, Germany
[5] John von Neumann Faculty of Informatics, Obuda University, 1034 Budapest, Hungary
[6] Institute of Information Society, University of Public Service, 1083 Budapest, Hungary
[7] Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, 81243 Bratislava, Slovakia
* Correspondence: karim.zare@iau.ac.ir
† These authors contributed equally to this work.

Abstract: This work considers a multifactor linear mixed model under heteroscedasticity in random-effect factors and the skew-normal errors for modeling the correlated datasets. We implement an expectation–maximization (EM) algorithm to achieve the maximum likelihood estimates using conditional distributions of the skew-normal distribution. The EM algorithm is also implemented to extend the local influence approach under three model perturbation schemes in this model. Furthermore, a Monte Carlo simulation is conducted to evaluate the efficiency of the estimators. Finally, a real data set is used to make an illustrative comparison among the following four scenarios: normal/skew-normal errors and heteroscedasticity/homoscedasticity in random-effect factors. The empirical studies show our methodology can improve the estimates when the model errors follow from a skew-normal distribution. In addition, the local influence analysis indicates that our model can decrease the effects of anomalous observations in comparison to normal ones.

1. Introduction

Linear mixed models (LMMs) are useful for the statistical analysis of correlated datasets such as longitudinal data. For simplicity, it is usually assumed that both random effects and random errors follow a normal distribution. Under these restrictions, there are several proposals for estimating LMM parameters in the literature; among these, one can refer to Harvill [1], Fellner [2], Khuri et al. [3] and Wu et al. [4]. For example, Harvill [1] and Fellner [2] obtained the maximum likelihood (ML) estimates of parameters in a multifactor normal LMM under heteroscedasticity of random-effect factors. They showed the estimates, in addition to being consistent, were asymptotically normally distributed. However, as pointed out by Zhong and Davidian [5], using these estimation methods may cause invalid statistical inferences when the data are asymmetric. Therefore, many authors have criticized the common use of the normality assumption (see, e.g., [6–10]).

From a practical perspective, the most frequently used method to achieve normality is to apply a transformation on the variables. Although such methods may provide

suitable empirical results, they should not be used if a more reasonable theoretical model is available [11]. Thus, it is of great interest to develop estimation methods in statistical models with flexible distribution assumptions. In this sense, one of the simplest and applicable distributions that provides skewness and contains a normal distribution is the skew-normal distribution introduced by Azzalini [12]. Although it is old, it is still used in statistical models due to its flexibility and simplicity, especially in complex models where it is difficult to use new generalized skewed distributions. Some new works in this research area are [13,14]. In the literature, many authors have studied inferences on parameters in LMMs (with only one random-effect factor) where the random effects or the model random errors follow asymmetric and non-normal distributions (see, e.g., [5,7–9,15–22]). Verbeke and Lesaffre [15], through an extended simulation when the random-effect distribution was misspecified, showed that the standard errors of the parameters needed to be corrected. Arellano-Valle et al. [18] considered an LMM when both the random-effect factor and the random errors follow the SN distribution. Due to complexity, they derived marginal distributions and implemented an EM algorithm to obtain the ML estimates. They indicated that the estimates had more efficiency than normal estimates when the normality assumption was violated. Lachos et al. [19] presented an LMM when the random effects followed a multivariate SN independent distribution. They derived the ML estimates of the parameters based on an efficient EM algorithm. They also investigated a technique to predict the response variable. Kheradmandi and Rasekh [20] followed [18] and obtained the ML estimates in the LMM when the fixed effects were measured with non-negligible errors. In this work, a multifactor SN–LMM was considered with different variances for the random-effect factors to show how heteroscedasticity, as a freer assumption in random-effect factors, can improve our statistical results. Here, a SN–LMM is an LMM with SN distribution in the model random errors.

A diagnostic analysis is a necessary step in statistical analysis after parameter estimation. The local influence approach, a pioneering work of Cook [23], is one of the most important diagnostic tools for assessing the stability of the estimation parameters. Due to the complicated calculations of Cook's local influence approach in statistical models with incomplete data, Zhu and Lee [24] developed Cook's approach to these models based on the conditional expectation of a complete-data log likelihood at the E-step of the EM algorithm. To see some applications of Zhu and Lee's approach, one can refer to [25–28]. Local influence analysis for LMMs based on SN distribution had been studied by Bolfarine [29], Montenegro et al. [30], and Zeller et al. [31]. All these works considered local influence diagnostics for an LMM based on SN distribution in the random-effect factor. Furthermore, all perturbation schemes were considered the same. In this work, besides parameter estimation, we developed Zhu and Lee's local influence diagnostic measures for the LMM under different assumptions on random effects and random errors that were mentioned before. A different perturbation scheme was also considered concerning the previous works. The rest of the paper is structured as follows. In Section 2, we present the model and obtain distributional facts about the variables that will help us use the EM algorithm. In Section 3, parameter estimation and random effects prediction are derived via the EM algorithm. In Section 4, the local influence diagnostic measures for the LMM are extended based on the methodology proposed by Zhu and Lee [24]. The basic building blocks of three perturbation schemes are also derived. In Section 5, a simulation study is performed to compare the normal LMM and the SN-LMM, and then a real dataset is analyzed to perform an illustrative comparison. Discussion and conclusions of this paper are given in Section 6.

2. The Model Definition

Consider the following LMM:

$$Y = X\beta + Ub + \varepsilon = X\beta + \sum_{i=1}^{m} U_i b_i + \varepsilon, \tag{1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, which are fixed effects; $\boldsymbol{X}$ and $\boldsymbol{U} = [\boldsymbol{U}_1|\boldsymbol{U}_2|\ldots|\boldsymbol{U}_m]$ are $n \times p$ and $n \times q$ known design matrices, respectively, where $\boldsymbol{U}_i$ is an $n \times q_i$ design matrix of the random-effect factor i; $\boldsymbol{b} = \left(\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top, \ldots, \boldsymbol{b}_m^\top\right)^\top$, where $\boldsymbol{b}_i$ is a $q_i \times 1$ vector of unobservable random effects from $N_q\left(\boldsymbol{0}, \sigma_i^2 \boldsymbol{I}_{q_i}\right)$, $i = 1, \ldots, m$; $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of unobservable random errors from $SN_n\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n, \boldsymbol{\lambda}_\varepsilon\right)$, that is n-dimensional SN distribution with skewness vector $\boldsymbol{\lambda}_\varepsilon \in \mathbb{R}^n$. The variances σ^2 and σ_i^2, $i = 1, \ldots, m$ are named variance components. We assume that $\boldsymbol{b}_i$, $i = 1, \ldots, m$, and $\boldsymbol{\varepsilon}$ are mutually independent. One may also write $\boldsymbol{b} \sim N_q\left(\boldsymbol{0}, \sigma^2 \boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Sigma}$ is a block diagonal matrix with the ith block being $\gamma_i \boldsymbol{I}_{q_i}$, for $\gamma_i = \sigma_i^2 / \sigma^2$, that are called the ratio of variance components.

The above assumptions conclude that $\boldsymbol{Y}|\boldsymbol{b} \sim SN_n\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{b}, \sigma^2 \boldsymbol{I}_n, \boldsymbol{\lambda}_\varepsilon\right)$ and so the joint distribution of the vectors $\boldsymbol{Y}$ and $\boldsymbol{b}$ is obtained as follows:

$$f(\boldsymbol{y}, \boldsymbol{b}) = 2\phi_q\left(\boldsymbol{b}\Big|\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}\right)\phi_n\left(\boldsymbol{y}\Big|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{b}, \sigma^2\boldsymbol{I}_n\right)$$

$$\times \Phi\left(\boldsymbol{\lambda}_\varepsilon^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{U}\boldsymbol{b})/\sigma\right) \tag{2}$$

where $\phi_n(\,|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the n-variate normal density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\Phi(\cdot)$ represents the cumulative distribution function of $N(0,1)$.

From (2), the marginal density of $\boldsymbol{Y}$ would be as follows:

$$f(\boldsymbol{y}) = \int_{\mathbb{R}^q} 2\phi_q\left(\boldsymbol{b}\Big|\boldsymbol{\Sigma}\boldsymbol{U}^\top\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \sigma^2\boldsymbol{\Sigma}\boldsymbol{T}\right)\phi_n\left(\boldsymbol{y}\Big|\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V}\right)$$

$$\times \Phi\left(\boldsymbol{\lambda}_\varepsilon^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{U}\boldsymbol{b})/\sigma\right)d\boldsymbol{b}$$

$$= 2\phi_n\left(\boldsymbol{y}\Big|\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V}\right)E\left(\Phi\left(\boldsymbol{\lambda}_\varepsilon^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{U}B)/\sigma\right)\right)$$

$$= 2\phi_n\left(\boldsymbol{y}\Big|\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V}\right)\Phi\left(\frac{\boldsymbol{\lambda}_\varepsilon^\top\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma\delta}\right); \ \boldsymbol{y} \in \mathbb{R}^n$$

where $\boldsymbol{V} = \boldsymbol{I}_n + \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top = \boldsymbol{I}_n + \sum_{i=1}^m \gamma_i \boldsymbol{U}_i\boldsymbol{U}_i^\top$, $\boldsymbol{T} = \left(\boldsymbol{I}_q + \boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\right)^{-1}$ so that $\boldsymbol{\Sigma}\boldsymbol{T}$ is a symmetric, non-singular matrix; $B \sim N_q\left(\boldsymbol{\Sigma}\boldsymbol{U}^\top\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \sigma^2\boldsymbol{\Sigma}\boldsymbol{T}\right)$; and where $\delta^2 = 1 + \boldsymbol{\lambda}_\varepsilon^\top\left(\boldsymbol{I}_n - \boldsymbol{V}^{-1}\right)\boldsymbol{\lambda}_\varepsilon$. Therefore, $\boldsymbol{Y}$ has a generalized SN distribution.

Based on the distribution of $\boldsymbol{Y}$, the log-likelihood function of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}_\varepsilon^\top\right)^\top$ is given by

$$l(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) \propto -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|\boldsymbol{V}|)$$

$$-\frac{1}{2\sigma^2}\left[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right] \tag{3}$$

$$+\log\Phi\left(\frac{\boldsymbol{\lambda}_\varepsilon^\top\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma\delta}\right),$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^\top$.

As can be seen, there is no obvious solution for the direct maximization of Equation (3), and the likelihood function has to be maximized numerically. Using numerical approaches, besides the high computational costs and lack of robustness to the starting values, causes some problems for the maximization due to the term $\log(\Phi(\cdot))$. Therefore, corresponding to previous studies in the field of using the skew family in modeling (see, e.g., [8,9,18–20,22,27]), an EM algorithm was applied to reduce computation complexity with high efficiency.

The EM algorithm, introduced by Dempster et al. [32], is a famous iterative algorithm for ML estimation in incomplete data models. One of the major reasons for its popularity is the M-step that includes maximization of a likelihood function based on complete data, which is often computationally simple. It is also not very sensitive to the starting parameter values.

Let $TN(\xi, \eta^2; a, b)$ represent the truncated normal distribution with parameters ξ and η^2 and truncation range (a, b). If the distribution of y is written as

$$f(y) = 2\phi_n\left(y\middle|X\beta, \sigma^2 V\right) \int_0^{+\infty} \phi\left(z\middle|\lambda_\varepsilon^\top V^{-1}(y - X\beta), \sigma^2\delta^2\right) dz; \ y \in \mathbb{R}^n,$$

the joint distribution of Y and the missing variable Z will be

$$f(y, z) = 2\phi_n\left(y\middle|X\beta, \sigma^2 V\right)\phi\left(z\middle|\lambda_\varepsilon^\top V^{-1}(y - X\beta), \sigma^2\delta^2\right); \ y \in \mathbb{R}^n, z > 0.$$

Based on the above joint distribution, the conditional distribution of $Z|y$ is obtained as

$$f(z|y) = \frac{\phi\left(z\middle|\lambda_\varepsilon^\top V^{-1}(y - X\beta), \sigma^2\delta^2\right)}{\int_0^{+\infty} \phi\left(z\middle|\lambda_\varepsilon^\top V^{-1}(y - X\beta), \sigma^2\delta^2\right) dz}; \ z > 0.$$

Hence, $Z|y \sim TN(\lambda_\varepsilon^\top V^{-1}(y - X\beta), \sigma^2\delta^2; 0, +\infty)$. Now, with the help of the properties of the truncated normal distribution [33], we have

$$E(Z|y) = \lambda_\varepsilon^\top V^{-1}(y - X\beta) + \sigma\delta W\left(\frac{\lambda_\varepsilon^\top V^{-1}(y - X\beta)}{\sigma\delta}\right), \tag{4}$$

and

$$E\left(Z^2\middle|y\right) = \left(\lambda_\varepsilon^\top V^{-1}(y - X\beta)\right)^2 + \sigma^2\delta^2$$
$$+ \sigma\delta\lambda_\varepsilon^\top V^{-1}(y - X\beta)W\left(\frac{\lambda_\varepsilon^\top V^{-1}(y - X\beta)}{\sigma\delta}\right), \tag{5}$$

To predict the random effects, we first need the conditional distribution of $b|y$ given by

$$\begin{aligned} f(b|y) &= \frac{f(y, b)}{f(y)} \\ &= \frac{1}{\alpha}\phi_q\left(b\middle|\Sigma U^\top V^{-1}(y - X\beta), \sigma^2\Sigma T\right)\Phi(\lambda_\varepsilon^\top(y - X\beta - Ub)/\sigma), \end{aligned}$$

where

$$\alpha = \Phi\left(\frac{\lambda_\varepsilon^\top V^{-1}(y - X\beta)}{\sigma\delta}\right).$$

Now, the conditional log-likelihood function of b—given θ, X and y—is obtained as

$$\begin{aligned} l^*(b; \theta, X, y) \ \propto \ &-\frac{q}{2}\log(\sigma^2) - \frac{1}{2}\log(|\Sigma T|) \\ &-\frac{1}{2\sigma^2}\left(b - \Sigma U^\top V^{-1}(y - X\beta)\right)^\top (\Sigma T)^{-1}\left(b - \Sigma U' V^{-1}(y - X\beta)\right) \\ &+\log\left(\Phi(\lambda_\varepsilon^\top(y - X\beta - Ub)/\sigma)\right) - \log(\alpha). \end{aligned}$$

As seen again, l^* like l has the term $\log(\Phi(\cdot))$ and so, to predict b, based on the ML method, we use the EM algorithm.

To do this, we first rewrote the conditional distribution of $b|y$ as

$$\begin{aligned} f(b|y) &= \frac{1}{\alpha}\phi_q\left(b\middle|\Sigma U^\top V^{-1}(y - X\beta), \sigma^2\Sigma T\right) \\ &\times \int_0^{+\infty} \phi\left(z^*\middle|\lambda_\varepsilon^\top(y - X\beta - Ub), \sigma^2\right) dz^*; \ b \in \mathbb{R}^q, \ y \in \mathbb{R}^n. \end{aligned}$$

So, the conditional distribution of b and the missing variable Z^* given y is equal to

$$\begin{aligned} &f(b, z^*|y) \\ &= \frac{1}{\alpha}\phi_q\left(b\middle|\Sigma U^\top V^{-1}(y - X\beta), \sigma^2\Sigma T\right) \\ &\times\phi\left(z^*\middle|\lambda_\varepsilon^\top(y - X\beta - Ub), \sigma^2\right); \ b \in \mathbb{R}^q, \ y \in \mathbb{R}^n, z^* > 0. \end{aligned} \tag{6}$$

The above equation concludes that the conditional distribution of the missing variable Z^* given b and y is equal to

$$f(z^*|b,y) = \frac{\phi\left(z^*|\lambda_\varepsilon^\top(y - X\beta - Ub), \sigma^2\right)}{\int_0^{+\infty} \phi\left(z^*|\lambda_\varepsilon^\top(y - X\beta - Ub), \sigma^2\right)dz^*}, \quad z^* > 0.$$

As seen, $Z^*|(b,y) \sim TN\left(\lambda_\varepsilon^\top(y - X\beta - Ub), \sigma^2; 0, +\infty\right)$ and hence,

$$E(Z^*|(b,y)) = \lambda_\varepsilon^\top(y - X\beta - Ub) + \sigma W\left(\lambda_\varepsilon^\top(y - X\beta - Ub)/\sigma\right), \tag{7}$$

and

$$\begin{aligned} E\left(Z^{*2}|(b,y)\right) &= \left(\lambda_\varepsilon^\top(y - X\beta - Ub)\right)^2 + \sigma^2 \\ &\quad + \sigma\lambda_\varepsilon^\top(y - X\beta - Ub)W\left(\lambda_\varepsilon^\top(y - X\beta - Ub)/\sigma\right). \end{aligned} \tag{8}$$

3. Parameter Estimation via the EM Algorithm

Now, let $y = (y_1, \ldots, y_n)'$ be a vector of observed responses and z be a missing observation. Then, the complete log-likelihood function associated with $y_c = (y', z)'$ will be obtained as

$$\begin{aligned} l_c(\theta; X, y_c) &= C - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|V|) \\ &\quad - \frac{1}{2\sigma^2}\left[(y - X\beta)^\top V^{-1}(y - X\beta)\right] \\ &\quad - \frac{1}{2}\log(\sigma^2\delta^2) - \frac{1}{2\sigma^2\delta^2}\left(z - \lambda_\varepsilon^\top V^{-1}(y - X\beta)\right)^2 \end{aligned}$$

where C does not depend on unknown parameters.

If $\hat{\theta}^{(r)} = \left(\hat{\beta}^{\top(r)}, \hat{\sigma}^{2(r)}, \hat{\gamma}^{\top(r)}, \hat{\lambda_\varepsilon}^{\top(r)}\right)^\top$ is the estimate in the rth iteration, then the expected complete log-likelihood function would be

$$\begin{aligned} &Q\left(\theta|\hat{\theta}^{(r)}\right) \\ &= E\left[l_c(\theta; X, y_c)|\hat{\theta}^{(r)}\right] \propto -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|V|) \\ &\quad - \frac{1}{2\sigma^2}\left[(y - X\beta)^\top V^{-1}(y - X\beta)\right] - \frac{1}{2}\log(\sigma^2\delta^2) \\ &\quad - \frac{1}{2\sigma^2\delta^2}\left(\hat{z^2}^{(r)} - 2\hat{z}^{(r)}\lambda_\varepsilon^\top V^{-1}(y - X\beta) + \left(\lambda_\varepsilon^\top V^{-1}(y - X\beta)\right)^2\right) \end{aligned} \tag{9}$$

where $\hat{z}^{(r)} = E\left[Z|y; \hat{\theta}^{(r)}\right]$ and $\hat{z^2}^{(r)} = E\left[Z^2|y; \hat{\theta}^{(r)}\right]$ are calculated by substituting $\hat{\theta}^{(r)}$ in Equations (4) and (5), respectively.

To obtain a new estimate $\hat{\theta}^{(r+1)}$, the M-step maximizes $Q\left(\theta|\hat{\theta}^{(r)}\right)$ with respect to θ. This was obtained as a solution of the following equations:

$$\frac{\partial Q\left(\theta|\hat{\theta}^{(r)}\right)}{\partial \beta} = 0,$$

$$\frac{\partial Q\left(\theta|\hat{\theta}^{(r)}\right)}{\partial \sigma^2} = 0,$$

and

$$\frac{\partial Q\left(\theta|\hat{\theta}^{(r)}\right)}{\partial \gamma_i} = 0; \quad i = 1, \ldots, m.$$

If we define $h = V^{-1}\lambda_\varepsilon$, then from the Equation (9), the estimation of β, is given by

$$
\begin{aligned}
\hat{\beta}^{(r+1)} &= \left(X^\top \left(\hat{V}^{(r)-1} + \hat{h}^{(r)}\hat{h}^{\top(r)}/\hat{\delta^2}^{(r)} \right) X \right)^{-1} \\
&\times \left(X^\top \left(\hat{V}^{(r)-1} + \hat{h}^{(r)}\hat{h}^{\top(r)}/\hat{\delta^2}^{(r)} \right) y - \hat{z}^{(r)} X^\top \hat{h}^{(r)}/\hat{\delta^2}^{(r)} \right)
\end{aligned}
$$

Before the estimation of variance components, we presented a realized value of the random effects. In a similar way for fixed effects (See Appendix A.1 for more details.), the ML prediction of b was given by

$$
\hat{b}^{(r+1)} = \hat{\Sigma}^{(r)} U^\top \hat{V}^{(r)-1} \left[\left(I_n + \frac{1}{\hat{\delta^2}^{(r)}} \hat{\lambda_\varepsilon}^{(r)} \hat{\lambda_\varepsilon}^{\top(r)} \hat{V}^{(r)-1} \right) \left(y - X\hat{\beta}^{(r)} \right) - \frac{1}{\hat{\delta^2}^{(r)}} \hat{z}^{*(r)} \hat{\lambda_\varepsilon}^{(r)} \right].
$$

where $\hat{z}^{*(r)} = E\left[Z^* | \left(\hat{b}^{(r)}, y \right); \hat{\theta}^{(r)} \right]$ and $\hat{z}^{*2(r)} = E\left[Z^{*2} | \left(\hat{b}^{(r)}, y \right); \hat{\theta}^{(r)} \right]$ were calculated by substituting $\hat{\theta}^{(r)}$ and $\hat{b}^{(r)}$ into Equations (7) and (8), respectively.

Then, for the ith random-effect factor, we have

$$
\hat{b}_i^{(r+1)} = \hat{\gamma}_i^{(r)} U_i^\top \hat{V}^{(r)-1} \left[\left(I_n + \frac{1}{\hat{\delta^2}^{(r)}} \hat{\lambda_\varepsilon}^{(r)} \hat{\lambda_\varepsilon}^{\top(r)} \hat{V}^{(r)-1} \right) \left(y - X\hat{\beta}^{(r)} \right) - \frac{1}{\hat{\delta^2}^{(r)}} \hat{z}^{*(r)} \hat{\lambda_\varepsilon}^{(r)} \right].
$$

From Equation (9), the estimates of variance components were derived as

$$
\hat{\sigma^2}^{(r+1)} = \frac{1}{n+1} \left[\left(y - X\hat{\beta}^{(r)} \right)^\top \left(\hat{V}^{(r)-1} + \hat{h}^{(r)}\hat{h}^{\top(r)}/\hat{\delta^2}^{(r)} \right) \left(y - X\hat{\beta}^{(r)} \right) + \left(\hat{z^2}^{(r)} - 2\hat{z}^{(r)}\hat{h}^{\top(r)} \left(y - X\hat{\beta}^{(r)} \right) \right)/\hat{\delta^2}^{(r)} \right].
$$

and

$$
\begin{aligned}
\hat{\sigma_i^2}^{(r+1)} = \frac{\hat{\gamma}_i^{(r)2}}{q_i - \operatorname{tr}(T_{ii})} \Bigg\{ & \left(y - X\hat{\beta}^{(r)} \right)^\top \hat{V}^{(r)-1} U_i U_i^\top \hat{V}^{(r)-1} \left(y - X\hat{\beta}^{(r)} \right) \\
& - \frac{\hat{\sigma^2}^{(r)}}{\hat{\delta^2}^{(r)}} \hat{h}^{\top(r)} U_i U_i^\top h \\
& + \frac{1}{\hat{\delta^2}^{(r)2}} \left[\hat{z^2}^{(r)} - 2\hat{z}^{(r)}\hat{h}^{\top(r)} \left(y - X\hat{\beta}^{(r)} \right) \right. \\
& \left. + \left(\hat{h}^{\top(r)} \left(y - X\hat{\beta}^{(r)} \right) \right)^2 \right] \hat{h}^{(r)'} U_i U_i' \hat{h}^{(r)} \\
& + \frac{2}{\hat{\delta^2}^{(r)}} \left[\left(y - X\hat{\beta}^{(r)} \right)^\top \hat{h}^{(r)} \hat{h}^{\top(r)} U_i U_i^\top \hat{V}^{(r)-1} \left(y - X\hat{\beta}^{(r)} \right) \right. \\
& \left. - \hat{z}^{(r)} \hat{h}^{\top(r)} U_i U_i^\top \hat{V}^{(r)-1} \left(y - X\hat{\beta}^{(r)} \right) \right\}
\end{aligned}
$$

where T_{ii} is ith diagonal block of the matrix T (See Appendix A.2 for more details.). By substituting $\hat{z}^{(r)}$ and $\hat{z^2}^{(r)}$ with $\hat{z}^{*(r)}$ and $\hat{z}^{*(r)2}$, respectively, in the above equation, one can obtain another estimator for σ_i^2 based on $\hat{b}_i^{(r+1)}$ similar to its corresponding estimator in the normal LMM. This estimator would be as follows:

$$
\tilde{\sigma_i^2}^{(r+1)} = \frac{1}{q_i - \operatorname{tr}(T_{ii})} \left(\hat{b}_i^{(r+1)} \hat{b}_i^{\top(r+1)} - \frac{\tilde{\sigma^2}^{(r)}}{\hat{\delta}^{(r)}} d_i^\top d_i \right),
$$

where $d_i = U_i^\top \hat{h}^{(r)}$.

Finally, the ML estimates of skewness parameters would be

$$
\hat{\lambda_\varepsilon}^{(r+1)} = \operatorname{argmax}_{\lambda_\varepsilon} \left(Q\left(\hat{\beta}^{(r+1)}, \hat{\sigma^2}^{(r+1)}, \hat{\gamma}^{(r+1)}, \lambda_\varepsilon \middle| \hat{\theta}^{(r)} \right) \right),
$$

where $Q\left(\hat{\boldsymbol{\beta}}^{(r+1)}, \hat{\sigma 2}^{(r+1)}, \hat{\boldsymbol{\gamma}}^{(r+1)}, \boldsymbol{\lambda}_\varepsilon \middle| \hat{\boldsymbol{\theta}}^{(r)}\right)$ is $Q\left(\boldsymbol{\theta} \middle| \hat{\boldsymbol{\theta}}^{(r)}\right)$ evaluated at updated $\hat{\boldsymbol{\beta}}^{(r+1)}, \hat{\sigma 2}^{(r+1)}$ and $\hat{\boldsymbol{\gamma}}^{(r+1)}$.

The above algorithm stops when a reasonable convergence rule is satisfied (e.g., $\hat{\boldsymbol{\theta}}^{(r+1)} - \hat{\boldsymbol{\theta}}^{(r)} < 10^{-6}$). A set of adequate starting values can be obtained by solving the normal LMM for $\boldsymbol{\beta}$, σ^2 and σ_i^2, $i = 1, \ldots, m$ and the sample skewness coefficient of the residuals or zero values for $\boldsymbol{\lambda}_\varepsilon$. But, as recommended in the literature, the EM algorithm should be run several times with different starting values.

4. Local Influence Analysis

Cook's local influence method was used to evaluate the influence of various minor model perturbations on the parameter estimates. Inspired by the general idea of the EM algorithm, Zhu and Lee [24] generalized the local influence diagnostic method to general statistical models with incomplete data based on a Q-function. Here, we briefly studied a natural extension of this procedure to SN–LMM. In this section, we assumed that the γ_i's were known. If the γ_i's were unknown, the ML estimates would have been placed back into $\boldsymbol{\Sigma}$, so the vector $\boldsymbol{\theta}$ would have been $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\lambda}_\varepsilon^\top\right)^\top$.

If we let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_n)$ be a n-dimensional vector of perturbations varying in open region $\Omega \in \mathbb{R}^n$, the perturbed complete-data log-likelihood function would be denoted by $l_c(\boldsymbol{\theta}, \boldsymbol{\omega}; \boldsymbol{X}, \boldsymbol{y}_c)$. It is assumed that there exists w_0, a vector of no perturbation, such that $l_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0; \boldsymbol{X}, \boldsymbol{y}_c) = l_c(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}_c)$ for all $\boldsymbol{\theta}$. Let $Q\left(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}) \middle| \hat{\boldsymbol{\theta}}^{(r)}\right)$ be the maximum value of the $Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \middle| \hat{\boldsymbol{\theta}}^{(r)}\right) = E(l_c(\boldsymbol{\theta}, \boldsymbol{\omega}; \boldsymbol{X}, \boldsymbol{y}_c))$, where $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denotes the ML estimate under $Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \middle| \hat{\boldsymbol{\theta}}^{(r)}\right)$. To evaluate the influence of minor perturbations on the ML estimate $\hat{\boldsymbol{\theta}}$, one may regard the Q-displacement function, defined as follows:

$$QD(\boldsymbol{\omega}) = 2\left\{ Q\left(\hat{\boldsymbol{\theta}} \middle| \hat{\boldsymbol{\theta}}^{(r)}\right) - Q\left(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}) \middle| \hat{\boldsymbol{\theta}}^{(r)}\right) \right\}$$

Zhu and Lee [24] suggested studying the local behavior of $QD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. Corresponding to their proposal, the normal curvature in the direction of some unit vector $\boldsymbol{d}$, given by $C_{QD,\boldsymbol{d}} = -2\boldsymbol{d}^\top \ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0} \boldsymbol{d}$, can be employed to summarize the local behavior of the Q-displacement function, where

$$-\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\omega}_0}^\top \left\{ -\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \right\}^{-1} \Delta_{\boldsymbol{\omega}_0}$$

in which

$$\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q\left(\boldsymbol{\theta} \middle| \hat{\boldsymbol{\theta}}^{(r)}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$$

and

$$\Delta_{\boldsymbol{\omega}} = \frac{\partial^2 Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \middle| \hat{\boldsymbol{\theta}}^{(r)}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(\boldsymbol{\omega})}.$$

Since most influence measures proposed in the statistical literature are closely related to a spectral decomposition of $-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}$, we used this expression to detect influential observations. Let $\sum_{k=1}^n \xi_k \boldsymbol{e}_k \boldsymbol{e}_k^\top$ be the spectral decomposition of $-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}$ where $\{(\xi_i, \boldsymbol{e}_i) : i = 1, \ldots, n\}$ are the eigenvalue–eigenvector pairs of the matrix $-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}$ with $\xi_1 \geq \ldots \geq \xi_q, \xi_{q+1} = \ldots = \xi_n = 0$ and $\left\{ \boldsymbol{e}_k = \left(e_{k1}, \ldots, e_{kq}\right) : k = 1, \ldots, n \right\}$ is the associated orthonormal basis.

Following Zhu and Lee [24] and Lu and Song [34], the assessment of influential observations was based on

$$M(0)_i = \sum_{k=1}^{q} \widetilde{\xi}_k e_{ki}^2, \ i = 1, \dots, n$$

where $\widetilde{\xi}_k = \xi_k / (\xi_1 + \dots + \xi_q)$. The influence measure $M(0)_i$ may be obtained through

$$B_{QD,d_i} = \frac{-2d_i^\top \ddot{Q}_{\omega_0} d_i}{\mathrm{tr}\left(-2\ddot{Q}_{\omega_0}\right)}$$

where d_i is an $n \times 1$ vector with the ith element equal to one and all other elements equal to zero. Moreover, corresponding to Lee and Xu [35], we used the cut-off point $1/n + c^* SM(0)$ to consider the ith observation as influential, where c^* is a constant, chosen according to the real application, and $SM(0)$ denotes the standard deviation of $\{M(0)_i : \ i = 1, \dots, n\}$.

4.1. The Hessian Matrix

To achieve the local influence diagnostic measures for a particular perturbation scheme, we needed to compute $\frac{\partial^2 Q(\theta | \hat{\theta}^{(r)})}{\partial \theta \partial \theta^\top}$. It follows from (9) that the Hessian matrix has elements given by

$$\frac{\partial^2 Q\left(\theta | \hat{\theta}^{(r)}\right)}{\partial \beta \partial \beta^\top} = -\frac{1}{\sigma^2} X^\top \left(V^{-1} + \frac{1}{\delta^2} h h^\top\right) X,$$

$$\frac{\partial^2 Q\left(\theta | \hat{\theta}^{(r)}\right)}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} X^\top \left\{ \left(V^{-1} + \frac{1}{\delta^2} h h^\top\right)(y - X\beta) - \frac{\hat{z}^{(r)}}{\delta^2} h \right\},$$

$$\begin{aligned}\frac{\partial^2 Q(\theta | \hat{\theta}^{(r)})}{\partial \beta \partial \lambda_\varepsilon^\top} &= \frac{2}{\sigma^2 \delta^4} \left\{ \hat{z}^{(r)} - h^\top (y - X\beta) \right\} X^\top h (\lambda_\varepsilon - h)^\top \\ &\quad - \frac{1}{\sigma^2 \delta^2} X^\top V^{-1} \left[\left\{ \hat{z}^{(r)} - h^\top (y - X\beta) \right\} I_n - \lambda_\varepsilon (y - X\beta)^\top V^{-1} \right],\end{aligned}$$

$$\frac{\partial^2 Q\left(\theta | \hat{\theta}^{(r)}\right)}{\partial \sigma^2 \partial \sigma^2} = \frac{n+1}{2\sigma^4} - \frac{1}{\sigma^6}(y - X\beta)^\top V^{-1}(y - X\beta) - \frac{1}{\sigma^6 \delta^2} \left[\hat{z^2}^{(r)} - 2\hat{z}^{(r)} h^\top (y - X\beta) + \left\{ h^\top (y - X\beta) \right\}^2 \right],$$

$$\begin{aligned}\frac{\partial^2 Q(\theta | \hat{\theta}^{(r)})}{\partial \sigma^2 \partial \lambda_\varepsilon^\top} &= -\frac{1}{\sigma^4 \delta^4} \left[\hat{z^2}^{(r)} - 2\hat{z}^{(r)} h^\top (y - X\beta) + \left\{ h^\top (y - X\beta) \right\}^2 \right] (\lambda_\varepsilon - h)^\top \\ &\quad - \frac{1}{\sigma^4 \delta^2} \left\{ \hat{z}^{(r)} - h^\top (y - X\beta) \right\} (y - X\beta)^\top V^{-1},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 Q(\theta | \hat{\theta}^{(r)})}{\partial \lambda_\varepsilon \partial \lambda_\varepsilon^\top} &= \frac{2}{\delta^4}(\lambda_\varepsilon - h)(\lambda_\varepsilon - h)^\top - \frac{1}{\delta^2}(I_n - V^{-1}) \\ &\quad - \frac{1}{\sigma^2 \delta^4} \left[\hat{z^2}^{(r)} - 2\hat{z}^{(r)} h^\top (y - X\beta) + \left\{ h^\top (y - X\beta) \right\}^2 \right] \left[\frac{4}{\delta^2}(\lambda_\varepsilon - h)(\lambda_\varepsilon - h)^\top - (I_n - V^{-1}) \right] \\ &\quad - \frac{2}{\sigma^2 \delta^4} \left\{ \hat{z}^{(r)} - h^\top (y - X\beta) \right\} (\lambda_\varepsilon - h)(y - X\beta)^\top V^{-1} \\ &\quad - \frac{1}{\sigma^2 \delta^2} V^{-1}(y - X\beta) \left[\frac{2}{\delta^2} \left\{ \hat{z}^{(r)} - h^\top (y - X\beta) \right\} (\lambda_\varepsilon - h)^\top + (y - X\beta)^\top V^{-1} \right].\end{aligned}$$

4.2. Perturbation Schemes

In this section, we present three distinct perturbation schemes for the model defined in (1).

4.2.1. Perturbation of the Response Variable

A perturbation of the response variable y is defined as $y_\omega = y + s_y \omega$, where s_y is the standard deviation of y. In this case, $\omega_0 = 0$ and

$$
\begin{aligned}
Q&\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}^{(r)}\right) \\
&\propto -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(|V|) \\
&\quad -\frac{1}{2\sigma^2}\left[(y_\omega - X\beta)^\top V^{-1}(y_\omega - X\beta)\right] - \frac{1}{2}\log(\sigma^2\delta^2) \\
&\quad -\frac{1}{2\sigma^2\delta^2}\left[\hat{z}^{2^{(r)}} - 2\hat{z}^{(r)}h^\top(y_\omega - X\beta) + \left\{h^\top(y_\omega - X\beta)\right\}^2\right].
\end{aligned}
\tag{10}
$$

From (10), the matrix

$$
\Delta_{\omega_0} = \left.\frac{\partial^2 Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}^{(r)}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\right|_{\omega = \omega_0}
$$

has the following elements.

$$
\Delta_\beta = \frac{s_y}{\sigma^2}X^\top\left(V^{-1} + \frac{1}{\delta^2}hh^\top\right)
$$

$$
\Delta_{\sigma^2} = \frac{s_y}{\sigma^4}(y - X\beta)^\top V^{-1} - \frac{s_y}{\sigma^4\delta^2}\left\{\hat{z}^{(r)} - h^\top(y - X\beta)\right\}h^\top
$$

$$
\begin{aligned}
\Delta_{\lambda_\varepsilon} = {}&-\frac{2s_y}{\sigma^2\delta^4}\left\{\hat{z}^{(r)} - h^\top(y - X\beta)\right\}(\lambda_\varepsilon - h)h^\top \\
&+\frac{s_y}{\sigma^2\delta^2}\left[\left\{\hat{z}^{(r)} - h^\top(y - X\beta)\right\}V^{-1} - V^{-1}(y - X\beta)h^\top\right]
\end{aligned}
$$

4.2.2. Perturbation of the kth Column of the Matrix X

We considered altering the kth column matrix X, i.e., x_k, by taking $x_{k\omega} = x_k + s_k\omega$ where s_k is the standard deviation of x_k and $\omega_0 = 0$ represents no perturbation. In this case, the perturbed Q-function took the form

$$
\begin{aligned}
Q&\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}^{(r)}\right) \\
&\propto -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(|V|) \\
&\quad -\frac{1}{2\sigma^2}\left[(y - X_\omega\beta)^\top V^{-1}(y - X_\omega\beta)\right] - \frac{1}{2}\log(\sigma^2\delta^2) \\
&\quad -\frac{1}{2\sigma^2\delta^2}\left[\hat{z}^{2^{(r)}} - 2\hat{z}^{(r)}h^\top(y - X_\omega\beta) + \left\{h^\top(y - X_\omega\beta)\right\}^2\right],
\end{aligned}
\tag{11}
$$

where $X_\omega = (x_{1\omega}, \ldots, x_{p\omega})$. It follows from (11), that the elements of the matrix Δ_{ω_0} were given by

$$
\Delta_\beta = \frac{s_t}{\sigma^2}\left\{u_t(y - X\beta)^\top V^{-1} - \beta_t X^\top V^{-1}\right\} - \frac{s_t}{\sigma^2\delta^2}\left[\left\{\hat{z}^{(r)} - h^\top(y - X\beta)\right\}u_t h^\top + \beta_t X^\top hh^\top\right]
$$

$$
\Delta_{\sigma^2} = -\frac{s_t\beta_t}{\sigma^4}(y - X\beta)^\top V^{-1} + \frac{s_t\beta_t}{\sigma^4\delta^2}\left\{\hat{z}^{(r)} - h^\top(y - X\beta)\right\}h^\top
$$

$$
\Delta_{\lambda_\varepsilon} = \frac{s_t\beta_t}{\sigma^2}\left[\left\{\hat{z}^{(r)} - h^\top(y - X\beta)\right\}\left\{\frac{2}{\delta^4}(\lambda_\varepsilon - h)h^\top - \frac{1}{\delta^2}V^{-1}\right\} + \frac{1}{\delta^2}V^{-1}(y - X\beta)h^\top\right]
$$

4.2.3. Perturbation of the Dispersion Matrix of the Errors

We modified the dispersion matrix of errors, i.e., $\sigma^2 I_n$, to $\sigma^2 D(\boldsymbol{\omega})$ where $D(\boldsymbol{\omega})$ is a diagonal matrix with diagonal elements $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_n)^\top$. The point representing no perturbation is $\boldsymbol{\omega}_0 = (1, 1, \cdots, 1)^\top$. In this case, the perturbed Q-function was obtained as

$$
\begin{aligned}
Q&\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}^{(r)} \right) \\
&\propto -\tfrac{n}{2} \log(\sigma^2) - \tfrac{1}{2} \log(|V(\boldsymbol{\omega})|) \\
&\quad -\tfrac{1}{2\sigma^2} \left[(\boldsymbol{y} - \boldsymbol{X\beta})^\top V(\boldsymbol{\omega})^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) \right] - \tfrac{1}{2} \log(\sigma^2 \delta^2(\boldsymbol{\omega})) \\
&\quad - \tfrac{1}{2\sigma^2 \delta^2(\boldsymbol{\omega})} \left[\hat{z^2}^{(r)} - 2\hat{z}^{(r)} \boldsymbol{h}(\boldsymbol{\omega})^\top (\boldsymbol{y} - \boldsymbol{X\beta}) + \left\{ \boldsymbol{h}(\boldsymbol{\omega})^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\}^2 \right],
\end{aligned}
\tag{12}
$$

where $V(\boldsymbol{\omega}) = D(\boldsymbol{\omega}) + \boldsymbol{U\Sigma U}^\top$, $\boldsymbol{h}(\boldsymbol{\omega}) = V^{-1}(\boldsymbol{\omega})\boldsymbol{\lambda}_\varepsilon$, and $\delta^2(\boldsymbol{\omega}) = 1 + \boldsymbol{\lambda}_\varepsilon^\top (D(\boldsymbol{\omega}) - D(\boldsymbol{\omega})V^{-1}(\boldsymbol{\omega})D(\boldsymbol{\omega}))\boldsymbol{\lambda}_\varepsilon$. From (12), we obtained the elements of $\Delta_{\boldsymbol{\omega}_0}$.

The kth column of the matrix $\Delta_{\boldsymbol{\beta}}$ was given by

$$
\begin{aligned}
-\tfrac{1}{\sigma^2} \boldsymbol{X}^\top \boldsymbol{c}_k \boldsymbol{c}_k^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \quad &+ \tfrac{1}{\sigma^2 \delta^4} \left\{ \boldsymbol{d}_k^\top (\boldsymbol{\lambda}_\varepsilon - \boldsymbol{h}) \right\}^2 \left\{ \hat{z}^{(r)} - \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} \boldsymbol{X}^\top \boldsymbol{h} \\
&- \tfrac{1}{\sigma^2 \delta^2} \left[\boldsymbol{\lambda}_\varepsilon^\top \boldsymbol{c}_k \boldsymbol{c}_k^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \boldsymbol{X}^\top \boldsymbol{h} \right. \\
&\quad \left. - \left\{ \hat{z}^{(r)} - \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} \boldsymbol{X}^\top \boldsymbol{c}_k \boldsymbol{c}_k^\top \boldsymbol{\lambda}_\varepsilon \right]
\end{aligned}
$$

where $\boldsymbol{c}_k$ is the kth column of matrix V^{-1}. Also, the kth element of the vector Δ_{σ^2} was obtained by

$$
\begin{aligned}
-\tfrac{1}{2\sigma^2} (\boldsymbol{y} - \boldsymbol{X\beta})^\top \boldsymbol{c}_k \boldsymbol{c}_k^\top (\boldsymbol{y} - \boldsymbol{X\beta}) &- \tfrac{1}{2\sigma^4 \delta^4} \left\{ \boldsymbol{d}_k^\top (\boldsymbol{\lambda}_\varepsilon - \boldsymbol{h}) \right\}^2 \left[\hat{z^2}^{(r)} - 2\hat{z}^{(r)} \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) + \left\{ \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\}^2 \right] \\
&+ \tfrac{1}{\sigma^4 \delta^2} \left\{ \hat{z}^{(r)} - \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} \boldsymbol{\lambda}_\varepsilon^\top \boldsymbol{c}_k \boldsymbol{c}_k^\top (\boldsymbol{y} - \boldsymbol{X\beta})
\end{aligned}
$$

Finally, the kth column of the matrix $\Delta_{\boldsymbol{\lambda}_\varepsilon}$ was achieved by

$$
\begin{aligned}
\boldsymbol{r}_k - \tfrac{1}{\sigma^2 \delta^2} &\left[\left[\hat{z^2}^{(r)} \quad - 2\hat{z}^{(r)} \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) + \left\{ \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\}^2 \right] \left(2\boldsymbol{r}_k + \tfrac{1}{\delta^2} H_k \boldsymbol{\lambda}_\varepsilon \right) \right. \\
&+ \tfrac{1}{\delta^2} \left\{ \boldsymbol{d}_k^\top (\boldsymbol{\lambda}_\varepsilon - \boldsymbol{h}) \right\}^2 \left\{ \hat{z}^{(r)} - \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} V^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) \\
&- \boldsymbol{\lambda}_\varepsilon^\top \boldsymbol{c}_k \boldsymbol{c}_k^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \left[\tfrac{2}{\delta^2} \left\{ \hat{z}^{(r)} - \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} (I_n - V^{-1}) \boldsymbol{\lambda}_\varepsilon \right. \\
&\left. + V^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) \right] + \left\{ \hat{z}^{(r)} - \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} \boldsymbol{c}_k \boldsymbol{c}_k^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \Big]
\end{aligned}
$$

where

$$
H_k = \left(I_n - V^{-1} \right) D(\boldsymbol{d}_k) \left(I_n - V^{-1} \right),
$$

and

$$
\boldsymbol{r}_k = \tfrac{1}{\delta^4} \left\{ \boldsymbol{d}_k^\top (\boldsymbol{\lambda}_\varepsilon - \boldsymbol{h}) \right\}^2 \left(I_n - V^{-1} \right) \boldsymbol{\lambda}_\varepsilon - \tfrac{1}{\delta^2} H_k \boldsymbol{\lambda}_\varepsilon.
$$

5. Empirical Studies

In this section, we presented a simulation study and a real data example. The **R** software (version 4.1.2) Vienna, Austria, [36] was used to conduct all programs.

5.1. Simulation Study

Here, we considered a Monte Carlo simulation to evaluate the performance of the ML estimates in finite sample sizes. The model was taken as follows:

$$Y_{ij} = x_{ij}^{(1)} \beta_1 + x_{ij}^{(2)} \beta_2 + b_{1j} + \varepsilon_{ij}; \ i = 1, \ldots, s; \ j = 1, \ldots, q,$$

where q, according to [37], is the number of independent clusters, and s is the cluster size in a longitudinal study; hence, the total sample size was $n = sq$. If this model were represented in a matrix form as in (1), we would have $m = 1$, $\boldsymbol{b} = \boldsymbol{b}_1 = (b_{11}, \ldots, b_{1q})^\top$; $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1q}, Y_{21}, \ldots, Y_{2q}, \ldots, Y_{s1}, \ldots, Y_{sq})^\top$; and $\boldsymbol{X} = \left(\boldsymbol{x}^{\top(1)}, \boldsymbol{x}^{\top(2)} \right)^\top$, where $\boldsymbol{x}^{(k)} = \left(x_{11}^{(k)}, \ldots, x_{1q}^{(k)}, x_{21}^{(k)}, \ldots, x_{2q}^{(k)}, \ldots, x_{s1}^{(k)}, \ldots, x_{sq}^{(k)} \right)^\top$ and $\boldsymbol{\varepsilon}$ has the same structure as $\boldsymbol{Y}$. We considered the following combinations for simulation: $q = 50$ or 100 and $s = 3$, which are usual sample sizes in longitudinal studies, $\beta_1 = -1$, $\beta_2 = 2$, $\boldsymbol{x}^{(k)} \sim N_n(0, \boldsymbol{I}_n)$, $\boldsymbol{b}_1 \sim N_q(0, \sigma_1^2 \boldsymbol{I}_q)$, $\boldsymbol{\varepsilon} \sim SN_n(0, \sigma^2 \boldsymbol{I}_n, \boldsymbol{\lambda}_\varepsilon = (\lambda_1 = -3, \lambda_2 = 0, \lambda_3 = 3)^\top \otimes \mathbf{1}_q / n)$, where $\mathbf{1}_q$ is a $q \times 1$ vector of 1s. To see the effect of variability in data on the estimation of the parameters, we took $\sigma_1^2 = 0.5^2$ and $\sigma^2 = 0.4^2$ as a low dispersion and $\sigma_1^2 = 0.9^2$ and $\sigma^2 = 0.8^2$ as a high dispersion in the response data. We also considered two cases for obtaining our estimates: (I) without taking into account skewness in the model and using the ML estimates under normal errors, which is called here the Normal estimates; and (II), using the estimates in this work named by the SN estimates. For each combination of the parameters, 1000 iterations were performed. We applied the mvtnorm and sn packages to generate random samples from the multivariate normal and SN distributions, respectively.

In each iteration, we obtained parameter estimates based on both scenarios, and then the mean and the standard deviation (SD) of the Normal and SN estimators were calculated. The summary results are presented in Tables 1–4.

Table 1. Mean and SD of each estimator for both cases with $n = 150$, $\sigma^2 = 0.4^2$ and $\sigma_1^2 = 0.5^2$.

| Parameter | True Value | Normal | | SN | |
		Mean	SD	Mean	SD
β_1	-1	-1.0031	0.0409	-0.9990	0.0385
β_2	2	1.9992	0.0411	2.0019	0.0406
σ^2	0.16	0.1871	0.0233	0.1580	0.0222
σ_1^2	0.25	0.2198	0.0611	0.2455	0.0575
λ_1	-3	-	-	-3.1239	0.3321
λ_2	0	-	-	-0.0083	0.0963
λ_3	3	-	-	2.9146	0.3275

Table 2. Mean and SD of each estimator for both cases with $n = 300$, $\sigma^2 = 0.4^2$ and $\sigma_1^2 = 0.5^2$.

| Parameter | True Value | Normal | | SN | |
		Mean	SD	Mean	SD
β_1	-1	-0.9989	0.0298	-1.0012	0.0268
β_2	2	1.9978	0.0297	2.0001	0.0274
σ^2	0.16	0.1893	0.0167	0.1587	0.0160
σ_1^2	0.25	0.2212	0.0389	0.2505	0.0433
λ_1	-3	-	-	-2.9564	0.2249
λ_2	0	-	-	0.0015	0.0635
λ_3	3	-	-	3.0650	0.2268

Table 3. Mean and SD of each estimator for both cases with $n = 150$, $\sigma^2 = 0.8^2$ and $\sigma_1^2 = 0.9^2$.

		Normal		SN	
Parameter	**True Value**	**Mean**	**SD**	**Mean**	**SD**
β_1	-1	-0.9995	0.0837	0.9988	0.0800
β_2	2	1.9989	0.0820	1.9980	0.0778
σ^2	0.64	0.7462	0.0942	0.6204	0.0868
σ_1^2	0.81	0.6872	0.2089	0.8138	0.1884
λ_1	-3	-	-	-2.9000	0.6506
λ_2	0	-	-	0.0056	0.2007
λ_3	3	-	-	2.9032	0.6644

**Table 4.** Mean and SD of each estimator for both cases with $n = 300$, $\sigma^2 = 0.8^2$ and $\sigma_1^2 = 0.9^2$.

		Normal		SN	
Parameter	**True Value**	**Mean**	**SD**	**Mean**	**SD**
β_1	-1	-0.9970	0.0576	-0.9982	0.0545
β_2	2	2.0019	0.0579	1.9981	0.0547
σ^2	0.64	0.7580	0.0670	0.6330	0.0644
σ_1^2	0.81	0.6883	0.1336	0.8059	0.1432
λ_1	-3	-	-	-3.0679	0.2716
λ_2	0	-	-	-0.0049	0.1386
λ_3	3	-	-	2.9507	0.2694

　　It can be seen from Tables 1–4, that the ML estimates of fixed effects were unbiased for both cases. However, the SN estimates had a lower dispersion compared to the Normal estimates. Both cases showed better results by increasing the sample size, but the SN estimates performed better. Increasing the values of variance components did not have much effect on the biases, but their SDs increased, and the precision of the estimates decreased. In addition, we observed that the Normal estimates of variance components were biased and hade larger SDs. In contrast, the SN estimates were unbiased and had smaller SDs. As was seen, increasing the values of variance components had a much worse effect on the biases and precision of the Normal estimates of the variance components. However, for the SN estimates, there were some rises in SDs, which was completely natural, but we still had unbiasedness. Furthermore, in both aspects (bias and SD), the SN estimates had better performance concerning the Normal estimates. For these, increasing the sample size did not affect demolishing biases; however, it decreased the SD to some extent. Nevertheless, it decreased bias and SD for the SN estimates. The simulation results in Tables 1–4 showed that the skewness estimates were unbiased. These estimates were made worse by increasing variability in the data and made better by increasing the sample size (see Figures 1 and 2). Furthermore, the histograms in Figures 1 and 2 indicate that the distribution of the estimators, even for small sample sizes, was approximately normal.
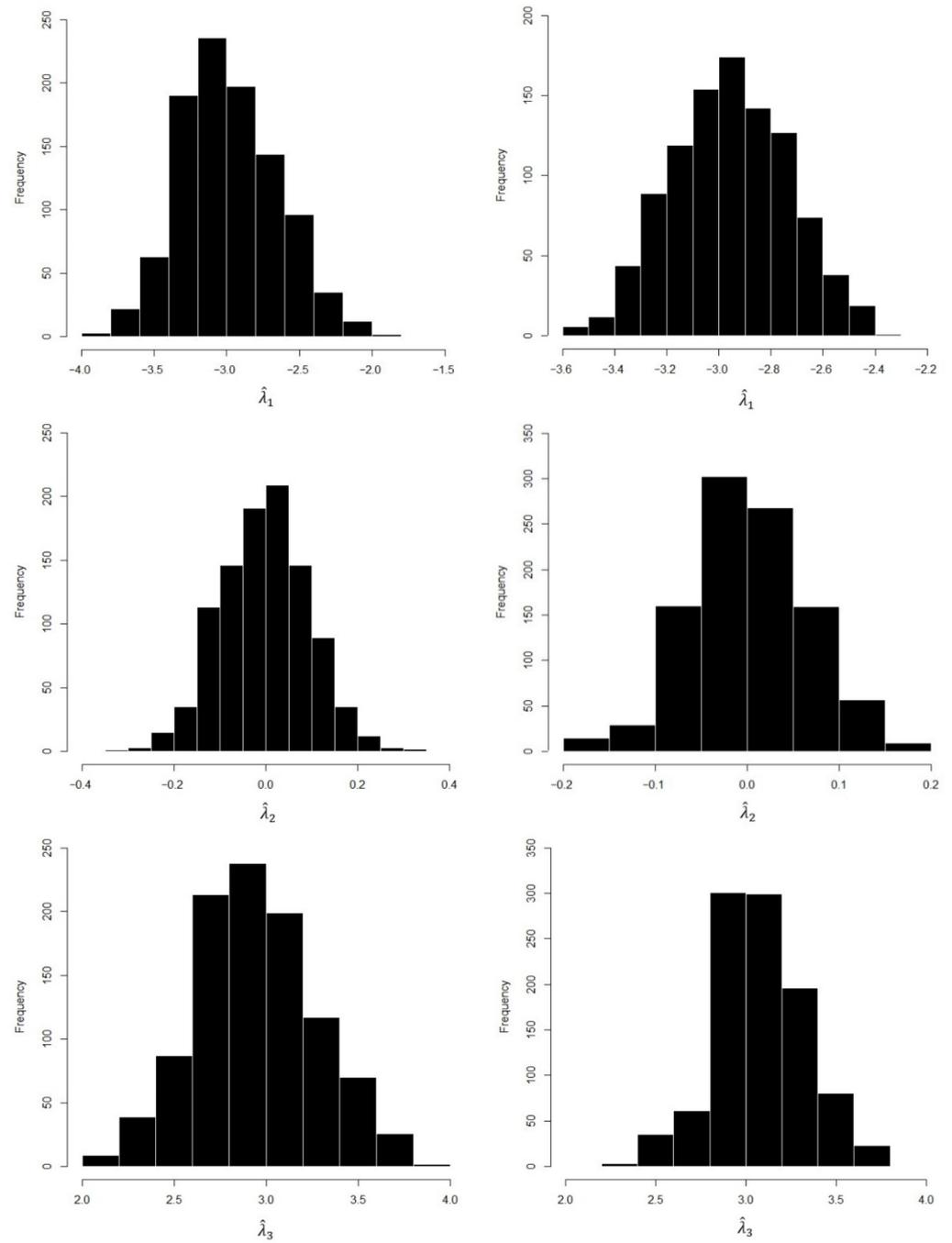
Figure 1. The histograms of the estimators of λ_1, λ_2 and λ_3 for $\sigma_1^2 = 0.5^2$ and $\sigma^2 = 0.4^2$. The left histograms are for $n = 150$ and the right histograms are for $n = 300$.
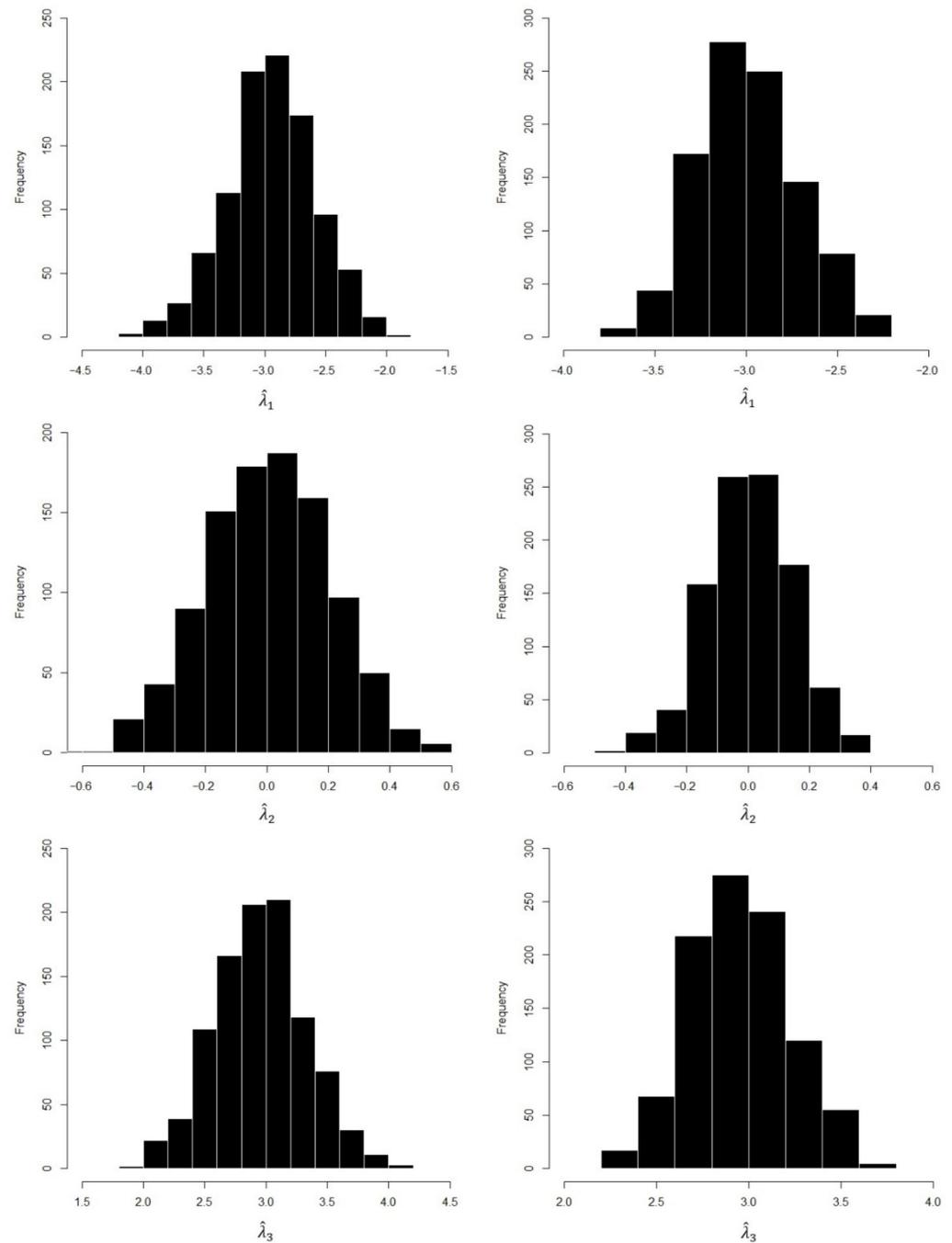
Figure 2. The histograms of the estimators of λ_1, λ_2 and λ_3 for $\sigma_1^2 = 0.8^2$ and $\sigma^2 = 0.9^2$. The left histograms are for $n = 150$ and the right histograms are for $n = 300$.

5.2. An Example

The metallic oxide data given by Fellner [2] was used to illustrate the usefulness of our method in applications. Following Fellner [2], a nested model was fitted with raw material for fixed effects, while the random effects were lot, sample, chemist, and analysis. We considered normal and SN distributions for the random errors in this model for comparative purposes, and both assumptions (heteroscedasticity and homoscedasticity) in the random-effect factors were considered. Therefore, we had four scenarios:

(a) Normal random errors and homoscedasticity in random-effect factors.
(b) SN random errors and homoscedasticity in random-effect factors.
(c) Normal random errors and heteroscedasticity in random-effect factors.

(d)   SN random errors and heteroscedasticity in random-effect factors.

The estimates of the parameters and some model selection criteria, including the log-likelihood values ($l(\boldsymbol{\theta})$), AIC, and BIC, are given in Table 5. In all cases, the Type 1 mean was approximately the same. The Type 2 mean was the same in the Normal cases (homoscedasticity and heteroscedasticity). This was approximately true in the SN cases as well. Under both assumptions in random-effect factors, the Type 2 mean in the SN case was greater than for the Normal case. In addition, the variance of the error term (analysis variance) and the lot variance were reduced in the SN case compared to the Normal case. As we expected, the model selection criteria showed that the fitted model under heteroscedasticity and SN random errors had the best fit.

Table 5. Metallic oxide summary results for four scenarios.

Parameter	(a)	(b)	(c)	(d)
Type 1 mean	3.865	3.858	3.863	3.856
Type 2 mean	3.064	3.107	3.064	3.134
Lot variance			0.607	0.565
Sample variance	0.175	0.123	0.043	0.044
Chemist variance			0.032	0.033
Analysis variance	0.045	0.041	0.043	0.037
Skewness parameter	-	−4.396	-	−3.674
$l(\boldsymbol{\theta})$	−122.762	−118.602	−91.322	−88.176
AIC	253.524	247.204	194.644	190.352

We continued by conducting the local influence approach to detect influential observations based on $M(0)$ for this dataset. Figures 3–6 depict, respectively, the index plots of $M(0)$ for all scenarios under perturbations of the response variable, the dispersion matrix of the errors, and the columns of matrix X. In all the perturbation schemes, we used $c^* = 3$ to construct the benchmarks. Furthermore, the influential observations are numbered in all figures.

Figure 3 indicates that observation #192 stood out as the most influential of the four scenarios under response variable perturbation. There were some suspicious points in the three first scenarios that were regular points in scenario 4. It could be claimed that the effects of these points in this scenario were controlled as the most flexible model in this work. Under the perturbation of the dispersion matrix of the errors in Figure 4, observation #192 again appeared as the most influential. As can be seen, the effects of other observations in scenario 4 were lower when compared with other scenarios. For the perturbation of the columns of matrix X, as depicted in Figures 5 and 6, the results were approximately the same. In these schemes, observation #192 seemed to be the most influential again. Furthermore, for the two first scenarios (the models under homoscedasticity), observation #191 was also influential. Except for the two last scenarios (the models under heteroscedasticity), observation #234 was an influential case. Generally, in all figures, suspicious points in the less flexible models were more numerous than in the more flexible models. In other words, when we used a more flexible model, we decreased the effects of anomalous observations in our statistical results.
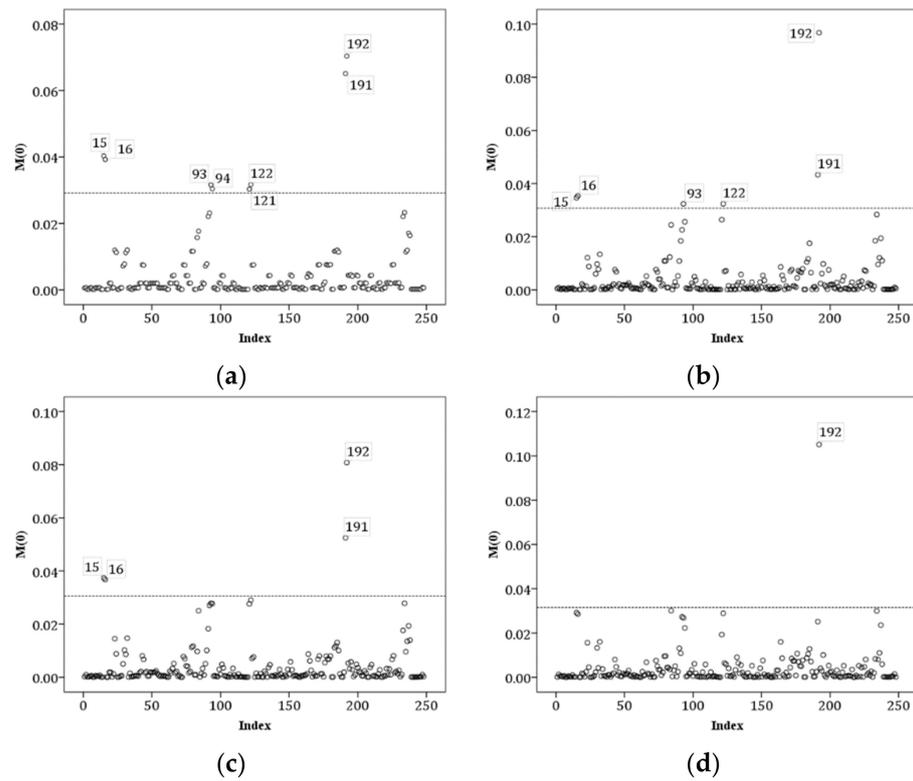
Figure 3. Index plots of $M(0)$ for the perturbation of the response variable for the four scenarios. Dotted lines show the yardstick for $M(0)$ with $c^* = 3$.
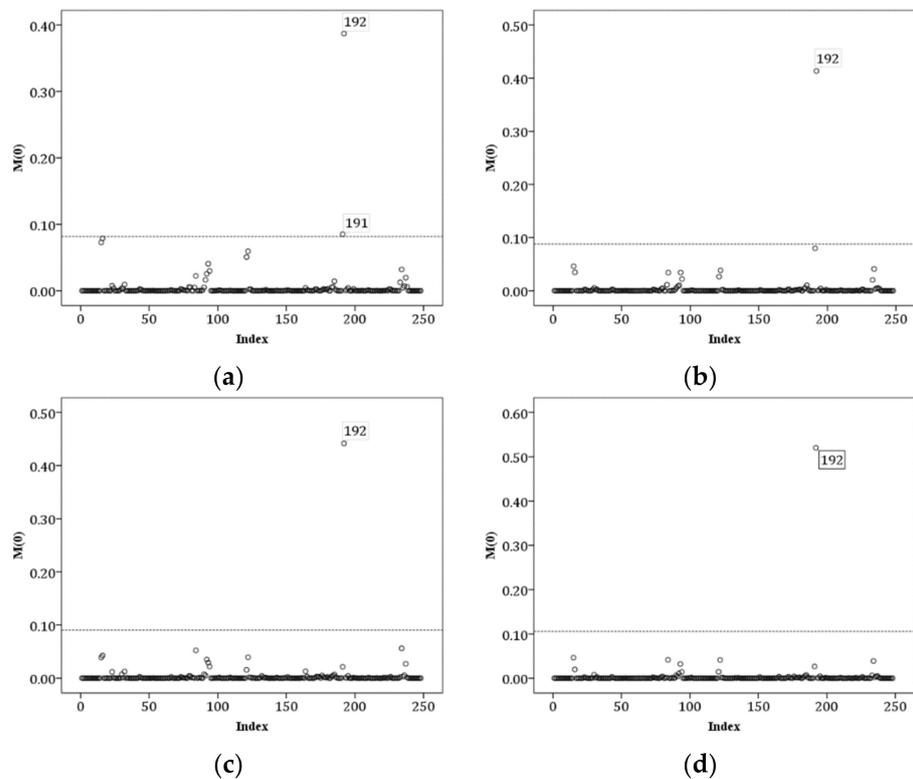
Figure 4. Index plots of $M(0)$ for the perturbation of the dispersion matrix of the errors for the four scenarios. Dotted lines show the yardstick for $M(0)$ with $c^* = 3$.
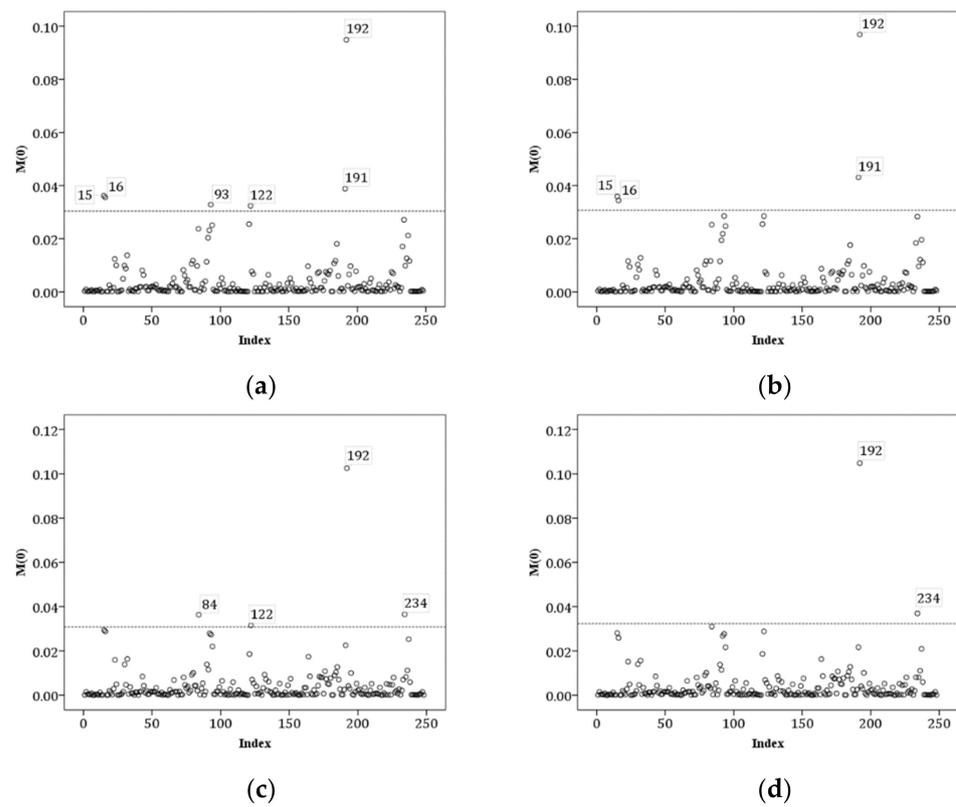
Figure 5. Index plots of $M(0)$ for the perturbation of the first column of X for the four scenarios. Dotted lines show the yardstick for $M(0)$ with $c^* = 3$.
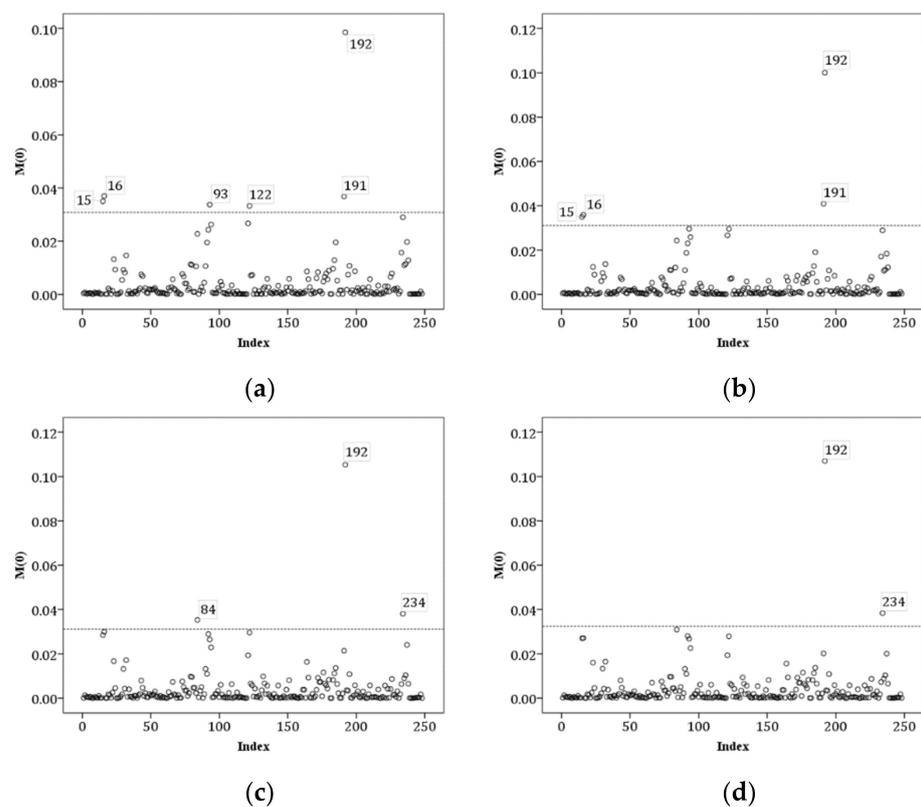
Figure 6. Index plots of $M(0)$ for the perturbation of the second column of X for the four scenarios. Dotted lines show the yardstick for $M(0)$ with $c^* = 3$.

6. Conclusions

The study of a multifactor normal LMM under heteroscedasticity was done by authors such as [1,2]. They showed that the ML method for estimation of the parameters performed well and the estimates had good properties such as consistency and asymptotic normal distribution. When the normality test of model errors or random effects was rejected, the derived estimates did not lead to satisfactory results. Therefore, LMMs (only with a random-effect factor) based on skewed distributions were presented by several authors. When the normality assumption did not hold for the model errors or the random-effect factor, these models performed well in comparison to those under the normality assumption. Additionally, diagnostic analyses showed these models decreased the effect of outliers. Recent research showed that new generalized skewed distributions usually have a better fit than simpler skewed distributions (see e.g., [38,39]). Clearly, using these distributions in LMMs can be also considered (see e.g., [40]), but, as mentioned before, the complexity of calculations in complicated models makes a case for using the simple but flexible SN distribution (See using SN distribution in some new complicated models [13,14]). Therefore, we considered SN distribution for the model errors in the multifactor LMM under heteroscedasticity in random-effect factors. Our main goals involved parameter estimation and the local influence method for the multifactor SN–LMM under heteroscedasticity in random-effect factors. At first, we expanded an EM-based algorithm, as many in the literature proposed, to estimate the model parameters. We also obtained a closed form to estimate variance components using this method. Then, we applied Zhu and Lee's approach to extend the local influence method to this model. Empirical studies, a simulation study and a real data example, were carried out to see the behavior of our estimators. Our findings followed previous results in this field. The simulation results—-consistency, low dispersion, and asymptotic normal distribution—-showed that the estimators performed well, even for finite sample sizes. It was also observed that ignorance of skewness when the error model followed from a skewed distribution like SN made unsuitable outcomes. Finally, through a real example, it was shown that taking into account both heteroscedasticity in random-effect factors and election a skewed distribution for random errors in the fitted model improved statistical results in comparison to other works that considered at most one of them. Additionally, in this case, we observed the robustness of the ML estimators through the local influence method. Finally, any skewed distributions contained symmetrical distributions in special cases. When the assumption of symmetry held for random variables in a sensitivity analysis of any model, skewed distributions were not recommended due to additional parameter costs. Extending this work when both random-effect factors and model errors have SN distribution is theoretically and computationally hard, but that will be our goal in a subsequent work. Moreover, generalized skewed distributions for the model errors or the random-effect factors in the multifactor LMM, along with diagnostic measures, are proposed for future work.

Author Contributions: Conceptualization, Z.N. and M.R.M.; Formal analysis, Z.N. and K.Z.; Investigation, Z.N., M.R.M., S.S. and A.M.; Methodology, Z.N., K.Z. and M.R.M.; Project administration, K.Z., M.R.M. and A.M.; Resources, M.R.M.; Software, M.R.M. and A.M.; Supervision, K.Z., M.R.M. and A.M.; Validation, S.S.; Visualization, A.M.; Writing—original draft, Z.N., K.Z., M.R.M. and S.S.; Writing—review & editing, A.M. All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

EM   Expectation Maximization
LMM   Linear Mixed Model
SN   Skew-Normal
ML   Maximum Likelihood
SD   Standard Deviation
AIC   The Akaike Information Criterion
BIC   The Bayesian Information Criterion

Appendix A.

Appendix A.1. ML Prediction of Random Effects

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ and $\boldsymbol{b} = \left(\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top, \ldots, \boldsymbol{b}_m^\top\right)^\top$ be vectors of observed responses and the random effects, respectively, and z^* be a missing observation. It follows from (6) that the conditional complete log-likelihood function of $\boldsymbol{b}$ given $\boldsymbol{\theta}$, $\boldsymbol{X}$ and $\boldsymbol{y}_c^* = \left(\boldsymbol{y}^\top, z^*\right)^\top$ may be expressed as

$$
\begin{aligned}
&l_c^*(\boldsymbol{b}; \boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{y}_c^*) \\
&= C - \frac{q+1}{2} \log(\sigma^2) - \frac{1}{2} \log(|\boldsymbol{\Sigma}\boldsymbol{T}|) \\
&\quad - \frac{1}{2\sigma^2} \left(\boldsymbol{b} - \boldsymbol{\Sigma}\boldsymbol{U}^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)^\top (\boldsymbol{\Sigma}\boldsymbol{T})^{-1} \left(\boldsymbol{b} - \boldsymbol{\Sigma}\boldsymbol{U}^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) \\
&\quad - \frac{1}{2\sigma^2} \left(z^* - \boldsymbol{\lambda}_\varepsilon^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{U}\boldsymbol{b})\right)^2 - \log(\alpha),
\end{aligned}
$$

where C does not depend on unknown parameters. If $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{b}}^{(r)}$ are the estimates in the rth iteration, the expected conditional complete log-likelihood function would be

$$
\begin{aligned}
&Q^*\left(\boldsymbol{b} \mid \hat{\boldsymbol{\theta}}^{(r)}, \hat{\boldsymbol{b}}^{(r)}\right) \\
&= E\left[l_c^*(\boldsymbol{b}; \boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{y}_c^*) \mid \hat{\boldsymbol{\theta}}^{(r)}, \hat{\boldsymbol{b}}^{(r)}\right] \\
&\propto -\frac{q+1}{2} \log(\sigma^2) - \frac{1}{2} \log(|\boldsymbol{\Sigma}\boldsymbol{T}|) \\
&\quad - \frac{1}{2\sigma^2} \left\{ \boldsymbol{b}^\top \left[(\boldsymbol{\Sigma}\boldsymbol{T})^{-1} + \boldsymbol{U}^\top \boldsymbol{\lambda}_\varepsilon \boldsymbol{\lambda}_\varepsilon^\top \boldsymbol{U} \boldsymbol{b} \right] \right. \\
&\qquad\qquad - 2\boldsymbol{b}^\top \left[\boldsymbol{U}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \hat{z^*}^{(r)} \boldsymbol{U}^\top \boldsymbol{\lambda}_\varepsilon + \boldsymbol{U}^\top \boldsymbol{\lambda}_\varepsilon \boldsymbol{\lambda}_\varepsilon^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right] \\
&\qquad\qquad + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \left(\boldsymbol{I}_n - \boldsymbol{V}^{-1} + \boldsymbol{\lambda}_\varepsilon \boldsymbol{\lambda}_\varepsilon^\top\right)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - 2\hat{z^*}^{(r)} \boldsymbol{\lambda}_\varepsilon^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&\qquad\qquad \left. + \hat{z^*}^{2(r)} \right\}
\end{aligned}
$$

Then, the M-step maximizes $Q^*\left(\boldsymbol{b} \mid \hat{\boldsymbol{\theta}}^{(r)}, \hat{\boldsymbol{b}}^{(r)}\right)$ with respect to $\boldsymbol{b}$. So, we solve the following equation:

$$
\begin{aligned}
&\frac{\partial Q^*\left(\boldsymbol{b} \mid \hat{\boldsymbol{\theta}}^{(r)}, \hat{\boldsymbol{b}}^{(r)}\right)}{\partial \boldsymbol{b}} \\
&= -\frac{1}{\sigma^2} \left[\left((\boldsymbol{\Sigma}\boldsymbol{T})^{-1} + \boldsymbol{U}^\top \boldsymbol{\lambda}_\varepsilon \boldsymbol{\lambda}_\varepsilon^\top \boldsymbol{U}\right) \boldsymbol{b} \right. \\
&\qquad\qquad \left. - \left(\boldsymbol{U}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \hat{z^*}^{(r)} \boldsymbol{U}' \boldsymbol{\lambda}_\varepsilon + \boldsymbol{U}^\top \boldsymbol{\lambda}_\varepsilon \boldsymbol{\lambda}_\varepsilon^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right) \right] = 0.
\end{aligned}
$$

Appendix A.2. ML Estimation of Variances of Random Effects

To find a solution for γ_i and hence for σ_i^2, $i = 1, \ldots, m$, in Equation (9) using the relations $|\boldsymbol{V}| = \left| \boldsymbol{I}_q + \boldsymbol{U}^\top \boldsymbol{U} \boldsymbol{\Sigma} \right| = |\boldsymbol{T}|^{-1}$, $\partial \boldsymbol{V}^{-1}/\partial \gamma_i = -\boldsymbol{V}^{-1} \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{V}^{-1}$, $\partial \boldsymbol{\Sigma}/\partial \gamma_i = \text{diag}(0, \ldots, 0, \boldsymbol{I}_{q_i}, 0, \ldots, 0)$ and

$$\partial \log \left| \boldsymbol{I}_q + \boldsymbol{U}^\top \boldsymbol{U} \boldsymbol{\Sigma} \right| / \partial \gamma_i = \text{tr}\left(\boldsymbol{T} \boldsymbol{U}^\top \boldsymbol{U} \frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i} \right) = \gamma_i^{-1}(q_i - \text{tr}(\boldsymbol{T}_{ii})),$$

we have

$$
\begin{aligned}
\frac{\partial Q\left(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r)} \right)}{\partial \gamma_i} \\
= -\tfrac{1}{2} \gamma_i^{-1}(q_i - \text{tr}(\boldsymbol{T}_{ii})) \\
+ \tfrac{1}{2\sigma^2\delta^2} \left[z^{2(r)} - 2\hat{z}^{(r)} \boldsymbol{h}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \left(\boldsymbol{h}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right)^2 \right] \boldsymbol{h}^\top \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{h} \\
+ \tfrac{1}{2\sigma^2} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1} \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \tfrac{1}{2\delta} \boldsymbol{h}^\top \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{h} \\
+ \tfrac{1}{\sigma^2\delta} \left[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{h} \boldsymbol{h}^\top \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \hat{z}^{(r)} \boldsymbol{h}^\top \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right]
\end{aligned}
$$

Then, by solving the equations

$$\frac{\partial Q\left(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r)} \right)}{\partial \gamma_i} = 0, \, i = 1, \ldots, m$$

the ML estimates of γ_i's are obtained. Thus, the ML estimate of σ_i^2 is taken to satisfy

$$\hat{\sigma_i^2} = \hat{\sigma}^2 \hat{\gamma}_i, \, i = 1, \ldots, m.$$

References

1. Harvill, D.A. Maximum likelihood approaches to variance component estimation and related problems (with discussion). *J. Ame. Stat. Assoc.* **1977**, *72*, 320–340. [CrossRef]
2. Fellner, W.H. Robust estimation of variance components. *Technometrics* **1986**, *28*, 51–60. [CrossRef]
3. Khuri, A.I.; Methew, T.; Sinha, B.K. *Statistical Tests for Mixed Linear Models*; John Wiley: New York, NY, USA, 1998.
4. Wu, M.X.; Yu, K.F.; Liu, A.Y.; Ma, T.F. Simultaneous optimal estimation in linear mixed models. *Metrika* **2012**, *75*, 471–489. [CrossRef]
5. Zhang, D.; Davidian, M. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **2001**, *57*, 795–802. [CrossRef] [PubMed]
6. Pinheiro, J.C.; Liu, C.H.; Wu, Y.N. Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *J. Comput. Gr. Stat.* **2001**, *10*, 249–276. [CrossRef]
7. Ghidey, W.; Lesaffre, E.; Eilers, P. Smooth random effects distribution in a linear mixed model. *Biometrics* **2004**, *60*, 945–953. [CrossRef]
8. Lin, T.I.; Lee, J.C. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Stat. Med.* **2008**, *27*, 1490–1507. [CrossRef]
9. Lachos, V.H.; Dey, D.K.; Cancho, V.G. Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *J. Stat. Plan. Inference* **2009**, *139*, 4098–4110. [CrossRef]
10. Ye, R.D.; Wang, T.; Gupta, A.K. Distribution of matrix quadratic forms under skew-normal settings. *J. Multivar. Anal.* **2014**, *131*, 229–239. [CrossRef]
11. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distributions. *J. Roy. Stat. Soc. Ser. B* **1999**, *61*, 579–602. [CrossRef]
12. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Statist.* **1985**, *12*, 171–178.
13. Hosseini, F.; Karimi, O. Approximate pairwise likelihood inference in SGLM models with skew normal latent variables. *J. Comput. Appl. Math.* **2021**, *398*, 113692. [CrossRef]
14. Ju, Y.; Yang, Y.; Hu, M.; Dai, L.; Wu, L. Bayesian Influence Analysis of the Skew-Normal Spatial Autoregression Models. *Mathematics* **2022**, *10*, 1306. [CrossRef]
15. Verbeke, G.; Lesaffre, E. The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Comput. Stat. Data. Anal.* **1997**, *23*, 541–556. [CrossRef]

16. Tao, H.; Palta, M.; Yandell, B.S.; Newton, M.A. An estimation method for the semiparametric mixed effects model. *Biometrics* **1999**, *55*, 102–110. [CrossRef]
17. Ma, Y.; Genton, M.G.; Davidian, M. Linear mixed models with flexible generalized skew-elliptical random effects. In *SKEW-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004; pp. 339–358.
18. Arellano-Valle, R.B.; Bolfarine, H.; Lachos, V.H. Skew-normal linear mixed models. *J. Data Sci.* **2005**, *3*, 415–438. [CrossRef]
19. Lachos, V.H.; Ghosh, P.; Arellano-Valle, R.B. Likelihood based inference for skew-normal independent linear mixed models. *Stat. Sin.* **2010**, *20*, 303–322.
20. Kheradmandi, A.; Rasekh, A. Estimation in skew-normal linear mixed measurement error models. *J. Multivar. Anal.* **2015**, *136*, 1–11. [CrossRef]
21. Maleki, M.; Wraith, D.; Arellano-Valle, R.B. A flexible class of parametric distributions for Bayesian linear mixed models. *Test* **2019**, *28*, 543–564. [CrossRef]
22. Ferreira, C.S.; Bolfarine, H.; Lachos, V.H. Linear mixed models based on skew scale mixtures of normal distributions. *Commun. Stat. Simul. Comput.* **2020**, 1–21. [CrossRef]
23. Cook, R.D. Assessment of local influence (with discussion). *J. Roy. Stat. Soc. Ser. B* **1986**, *48*, 133–169. [CrossRef]
24. Zhu, H.; Lee, S. Local influence for incomplete-data models. *J. Roy. Stat. Soc. Ser. B* **2001**, *63*, 111–126. [CrossRef]
25. Montenegro, L.C.; Bolfarine, H.; Lachos, V.H. Influence diagnostics for a skew extension of the Grubbs measurement error model. *Comm. Stat. Simul. Comput.* **2009**, *38*, 667–681. [CrossRef]
26. Zeller, C.B.; Lachos, V.H.; Vilca, F.V. Influence diagnostics for Grubbs's model with asymmetric heavy-tailed distributions. *Stat. Pap.* **2014**, *55*, 671–690. [CrossRef]
27. Ferreira, C.S.; Lachos, V.H.; Bolfarine, H. Inference and diagnostics in skew scale mixtures of normal regression models. *J. Stat. Comput. Simul.* **2015**, *85*, 517–537. [CrossRef]
28. Massuia, M.B.; Cabral, C.R.B.; Matos, L.A.; Lachos, V.H. Inference diagnostics for student-t censored linear regression models. *Statistics* **2015**, *49*, 1074–1094. [CrossRef]
29. Bolfarine, H.; Montenegro, L.C.; Lachos, V.H. Influence diagnostics for skew-normal linear mixed models. *Sankhyā Indian J. Stat.* **2007**, *69*, 648–670.
30. Montenegro, L.C.; Lachos, V.H.; Bolfarine, H. Local influence analysis for skew-normal linear mixed models. *Commun. Stat. Theory Methods* **2009**, *38*, 484–496. [CrossRef]
31. Zeller, C.B.; Labra, F.V.; Lachos, V.H.; Balakrishnan, N. Influence analyses of skew-normal/independent linear mixed models. *Comput. Stat. Data Anal.* **2010**, *54*, 1266–1280. [CrossRef]
32. Dempster, A.; Laird, N.; Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. Ser. B* **1977**, *39*, 1–22. [CrossRef]
33. Johnson, N.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1994.
34. Lu, B.; Song, X. Local influence of multivariate probit latent variable models. *J. Multivar. Anal.* **2006**, *97*, 1783–1798. [CrossRef]
35. Lee, S.Y.; Xu, L. Influence analysis of nonlinear mixed-effects models. *Comput. Stat. Data Anal.* **2004**, *45*, 321–341. [CrossRef]
36. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: https://www.R-project.org/ (accessed on 1 November 2021).
37. Wang, N.; Lin, X.; Gutierrez, R.G.; Carroll, R.J. Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *J. Amer. Statist. Assoc.* **1998**, *93*, 249–261. [CrossRef]
38. Labra, F.V.; Garay, A.M.; Lachos, V.H.; Ortega, E.M.M. Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *J. Stat. Plan. Inference* **2012**, *142*, 2149–2165. [CrossRef]
39. Schumacher, F.L.; Dey, D.K.; Lachos, V.H. Approximate Inferences for Nonlinear Mixed Effects Models with Scale Mixtures of Skew-Normal Distributions. *J. Stat. Theory Pract.* **2021**, *15*, 60. [CrossRef]
40. Schumacher, F.L.; Lachos, V.H.; Matos, L.A. Scale mixture of skew-normal linear mixed models with within-subject serial dependence. *Stat. Med.* **2021**, *40*, 1790–1810. [CrossRef] [PubMed]