Skew-Normal Distribution in Multivariate Null Intercept Measurement Error Model

F. V. Labra[1,*], **R. Aoki**[2], **V. Garibay**[2], and **V.H. Lachos**[1]

[1]Departamento de Estatística, Universidade Estadual de Campinas,
Caixa Postal 6065–CEP 13083–859, Campinas, SP, Brasil

[2]Departamento de Matemática Aplicada e Estatística, ICMC, Universidade de São paulo,
Caixa Postal 668, São Carlos, São Paulo, Brasil

Abstract

In this paper we discuss inferential aspects of the multivariate null intercept measurement error model where the unobserved value of the covariate (latent variable) follows a skew-normal distribution. First, closed form expressions of the marginal likelihood, the score function and the observed information matrix of the observed quantities are presented allowing direct inference implementation. Then, we indicate how maximum likelihood estimators of the parameter vector may be obtained via the ECM algorithm. Additionally, an EM-type algorithm for evaluating the restricted maximum likelihood estimate under equality constraints on the regression coefficients is examined. In order to discuss some diagnostic techniques in this type of models, we derive the appropriate matrices to assess the local influence on the parameters estimate under different perturbation schemes. The results and methods are applied to a dental clinical trial presented in Hadgu and Koch (1999).

Key Words: *Skew-normal distribution; EM algorithm; Skewness; Multivariate null intercept model; Measurement error; Local influence.*

1 Introduction

The development of parametric families and the study of their properties have ever been a persistent theme of the statistical literature. A substantial part of this recent literature is broadly related to the skew-normal distribution, which represents a superset of the normal family and has a shape parameter that defines the direction of the asymmetry of the distribution. Advantages of using such general structures in practice, include easiness of interpretation, as well as estimation efficiency. Motivation was originated from real data sets presenting clear indication of skewness (not following the symmetric normal law) in diverse areas, such as, engineering,

* *Correspondence:* Filidor Vilca Labra, Departamento de Estatística, IMECC, Universidade de Campinas, Caixa Postal 6065, CEP 13083-859, São Paulo, SP, Brazil. E-mail: fily@ime.unicamp.br

medicine, psychology and agriculture, among others. In this paper we use a real data set from a dental clinical trial in which the outcome measurements (plaque index) are typically not symmetric so that it seems more adequate to consider a skew-normal distribution to describe the behavior of these measurements.

As discussed by Arellano-Valle et al. (2005), we say that a k-dimensional random vector $\mathbf{X}$ has a *standardized multivariate skew-normal distribution* with skewness vector $\boldsymbol{\lambda}$, namely $\mathbf{X} \sim SN_k(\boldsymbol{\lambda})$, if its probability density function (pdf) is given by

$$f_{\mathbf{X}}(\mathbf{x}) = 2\phi_k(\mathbf{x})\Phi_1(\boldsymbol{\lambda}^\top \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^k, \tag{1}$$

where, as usual, $\phi_k(.)$ and $\Phi_k(.)$ denote, respectively, the probability density function (pdf) and cumulative distribution function (cdf) of the $N_k(\mathbf{0}, \mathbf{I}_k)$ distribution. More generally, for the $N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$, we denote such functions by $\phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\Psi})$ and $\Phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\Psi})$, respectively. Now, as a location-scale extension of (1), we consider the distribution of $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}\mathbf{X}$, yielding the following pdf

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_k(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Psi})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Psi}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^k, \tag{2}$$

which will be denoted by $\mathbf{Y} \sim SN_k(\boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, with stochastic representation given by

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_k - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{with} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}}, \tag{3}$$

where $T_0 \sim N_1(0, 1)$ and $\mathbf{T}_1 \sim N_k(\mathbf{0}, \mathbf{I}_k)$ are independent, and " $\stackrel{d}{=}$ " meaning "distributed as". For more details on this approach, see Arellano–Valle and Genton (2005) and Arellano–Valle et al. (2005). Note that when $k = 1$ we obtain the univariate skew-normal distribution introduced by Azzalini (1985) and (3) is reduced to the stochastic representation obtained in Henze (1986).

Error-in-variables regression models constitute an attractive alternative to modeling many practical experimental problems, specially when the same responses are observed on the same units under different experimental conditions. A wide bibliography can be found in Fuller (1987) and Cheng and Van Ness (1999). Recently, Aoki et al. (2003) has discussed a multivariate symmetric null intercept error-in-variables regression model with a dependency structure between the response variables within the same group appropriate to longitudinal data studies. A generalization of that proposed model can be written as

$$\mathbf{X}_i = \mathbf{x}_i + \boldsymbol{\delta}_i, \tag{4}$$

$$\mathbf{y}_{k_i} = \mathbf{x}_i\beta_{k_i} + \boldsymbol{\epsilon}_{k_i}, \tag{5}$$

$i = 1, \ldots, p$, $k = 1, \ldots, m$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})^\top$, $\mathbf{y}_{k_i} = (y_{k_{i1}}, \ldots, y_{k_{in_i}})^\top$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{in_i})^\top$, $\boldsymbol{\epsilon}_{k_i} = (\epsilon_{k_{i1}}, \ldots, \epsilon_{k_{in_i}})^\top$, $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{in_i})^\top$. In the dental clinical data set, presented in Hadgu and Koch (1999), 105 volunteers with preexisting dental plaque were randomized to two experimental mouth rinses (A and B) or a control mouth rinse with double blinding and evaluated with respect to the dental plaque index at baseline ($\mathbf{X}_i$), after three months ($\mathbf{y}_{1_i}$) and after six months ($\mathbf{y}_{2_i}$) from the baseline with the use of the corresponding mouth rinses (A, B or control C). Considering this data set, we have $i = 1, 2, 3$ representing the control mouth rinse, the experimental mouth rinse A and experimental mouth rinse B, respectively. We have $m = 2$, with $k = 1$ ($k = 2$) representing three months from the baseline (six months from the baseline). As the covariates are measured imprecisely, Aoki et al. (2003) proposed the use of the measurement error model and since null pretest dental plaque index implies null expected post test values, the null intercept model was considered. To account for the possible dependence of the within subject measurements a structural model was considered. Typically, it is assumed that $\delta_{ij} \overset{iid}{\sim} N_1(0, \sigma^2)$, $\epsilon_{kij} \overset{ind}{\sim} N_1(0, \sigma^2_{e_i})$, δ_{ij} and ϵ_{kij} are not correlated and independent of $x_{ij} \overset{iid}{\sim} N_1(\mu_x, \sigma^2_x)$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$ $k = 1, \ldots, m$. As the observed quantities are not symmetric, the dental plaque index in the baseline (x_{ij}) may require data transformation in order to be better approximated by the normal distribution. Azzalini and Dalla–Valle (1996) give several reasons to avoid variables transformation if a more suitable theoretical model is found. Thus, the main objective of this paper is the study of inference and influence diagnostics in the multivariate null intercept measurement error regression model defined in (4)–(5) with the assumption that the unknown quantity x_{ij} (latent variable) follows a univariate skew-normal distribution, implying that the observation vector $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, follows a multivariate skew-normal distribution within the setup defined in (2).

Influence diagnostic is an important step in the analysis of a data set, as it provides us indication of bad model fitting or influential observations. This analysis has received a great deal of attention since the paper by Cook (1977). Typically the analysis is based on the case-weight perturbation scheme where the case (observation) is either deleted or retained. Cook (1986) proposed a method of assessing the local influence of minor perturbations of a statistical model. Since then several papers have been written with respect to the local influence approach which is considered by some authors in measurement error regression models. For example, Kelly (1984) derived the influence functions for the model parameters. Wellman and Gunst (1991) shows the need for influence diagnostic in such models using the influence functions. Abdullah (1995) applied some diagnostic methods in regression analysis to the functional model. Kim (2000) applied local influence methods in structural measurement error models. In a normal context, Aoki, et al. (2001) proposed a null intercept measurement error regression models with additional within-subjects correlation structure to analyze data from pretest/posttest. Labra et al. (2005) applied the local influence methodology in that model considering the Student-t distribution. More recently, Lachos et al. (2006) applied the local influence method in the skew-normal null intercept measurement error model without considering longitudinal structure, i.e, $m = 1$. Here, we extend those results to a model that allows the longitudinal structure.

The paper is organized as follows. In Section 2 the multivariate null intercept measurement error model under the skew-normal distribution is defined (SN-MEM, hereafter). Moreover, the score function and the observed information matrix are derived algebraically allowing a direct implementation of inferences. On the other hand, these matrices can be obtained numerically by using the marginal likelihood function and statistical software, as the Ox program. In Section 3 the ECM-algorithm for maximum likelihood estimation (MLE) is developed by exploring statistical properties of the considered model, yielding closed form expressions for the equations in the CM-step. Then, we present an EM-type algorithm for evaluating the restricted maximum likelihood estimate under equality constraints on the regression coefficients. Also, we discuss hypothesis testing considering the likelihood ratio, score and Wald test statistics with special emphasis in the regression coefficients. Section 4 contains the main concepts of local influence and the related concepts of diagnostics. Considering the model proposed in Section 2, we derive the appropriate matrices to obtain the normal curvature under various perturbation schemes. Finally, in Section 5 applications of the results and methods are illustrated with a numerical example and in Section 6 some final conclusions are discussed.

2 The skew-normal multivariate null intercept measurement error model

To specify the null intercept measurement error model in the multivarite skew-normal class, notice that we can write the linear model defined in (4)-(5) as,

$$\mathbf{z}_{ij} = \boldsymbol{\beta}_{0i} x_{ij} + \boldsymbol{\zeta}_{ij}, \tag{6}$$

where $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$ is the vector of observations; $\boldsymbol{\beta}_{0i} = (1, \boldsymbol{\beta}_i^\top)^\top$, with $\boldsymbol{\beta}_i = (\beta_{1_i}, \ldots, \beta_{m_i})^\top$ is the vector of the regression parameters and $\boldsymbol{\zeta}_{ij} = (\delta_{ij}, \epsilon_{1_{ij}}, \ldots, \epsilon_{m_{ij}})^\top$ is the vector of random errors . Here, the linear model defined in (6) will be called SN-MEM if

$$\boldsymbol{\zeta}_{ij} \overset{ind}{\sim} N_{m+1}(\mathbf{0}, \mathrm{D}(\boldsymbol{\phi}_i)) \quad \text{and} \quad x_{ij} \overset{iid}{\sim} SN_1(\mu_x, \phi_x, \lambda_x), \tag{7}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$, where $\mathrm{D}(\boldsymbol{\phi}_i)$ is a diagonal matrix with the diagonal elements given by $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_i = (\sigma_u^2, \sigma_{e_i}^2, \ldots, \sigma_{e_i}^2)^\top$ with dimension $(m+1) \times 1$. The above model considers, for instance, that in the case of the Hadgu and Koch (1999) data set the dental plaque index may not be symmetrically distributed in the population. On the other hand, the errors $\boldsymbol{\zeta}_{ij}$, are related to measurement errors so that it is expected to be normally distributed. The asymmetry parameter λ_x incorporates skewness in the latent variable x_{ij} and consequently in the observed quantities $\mathbf{z}_{ij}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, which can be shown to have marginally a $m+1$-variate skew-normal distribution. If $\lambda_x = 0$, then the asymmetric model reduces to the multivariate normal null intercept measurement error model (N-MEM), so this construction allows a continuous variation from normality to non-normality. Note from (3) that, the regression set up defined in (6)-(7) can be written hierarchically as

$$\begin{aligned} \mathbf{z}_{ij} \mid x_{ij} &\overset{ind}{\sim} N_{m+1}(\boldsymbol{\beta}_{0i} x_{ij}, \mathrm{D}(\boldsymbol{\phi}_i)), && (8)\\ x_{ij} \mid T_{ij} = t_{ij} &\overset{ind}{\sim} N_1(\mu_x + \phi_x^{1/2} \delta_x t_{ij}, \phi_x(1 - \delta_x^2)), && (9)\\ T_{ij} &\overset{iid}{\sim} HN_1(0, 1), && (10) \end{aligned}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$, all independent, where $HN_1(0, 1)$ denote the standardized univariate half-normal distribution and $\delta_x = \lambda_x/(1+\lambda_x^2)^{1/2}$. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\sigma}_e^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x)^\top \in \mathbb{R}^{(m+1)p+4}$, with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top)^\top$, $\boldsymbol{\beta}_i = (\beta_{1_i}, \ldots, \beta_{m_i})^\top$, $\boldsymbol{\sigma}_e^2 = (\sigma_{e_1}^2, \ldots, \sigma_{e_p}^2)^\top$ is based on the marginal distribution for the response $\mathbf{z}_{ij}$, which is given by

$$f_{\mathbf{Z}_{ij}}(\mathbf{z}_{ij}|\boldsymbol{\theta}) = 2\phi_{m+1}(\mathbf{z}_{ij}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\Phi_1(\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)), \tag{11}$$

$i = 1, \ldots, p,\ j = 1, \ldots, n_i$, i.e., $\mathbf{z}_{ij} \overset{ind}{\sim} SN_{m+1}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i)$, where

$$\boldsymbol{\mu}_i = \boldsymbol{\beta}_{0i}\mu_x,\ \ \boldsymbol{\Sigma}_i = \mathrm{D}(\boldsymbol{\phi}_i) + \phi_x\boldsymbol{\beta}_{0i}\boldsymbol{\beta}_{0i}^\top,\ \ \bar{\boldsymbol{\lambda}}_i = \frac{\lambda_x\phi_x\boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\beta}_{0i}}{\sqrt{\phi_x + \lambda_x^2\Lambda_i}},\ \text{with}\ \Lambda_i = \frac{\phi_x}{1 + \phi_x\boldsymbol{\beta}_{0i}^\top\mathrm{D}^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}}. \tag{12}$$

It follows that the log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{z} = (\mathbf{z}_{11}^\top, \ldots, \mathbf{z}_{1n_1}^\top, \ldots, \mathbf{z}_{p1}^\top, \ldots, \mathbf{z}_{pn_p}^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\ell_{ij}(\boldsymbol{\theta}), \tag{13}$$

where $\ell_{ij}(\boldsymbol{\theta}) = \log(2) - \frac{(m+1)}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \dfrac{1}{2}g_{ij} + \log(K_{ij})$, with

$$g_{ij} = (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i), K_{ij} = \Phi_1(\bar{\boldsymbol{\lambda}}_i^\top\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)) \tag{14}$$

and $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, $\bar{\boldsymbol{\lambda}}_i$ as in (12).

2.1 Score function

From (11), we have after some algebraic manipulations that the expression of K_{ij} given in (14) can be written as $K_{ij} = \Phi_1(A_i a_{ij})$, with $A_i = \dfrac{\lambda_x\Lambda_i}{\sqrt{\phi_x + \lambda_x^2\Lambda_i}}$, $a_{ij} = (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$, $\Lambda_i = \dfrac{\phi_x}{c_i}$ and $c_i = 1 + \phi_x\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. The score function is given by

$$U(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\partial}{\partial\boldsymbol{\theta}}\sum_{i=1}^{p}\sum_{j=1}^{n_i}l_{ij}(\boldsymbol{\theta}) = \left(\left(\sum_{j=1}^{n_i}U_{ij}(\boldsymbol{\theta}_1)\right)^\top, \left(\sum_{i=1}^{p}\sum_{j=1}^{n_i}U_{ij}(\boldsymbol{\theta}_2)\right)^\top\right)^\top, \tag{15}$$

where $U_{ij}(\boldsymbol{\theta}_1) = \dfrac{\partial\ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1} = (U_{ij}(\boldsymbol{\beta}_1), \cdots, U_{ij}(\boldsymbol{\beta}_p), U_{ij}(\sigma_{e_1}^2), \cdots, U_{ij}(\sigma_{e_p}^2))^\top$ and $U_{ij}(\boldsymbol{\theta}_2) = \dfrac{\partial\ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2} = (U_{ij}(\sigma_u^2), U_{ij}(\mu_x), U_{ij}(\phi_x), U_{ij}(\lambda_x))^\top$, with

$$U_{ij}(\boldsymbol{\gamma}) = \frac{\partial\ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}} = -\frac{1}{2}\frac{\partial\log|\boldsymbol{\Sigma_i}|}{\partial\boldsymbol{\gamma}} - \frac{1}{2}\frac{\partial g_{ij}}{\partial\boldsymbol{\gamma}} + \frac{\partial\log K_{ij}}{\partial\boldsymbol{\gamma}}, \tag{16}$$

where $\dfrac{\partial\log K_{ij}}{\partial\boldsymbol{\gamma}} = W_{\Phi_1}(A_i a_{ij})\left[A_i\dfrac{\partial a_{ij}}{\partial\boldsymbol{\gamma}} + a_{ij}\dfrac{\partial A_i}{\partial\boldsymbol{\gamma}}\right]$, with $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $u \in \mathbb{R}$, $\boldsymbol{\gamma} = \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_p, \sigma_{e_1}^2, \cdots, \sigma_{e_p}^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. Analytical expressions for the above derivatives are given in the Appendix A.

2.2 The observed information matrix

The matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{L} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\ell_{ij}(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{j=1}^{n_i}\frac{\partial^2 \ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\top} & \sum_{j=1}^{n_i}\frac{\partial^2 \ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top} \\ \sum_{j=1}^{n_i}\frac{\partial^2 \ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_1^\top} & \sum_{i=1}^{p}\sum_{j=1}^{n_i}\frac{\partial^2 \ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} \end{pmatrix}, \quad (17)$$

with $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^\top, \cdots, \boldsymbol{\beta}_p^\top, \sigma_{e1}^2, \cdots, \sigma_{ep}^2)^\top$ and $\boldsymbol{\theta}_2 = (\sigma_u^2, \mu_x, \phi_x, \lambda_x)$. From (16) it follows that the observed, per element, information matrix is given by

$$\mathbf{J}_{ij} = -\left[\frac{\partial^2 \ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top}\right], \quad (18)$$

where $\frac{\partial^2 \ell_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} = -\frac{1}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} - \frac{1}{2}\frac{\partial^2 g_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} + \frac{\partial^2 \log K_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top}$, with

$$\begin{aligned} \frac{\partial^2 \log K_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} &= W_{\Phi_1}(A_i a_{ij})\left[\frac{\partial A_i}{\partial\boldsymbol{\gamma}}\frac{\partial a_{ij}}{\partial\boldsymbol{\tau}^\top} + A_i\frac{\partial^2 a_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} + \frac{\partial a_{ij}}{\partial\boldsymbol{\gamma}}\frac{\partial A_i}{\partial\boldsymbol{\tau}^\top} + a_{ij}\frac{\partial^2 A_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top}\right] \\ &\quad + \Delta_{\Phi_1}(A_i a_{ij})\left[A_i\frac{\partial a_{ij}}{\partial\boldsymbol{\gamma}} + a_{ij}\frac{\partial A_i}{\partial\boldsymbol{\gamma}}\right]\left[A_i\frac{\partial a_{ij}}{\partial\boldsymbol{\tau}^\top} + a_{ij}\frac{\partial A_i}{\partial\boldsymbol{\tau}^\top}\right], \end{aligned}$$

$\Delta_{\Phi_1}(u) = W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, $\boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_p, \sigma_{e1}^2, \cdots, \sigma_{ep}^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x$. The results of this derivatives are given in the Appendix B. Asymptotic confidence intervals and testing on the ML estimates can be obtained considering this matrix. In the next section we discuss the MLE of the vector of parameters $\boldsymbol{\theta}$ using the ECM algorithm. Note that algorithms such as Newton-Raphson can be implemented using the above results.

3 MLE via the ECM-algorithm and the hypothesis testing

3.1 Maximum likelihood estimation

The EM algorithm (Dempster, Laird and Rubin (1977)) is a popular iterative algorithm for ML estimation in models with incomplete data. More specifically, let $\mathbf{z}$ denote the observed data and $\mathbf{s}$ denote the missing data. The complete data $\mathbf{r} = (\mathbf{z}, \mathbf{s})$ is $\mathbf{z}$ augmented with $\mathbf{s}$. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{s})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the complete data log-likelihood function and by $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$ the expected value of the complete data log-likelihood function with respect to the unknown data $\mathbf{s}$ given the observed data $\mathbf{z}$ and the current parameters estimates. That is, we define:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{s})|\mathbf{z}, \boldsymbol{\theta}^{(m-1)}],$$

where $\boldsymbol{\theta}^{(m-1)}$ is the vector of current estimate of the parameters, which are used to evaluate the expectation and $\boldsymbol{\theta}$ are the new parameters that we optimize to increase Q.

Each iteration of the EM algorithm involves two steps, the expectation step and the maximization step:

E-step: Compute $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$ as a function of $\boldsymbol{\theta}$;

M-step: Find $\boldsymbol{\theta}^{(m)}$ such that $Q(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m-1)}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)})$.

These two steps are repeated until convergence. Each iteration of the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta})$ and the EM algorithm typically converges to a local or global maximum of the likelihood function.

An extension of this method was proposed by Meng (1993) and called the ECM algorithm. The ECM algorithm replaces each M-step of the EM algorithm by a sequence of S conditional maximization steps, which is called the CM-steps, each of which maximizes Q over $\boldsymbol{\theta}$ but with some vector function of $\boldsymbol{\theta}$, $g_s(\boldsymbol{\theta})$ $(s = 1, \cdots, S)$ fixed at its previous value.

Let $\mathbf{z} = (\mathbf{z}_{11}^\top, \ldots, \mathbf{z}_{1n_1}^\top, \ldots, \mathbf{z}_{p1}^\top, \ldots, \mathbf{z}_{pn_p}^\top)^\top$, with $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, $\mathbf{x} = (x_{11}, \ldots, x_{1n_1}, \ldots, x_{p1}, \ldots, x_{pn_p})^\top$ and $\mathbf{t} = (t_{11}, \ldots, t_{1n_1}, \ldots, t_{p1}, \ldots, t_{pn_p})^\top$. In the following we implement the ECM algorithm for the SN-MEM using double augmentation by considering that $(\mathbf{x}, \mathbf{t})$ are missing data. Thus, under the hierarchical representation (8)-(10), with $\upsilon^2 = \phi_x(1 - \delta_x^2)$ and $\tau = \phi_x^{1/2}\delta_x$, it follows that the complete log-likelihood function associated with $(\mathbf{z}, \mathbf{x}, \mathbf{t})$ is

$$\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}, \mathbf{t}) \propto -\sum_{i=1}^{p} \frac{n_i}{2} \log(|D(\boldsymbol{\phi}_i)|) - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i} x_{ij})^\top D^{-1}(\boldsymbol{\phi}_i)(\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i} x_{ij})$$
$$-\frac{N}{2} \log(\upsilon^2) - \frac{1}{2\upsilon^2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (x_{ij} - \mu_x - \tau t_{ij})^2 - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} t_{ij}^2, \quad (19)$$

where $N = \sum_{i=1}^{p} n_i$. Let $\widehat{x}_{ij} = E[x_{ij}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}]$, $\widehat{x^2}_{ij} = E[x_{ij}^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}]$, $\widehat{t}_{ij} = E[T_{ij}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}]$, $\widehat{t^2}_{ij} = E[T_{ij}^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}]$ and $\widehat{tx}_{ij} = E_{t_{ij}, x_{ij}}[T_{ij} x_{ij}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}]$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. Considering double conditional expectations and the moments of the truncated normal distributions (see Johnson et al., 1994, Section 10.1) we obtain that

$$\begin{aligned}
\widehat{t}_{ij} &= \widehat{\mu}_{T_{ij}} + W_{\Phi_1}(\frac{\widehat{\mu}_{T_{ij}}}{\widehat{N}_{Ti}})\widehat{N}_{Ti}, \\
\widehat{t^2}_{ij} &= \widehat{\mu}_{T_{ij}}^2 + \widehat{N}_{Ti}^2 + W_{\Phi_1}(\frac{\widehat{\mu}_{T_{ij}}}{\widehat{N}_{Ti}})\widehat{N}_{Ti}\widehat{\mu}_{T_{ij}}, \\
\widehat{x}_{ij} &= \widehat{c}_{ij} + \widehat{d}_i \, \widehat{t}_{ij}, \\
\widehat{x^2}_{ij} &= \widehat{M_i^2} + \widehat{c}_{ij}^2 + 2\widehat{c}_{ij} \, \widehat{d}_i \, \widehat{t}_{ij} + \widehat{d}_i^{\,2} \, \widehat{t^2}_{ij}, \quad \text{and} \\
\widehat{tx}_{ij} &= \widehat{c}_{ij} \, \widehat{t}_{ij} + \widehat{d}_i \, \widehat{t^2}_{ij},
\end{aligned} \quad (20)$$

where $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $\widehat{N}_{Ti}^2 = [1 + \widehat{\tau}^2 \widehat{\boldsymbol{\beta}}_{0i}^\top (D(\widehat{\boldsymbol{\phi}}_i) + \widehat{\upsilon}^2 \widehat{\boldsymbol{\beta}}_{0i} \widehat{\boldsymbol{\beta}}_{0i}^\top)^{-1} \widehat{\boldsymbol{\beta}}_{0i}]^{-1}$, $\widehat{\mu}_{T_{ij}} = \widehat{\tau} \widehat{N}_{Ti}^2 \widehat{\boldsymbol{\beta}}_{0i}^\top (D(\widehat{\boldsymbol{\phi}}_i) + \widehat{\upsilon}^2 \widehat{\boldsymbol{\beta}}_{0i} \widehat{\boldsymbol{\beta}}_{0i}^\top)^{-1} (\mathbf{z}_{ij} - \widehat{\boldsymbol{\beta}}_{0i} \widehat{\mu}_x)$, $\widehat{M_i^2} = \widehat{\upsilon}^2 [1 + \widehat{\upsilon}^2 \widehat{\boldsymbol{\beta}}_{0i}^\top D^{-1}(\widehat{\boldsymbol{\phi}}_i) \widehat{\boldsymbol{\beta}}_{0i}]^{-1}$, $\widehat{c}_{ij} = \widehat{\mu}_x + \widehat{M_i^2} \widehat{\boldsymbol{\beta}}_{0i}^\top D^{-1}(\widehat{\boldsymbol{\phi}}_i)(\mathbf{z}_{ij} - \widehat{\boldsymbol{\beta}}_{0i} \widehat{\mu}_x)$ and $\widehat{d}_i = \widehat{\tau}(1 - \widehat{M_i^2} \widehat{\boldsymbol{\beta}}_{0i}^\top D^{-1}(\widehat{\boldsymbol{\phi}}_i) \widehat{\boldsymbol{\beta}}_{0i})$.

It follow that the conditional expectation in the E-step has the form

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{z},\mathbf{x},\mathbf{t})|\mathbf{z},\widehat{\boldsymbol{\theta}}] \propto -\sum_{i=1}^{p}\frac{n_i}{2}\log(|D(\boldsymbol{\phi}_i)|)$$
$$-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\mathbf{z}_{ij}-\boldsymbol{\beta}_{0i}\widehat{x}_{ij})^{\top}D^{-1}(\boldsymbol{\phi}_i)(\mathbf{z}_{ij}-\boldsymbol{\beta}_{0i}\widehat{x}_{ij})-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{x^2}_{ij}-(\widehat{x}_{ij})^2)\boldsymbol{\beta}_{0i}^{\top}\mathrm{D}^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$$
$$-\frac{N}{2}\log(\upsilon^2)-\frac{1}{2\upsilon^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{x^2}_{ij}+\mu_x^2+\tau^2\widehat{t^2}_{ij}-2\widehat{x}_{ij}\mu_x-2\tau\widehat{tx}_{ij}+2\tau\mu_x\widehat{t}_{ij}), \tag{21}$$

which leads to the following ECM algorithm:

E-step: Given $\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}$, compute $\widehat{t}_{ij}$, $\widehat{t^2}_{ij}$, $\widehat{x}_{ij}$, $\widehat{x^2}_{ij}$ and $\widehat{tx}_{ij}$ for $i=1,\ldots,p$ and $j=1,\ldots,n_i$ using (20).

CM-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $E[\ell_c(\boldsymbol{\theta}|\mathbf{z},\mathbf{x},\mathbf{t})|\mathbf{z},\widehat{\boldsymbol{\theta}}]$ over $\boldsymbol{\theta}$, which leads to

$$\widehat{\beta}_{k_i} = \frac{\sum_{j=1}^{n_i}y_{k_{ij}}\widehat{x}_{ij}}{\sum_{j=1}^{n_i}\widehat{x^2}_{ij}},\quad k=1,\ldots,m,\ i=1,\ldots,p, \tag{22}$$

$$\widehat{\sigma^2}_{e_i} = \frac{1}{mn_i}\Big(\sum_{k=1}^{m}\sum_{j=1}^{n_i}y_{k_{ij}}^2-\frac{\sum_{k=1}^{m}(\sum_{j=1}^{n_i}y_{k_{ij}}\widehat{x}_{ij})^2}{\sum_{j=1}^{n_i}\widehat{x^2}_{ij}}\Big),\quad i=1,\ldots,p, \tag{23}$$

$$\widehat{\sigma}_u^2 = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}^2-2X_{ij}\widehat{x}_{ij}+\widehat{x^2}_{ij}), \tag{24}$$

$$\widehat{\mu}_x = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{x}_{ij}-\tau\widehat{t}_{ij}), \tag{25}$$

$$\widehat{\upsilon}^2 = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{x^2}_{ij}+\mu_x^2+\tau^2\widehat{t^2}_{ij}-2\mu_x\widehat{x}_{ij}-2\tau\widehat{tx}_{ij}+2\tau\mu_x\widehat{t}_{ij}),\ \text{ and} \tag{26}$$

$$\widehat{\tau} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{tx}_{ij}-\mu_x\widehat{t}_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t^2}_{ij}}. \tag{27}$$

Notice that we have closed form expressions for $\widehat{\beta}_{k_i}$, $\widehat{\sigma^2}_{e_i}$ and $\widehat{\sigma}_u^2$, $k=1,\cdots,m$, $i=1,\cdots,p$. Although the joint maximization of the vector of parameters $\boldsymbol{\theta}$ are not in closed form, we note that the conditional maximum likelihood estimate of μ_x in the mth iteration given $\tau=\tau^{(m-1)}$ would be given by the equation (25). On the other hand, given $\mu_x=\mu_x^{(m)}$, the conditional maximum likelihood estimate of τ in the mth iteration is obtained by using the equation (27). Also the conditional maximum likelihood estimate of υ^2 in the mth iteration, given $\tau=\tau^{(m)}$ and $\mu_x=\mu_x^{(m)}$ would be given by the equation (26). So considering the ECM algorithm, the obtention of the maximum likelihood estimate is straight forward in our case.

Note that (22) can be written, alternatively, as $\widehat{\boldsymbol{\beta}}=\boldsymbol{\Delta}_x^{-1}\boldsymbol{\delta}_{xy}$, where $\boldsymbol{\Delta}_x=\mathrm{diag}(\sum_{j=1}^{n_1}\widehat{x^2}_{1j},\ldots,\sum_{j=1}^{n_p}\widehat{x^2}_{pj},\ldots,\sum_{j=1}^{n_1}\widehat{x^2}_{1j},\ldots,\sum_{j=1}^{n_p}\widehat{x^2}_{pj})$ and $\boldsymbol{\delta}_{xy}=(\sum_{j=1}^{n_1}y_{1_{1j}}\widehat{x}_{1j},\ldots,\sum_{j=1}^{n_p}y_{1_{pj}}\widehat{x}_{pj},\sum_{j=1}^{n_1}y_{2_{1j}}\widehat{x}_{1j},\ldots,\sum_{j=1}^{n_p}y_{2_{pj}}\widehat{x}_{pj},\ldots,\sum_{j=1}^{n_1}y_{m_{1j}}\widehat{x}_{1j},\ldots,\sum_{j=1}^{n_p}y_{m_{pj}}\widehat{x}_{pj})^{\top}$. The shape and scale parameters of the latent variable x, can be estimated by noting that $\tau/\upsilon = \lambda_x$, and $\phi_x = \tau^2 + \upsilon^2$. Starting values are often chosen to be the corresponding estimates under a normal assumption, where the starting value for the asymmetric parameter is set to be 0 and, as recommended in the literature, it is useful to run the ECM-algorithm several times with different starting values. Inspection of information criteria such as Akaike Information Criterion (AIC, $-\ell(\widehat{\boldsymbol{\theta}}) + P$), Schawarz's Bayesian Information Criterion (BIC, $-\ell(\widehat{\boldsymbol{\theta}}) + 0.5\log(mN)P$), and the Hannan-Quinn Criterion (HQ, $-\ell(\widehat{\boldsymbol{\theta}}) + \log(\log(mN))P$), where P is the number of free parameters in the model and $N = \sum_{i=1}^{p} n_i$, can be used in practice to select between N-MEM and SN-MEM fits. In the next section we discuss the EM-algorithm for evaluating the restricted MLE in the SN-MEM with especial emphasis in the slope parameter.

3.1.1 Restricted estimation

The main objective of the analysis in this type of models is to estimate $\boldsymbol{\beta}$ and to test the hypothesis $H_{01} : \beta_{k_1} = \ldots = \beta_{k_p}$ or $H_{02} : \beta_{1_i} = \ldots = \beta_{m_i}$, $i = 1, \ldots, p$, $k = 1, \ldots, m$. For instance, in the dental clinical trial the interest was to know if the experimental mouth rinse A and B are more efficient than the control mouth rinse C after three and after six months from the baseline leading to the following testing hypothesis: $H_{01} : \beta_{1_1} = \beta_{1_2} = \beta_{1_3}$ and $H_{02} : \beta_{2_1} = \beta_{2_2} = \beta_{2_3}$, respectively. Another question of interest was to know if the experimental mouth rinses A and B were long lasting, i.e., $H_{03} : \beta_{1_2} = \beta_{2_2}$ and $H_{04} : \beta_{1_3} = \beta_{2_3}$, respectively. So, it is important to also find an estimation procedure under the restriction imposed by H_0 and use these results to construct likelihood ratio and score test statistics for testing the above hypothesis. As such, suppose that our interest centers in estimating the parameters $\boldsymbol{\beta}$ under q linearly independent restrictions defined as $\mathbf{C}_s^\top \boldsymbol{\beta} - d_s = 0$, where $\mathbf{C}_s$, $s = 1, \ldots, q$, are $mp \times 1$ vectors and d_s, $s = 1, \ldots, q$, are scalars, both of which are known and fixed. The problem here is to maximize the complete log-likelihood function $E[\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}, \mathbf{t})|\mathbf{z}, \widehat{\boldsymbol{\theta}}]$ subject to the linear constraints $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$, where $\mathbf{C} = (\mathbf{C}_1, \ldots, \mathbf{C}_q)^\top$ and $\mathbf{d} = (d_1, \ldots, d_q)^\top$. Similarly as Nyquist (1991), we will apply the methodology of penalty functions by considering the quadratic penalty function

$$P(\boldsymbol{\theta}, \boldsymbol{\Upsilon}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}, \mathbf{t})|\mathbf{z}, \widehat{\boldsymbol{\theta}}] - \frac{1}{2}\sum_{s=1}^{q} \gamma_s (d_s - \mathbf{C}_s^\top \boldsymbol{\beta})^2, \tag{28}$$

where, $\boldsymbol{\Upsilon} = (\gamma_1, ...\gamma_q)^\top$. The procedure consists in finding the solution of $max_{\boldsymbol{\theta}} P(\boldsymbol{\theta}, \boldsymbol{\Upsilon})$ for positive and fixed values of γ_s, $s = 1, \ldots, q$. The solution for $\boldsymbol{\beta}$ will be denoted by $\boldsymbol{\beta}(\boldsymbol{\Upsilon})$. The restricted equality estimate $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_c = \lim_{\gamma_1, \ldots, \gamma_q \to \infty} \boldsymbol{\beta}(\boldsymbol{\Upsilon}).$$

Using a similar approach of that given in Nyquist (1991), one may show that $\boldsymbol{\beta}(\boldsymbol{\Upsilon})$ is the solution of the following iterative process:

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_c^{(r+1)} &= \left(\boldsymbol{\Delta}_x^{(r)}\right)^{-1} \boldsymbol{\delta}_{xy}^{(r)} + \left(\boldsymbol{\Delta}_x^{(r)}\right)^{-1} \mathbf{C}^\top [\mathbf{C} \left(\boldsymbol{\Delta}_x^{(r)}\right)^{-1} \mathbf{C}^\top]^{-1} [\mathbf{d} - \mathbf{C} \left(\boldsymbol{\Delta}_x^{(r)}\right)^{-1} \boldsymbol{\delta}_{xy}^{(r)}] \\
&= \widehat{\boldsymbol{\beta}}^{(r)} + \left(\boldsymbol{\Delta}_x^{(r)}\right)^{-1} \mathbf{C}^\top [\mathbf{C} \boldsymbol{\Delta}_x^{(r)^{-1}} \mathbf{C}^\top]^{-1} [\mathbf{d} - \mathbf{C}\widehat{\boldsymbol{\beta}}^{(r)}]
\end{aligned} \tag{29}$$

for $r = 0, 1, \ldots$, where $\widehat{\boldsymbol{\beta}}^{(r)}$, $\boldsymbol{\delta}_{xy}^{(r)}$ and $\boldsymbol{\Delta}_x^{(r)}$ are obtained using the CM-step given in the previous Section. The ECM algorithm for estimating the parameters of the model (6)-(7) under the restriction $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, denoted by $\widetilde{\boldsymbol{\theta}}_c$, follows the same procedures given in (20)-(27), replacing $\widehat{\boldsymbol{\beta}}$ by $\widetilde{\boldsymbol{\beta}}_c$ in the CM-step of the algorithm. Note from (29) that the problem of testing linear inequality hypotheses of the form $H_0 : \mathbf{C}\boldsymbol{\beta} - \mathbf{d} \geq \mathbf{0}$ can easily be treated using conditions given in Fahrmeir and Klinger (1994) which guarantee that $\widetilde{\boldsymbol{\beta}}_c$ corresponds to the inequality restricted estimate. See Cysneiros and Paula (2004) for further details on this approach

3.2 Hypothesis testing

Let $\mathbf{J}_0 = \left[\sum_{i=1}^{p} \frac{n_i}{N} E\left(\mathbf{J}_{ij}(\boldsymbol{\theta})\right) |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$, where $N = \sum_{i=1}^{p} n_i$, $\boldsymbol{\theta}_0$ is the true parameter vector and $\mathbf{J}_{ij}(\boldsymbol{\theta})$ as presented in Section 2. Then under regularity conditions (Bradley and Gart (1962)), $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ has asymptotically as $N \to \infty$ the multivariate normal distribution with zero means and variance-covariance matrix $\mathbf{J}_0^{-1}$, with $\hat{\boldsymbol{\theta}}$ denoting the MLE of $\boldsymbol{\theta}$. Considering this distribution we may obtain the asymptotic confidence intervals and hypothesis testing for the parameter $\boldsymbol{\theta}$. In the context of dental clinical trial the hypothesis of interest concerns $\boldsymbol{\beta}$.

Next, we conduct inference regarding the regression coefficients $\boldsymbol{\beta}$. As discussed in the previous section, to test H_{01}, H_{02}, H_{03} and H_{04} we may consider the likelihood ratio (ξ_{LR}), score (ξ_{SR}) or Wald (ξ_W) test statistics. These tests, which are sometimes called the classical tests are particulary useful when the parameter space is multidimensional and they are asymptotically equivalent under the null hypothesis. Let $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ be the ML estimates of $\boldsymbol{\theta} \in \mathbb{R}^{mp+p+4}$ under the unrestricted model and under the null hypothesis, respectively. We notice that the hypothesis of interest can be written as $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, where $\mathbf{C}$ is a $q \times mp$ dimensional matrix with $rank(\mathbf{C}) = q \leq mp$ and $\mathbf{d}$ is a $q \times 1$, known vector. Thus, the statistics ξ_{LR}, ξ_{SR} and ξ_W can be written as

$$\xi_{LR} = 2[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})],\ \xi_{SR} = [\mathbf{U}_\beta(\widetilde{\boldsymbol{\theta}})]^\top[\mathbf{J}_{\beta\beta}^{-1}(\widetilde{\boldsymbol{\theta}})][\mathbf{U}_\beta(\widetilde{\boldsymbol{\theta}})]$$

and

$$\xi_W = [\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}]^\top[\mathbf{C}\mathbf{J}_{\beta\beta}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{C}^\top]^{-1}[\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}],$$

where $\ell(\boldsymbol{\theta})$ is the log-likelihood function, $\mathbf{U}_\beta$ and $\mathbf{J}_{\beta\beta}^{-1}$ corresponds to the partition of $U(\boldsymbol{\theta})$ defined in (15) and $\mathbf{J}$ defined in (18) as $U(\boldsymbol{\theta}) = (U_{\boldsymbol{\beta}}^\top, U_{\boldsymbol{\theta}-\boldsymbol{\beta}}^\top)^\top$ and $\mathbf{J}^{-1} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} & \mathbf{J}_{\boldsymbol{\beta},\boldsymbol{\theta}-\boldsymbol{\beta}}^{-1} \\ \mathbf{J}_{\boldsymbol{\theta}-\boldsymbol{\beta},\boldsymbol{\beta}}^{-1} & \mathbf{J}_{\boldsymbol{\theta}-\boldsymbol{\beta},\boldsymbol{\theta}-\boldsymbol{\beta}}^{-1} \end{bmatrix}$, with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \cdots, \boldsymbol{\beta}_p^\top)^\top$ and $\boldsymbol{\theta} - \boldsymbol{\beta} = (\sigma_{e_1}^2, \cdots, \sigma_{e_p}^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x)^\top$. Under the null hypothesis, the three statistics follow asymptotically a chi-square distribution with q degrees of freedom (χ_q^2).

4 Local influence

Outliers and detection of influential observations is an important step in the analysis of a data set. There are several alternatives to evaluate the influence of perturbations in the data and/or in the model on the parameter estimate, see for example, Cook and Weisberg (1982), Chatterjee and Hadi (1988).

Case deletion is a common way to assess the effect of an observation on the estimation process. This is a global influence analysis, since the effect of the observation is evaluated by eliminating it from the data set. Alternatively, local influence is based more on geometric differentiation rather than on the elimination of the observations. A differential comparison of estimators is used before and after perturbing the data and/or model assumptions. In order to evaluate the robustness of the maximum likelihood estimator, to possible atypical observations in the data set, we use the local influence concept introduced by Cook (1986).

Let $l(\boldsymbol{\theta})$ and $l(\boldsymbol{\theta}|\boldsymbol{\omega})$ denote, respectively, the log-likelihood function from the postulated model and the log-likelihood function of the perturbed model, with $\boldsymbol{\theta} \in \mathbb{R}^p$ and $\boldsymbol{\omega}$ a $q \times 1$ vector of perturbations restricted to some open subset of $\mathbb{R}^q$. Denoting the vector of no perturbation by $\boldsymbol{\omega}_0$, we assume that $l(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = l(\boldsymbol{\theta})$. To asses the influence of the perturbations on the maximum likelihood estimate of $\boldsymbol{\theta}$, one may consider the likelihood displacement defined as

$$LD(\boldsymbol{\omega}) = 2[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_\omega)],$$

where $\widehat{\boldsymbol{\theta}}_\omega$ ($\widehat{\boldsymbol{\theta}}$) denotes the maximum likelihood estimator under the model $l(\boldsymbol{\theta}|\boldsymbol{\omega})$ ($l(\boldsymbol{\theta})$). The idea of local influence (Cook, 1986) is concerned with characterizing the behavior of $LD(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$. The procedure consists in selecting a unit direction $\boldsymbol{l}$, $||\boldsymbol{l}|| = 1$, and then to consider the plot of $LD(\boldsymbol{\omega}_0 + a\boldsymbol{l})$ against $a \in \mathbb{R}$. This plot is called *lifted line*. Notice that since $LD(\boldsymbol{\omega}_0) = 0$, $LD(\boldsymbol{\omega}_0 + a\boldsymbol{l})$ has a local minimum at $a = 0$. Each lifted line can be characterized by considering the normal curvature $C_l(\boldsymbol{\theta})$ around $a = 0$. The suggestion is to consider the direction $\boldsymbol{l}_{\max}$ corresponding to the largest curvature $C_{l\max}(\boldsymbol{\theta})$. The index plot of $\boldsymbol{l}_{\max}$ may reveal those observations that under small perturbations exert notable influence on $LD(\boldsymbol{\omega})$. Cook (1986) showed that the normal curvature at the direction $\boldsymbol{l}$ takes the form

$$C_l(\boldsymbol{\theta}) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|, \tag{30}$$

where $-\mathbf{L}$ is the observed information matrix for the postulated model ($\boldsymbol{\omega} = \boldsymbol{\omega}_0$) and $\boldsymbol{\Delta}$ is the $p \times q$ matrix with elements

$$\Delta_{rs} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_r \partial\omega_s},$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0, r = 1, ..., p$ and $s = 1, ..., q$. Therefore, the maximization of (30) is equivalent to finding the largest absolute eigenvalue, $C_{l\max}$, of the matrix $\mathbf{B} = -\boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta}$ and $\boldsymbol{l}_{\max}$ is the corresponding eigenvector. In some situations, it may be of interest to assess the influence on a subset $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. For example, one may have interest on $\boldsymbol{\theta}_1 = (\mu_x, \boldsymbol{\beta}^\top)^\top$ or $\boldsymbol{\theta}_1 = \lambda_x$. In such situations, the curvature at the direction $\boldsymbol{l}$ is given by

$$C_l(\boldsymbol{\theta}_1) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top (\mathbf{L}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta} \boldsymbol{l}|, \tag{31}$$

where

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{22}^{-1} \end{pmatrix},$$

and $\mathbf{L}_{22}$ is obtained from the partition of $\mathbf{L}$ according to the partition of $\boldsymbol{\theta}$. The eigenvector $\boldsymbol{l}_{\max}$ corresponds to the largest absolute eigenvalue of the matrix $\mathbf{B} = \boldsymbol{\Delta}^\top (\mathbf{L}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}$. Another important direction, according to Escobar and Meeker (1992) (see also Verbeke and Molenberghs, 2000) is $\boldsymbol{l} = \boldsymbol{e}_k$, a $N \times 1$ vector of zeros with a one in the k-th position.

In that case, the normal curvature called the total local influence of subject k, is given by $C_k = 2|\boldsymbol{e}_k^\top \mathbf{B}\boldsymbol{e}_k| = 2|b_{kk}|$, where b_{kk} is the kth diagonal element of $\mathbf{B}$, $k = 1, ..., N$. In the case of the SN-MEM we have that $N = \sum_{i=1}^{p} n_i$.

In order to compare local and global influence, we may use the Cook's distance (D_{ij}) and the likelihood displacement (LD_{ij}), which are defined, respectively, as

$$D_{ij} = (\widehat{\boldsymbol{\theta}}_{(ij)} - \widehat{\boldsymbol{\theta}})^\top (-\,\boldsymbol{L})(\widehat{\boldsymbol{\theta}}_{(ij)} - \widehat{\boldsymbol{\theta}})/(mp + p + 4), \tag{32}$$

$$LD_{ij} = 2\left[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{(ij)})\right], \tag{33}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$, where $\widehat{\boldsymbol{\theta}}_{(ij)}$ denotes the ML estimates without the case ij. See Zhao and Lee (1998) for details.

4.1 Curvature derivation for SN-MEM

In this Section we derive the $\boldsymbol{\Delta}$ matrix for different perturbation schemes.

4.1.1 Case weight perturbation

Notice that the logarithm of the likelihood function for the model (6)-(7) is given by (13), where $\ell_{ij}(\boldsymbol{\theta})$ is the contribution of the ijth observation (equally weighted) to the likelihood, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. A perturbed log-likelihood function - allowing different weights for different observations - can be defined by

$$\ell(\boldsymbol{\theta}/\boldsymbol{\omega}) = \sum_{i=1}^{p}\sum_{j=1}^{n_i} \omega_{ij}\ell_{ij}(\boldsymbol{\theta}), \tag{34}$$

where, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x)^\top$ and $\boldsymbol{\omega} = (\omega_{11}, \ldots, \omega_{1n_1}, \ldots, \omega_{p1} \ldots \omega_{pn_p})^\top$. $\boldsymbol{\omega}$ is the vector of weights corresponding to the contribution of each observation to the likelihood, $\boldsymbol{\omega}_0 = \mathbf{1}_N = (1, \ldots, 1)^\top$, with $N = \sum_{i=1}^{p} n_i$ (no perturbation vector). This perturbation scheme is intended to evaluate whether the contribution of the observations with differing weights affect the maximum likelihood estimate of $\boldsymbol{\theta}$. Perhaps, this is the method most commonly used to evaluate the influence of a small modification of the model. Thus, using (34) it follows, after some algebraic manipulation, that the delta matrix is given by

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{11}(\boldsymbol{\theta}), \ldots, \boldsymbol{\Delta}_{1n_1}(\boldsymbol{\theta}), \ldots, \boldsymbol{\Delta}_{p1}(\boldsymbol{\theta}), \ldots, \boldsymbol{\Delta}_{pn_p}(\boldsymbol{\theta})), \tag{35}$$

where, $\boldsymbol{\Delta}_{ij} = \dfrac{\partial \ell_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, with individual elements given by

$$\frac{\partial l_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2}\frac{\partial \log|\boldsymbol{\Sigma_i}|}{\partial \boldsymbol{\gamma}} - \frac{1}{2}\frac{\partial g_{ij}}{\partial \boldsymbol{\gamma}} + \frac{\partial \log(K_{ij})}{\partial \boldsymbol{\gamma}}, \quad \boldsymbol{\gamma} = \boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x. \tag{36}$$

The components of $\frac{\partial l_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}$, i.e., $\dfrac{\partial \log|\boldsymbol{\Sigma_i}|}{\partial \boldsymbol{\gamma}}$, $\dfrac{\partial g_{ij}}{\partial \boldsymbol{\gamma}}$ and $\dfrac{\partial \log(K_{ij})}{\partial \boldsymbol{\gamma}}$ are presented in the Appendix A. The above $\boldsymbol{\Delta}$ matrix is to be evaluated at $\hat{\boldsymbol{\theta}}$.

4.1.2 Response variables perturbation

In this subsection, our interest is to detect the sensitivity of the model when $y_{k_{ij}}$ is perturbed. The perturbation considered here, is given by

$$y_{k_{ij}}(\omega_{ij}) = y_{k_{ij}} + S_k\omega_{ij}, \tag{37}$$

where S_k is a sequence of scale factors $S_1, \ldots, S_m$, which can be taken, for example, as the sample standard deviation of the observations indexed by k and $\boldsymbol{\omega} = (\omega_{11}, \ldots, \omega_{1n_1}, \ldots, \omega_{p1}, \ldots, \omega_{pn_p})^\top$. The no perturbation case follows by taking $\boldsymbol{\omega}_o = \mathbf{0}$ and the perturbed log-likelihood function can be obtained from (13) with $y_{k_{ij}}$ replaced by $y_{k_{ij}}(\omega_{ij})$, $i = 1, \ldots, p, j = 1, \ldots, n_i$. Then

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{p}\sum_{j=1}^{n_i} \ell_{ij}(\boldsymbol{\theta}|\omega_{ij}), \tag{38}$$

where $\ell_{ij}(\boldsymbol{\theta}|\omega_{ij}) \propto -\dfrac{1}{2}\log|\boldsymbol{\Sigma}_i| - \dfrac{1}{2}g_{ij}(\omega_{ij}) + \log(K_{ij}(\omega_{ij}))$ with $g_{ij}(\omega_{ij}) = (\mathbf{z}_{ij}(\omega_{ij}) - \boldsymbol{\mu}_i)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_{ij}(\omega_{ij}) - \boldsymbol{\mu}_i)$, and $K_{ij}(\omega_{ij}) = \Phi_1\left(A_i a_{ij}(\omega_{ij})\right)$ with $a_{ij}(\omega_{ij}) = (\mathbf{z}_{ij}(\omega_{ij}) - \boldsymbol{\mu}_i)^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$.

Differentiating $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ leads to $\boldsymbol{\Delta}$ as defined in (35), where

$$\begin{aligned}\boldsymbol{\Delta}_{ij}(\boldsymbol{\theta}) \;&=\; -\frac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}} + W_{\Phi_1}\left(A_i a_{ij}(\omega_{ij})\right)\left[\frac{\partial A_i}{\partial \boldsymbol{\theta}}Q_{ij}(\omega_{ij}) + A_i\frac{\partial Q_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}\right] \\ &\quad + A_i W'_{\Phi_1}(A_i a_{ij}(\omega_j))Q_{ij}(\omega_{ij})[A_i\frac{\partial a_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}} + a_{ij}(\omega_{ij})\frac{\partial A_i}{\partial \boldsymbol{\theta}}],\end{aligned} \tag{39}$$

with $P_{ij}(\omega_{ij}) = (\mathbf{z}_{ij}(\omega_{ij}) - \boldsymbol{\mu}_i)^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{d}$, $Q_{ij}(\omega_{ij}) = \boldsymbol{d}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$, $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, $\boldsymbol{d} = \dfrac{\partial \mathbf{z}_{ij}(\omega_{ij})}{\partial \omega_{ij}} = (d_1, \boldsymbol{d}_2^\top)^\top$ a $(m+1)\times 1$ vector and $\dfrac{\partial a_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ is as in the unperturbed case, replacing $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$ by $\mathbf{z}_{ij}(\omega_{ij}) = (X_{ij}, y_{1_{ij}} + S_1\omega_{ij}, \ldots, y_{m_{ij}} + S_m\omega_{ij})^\top$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$.

The expressions for $\dfrac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ evaluated at $\boldsymbol{\omega}_o = \mathbf{0}$ is given by

$$\begin{aligned}\frac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\beta}_i} &= -\frac{1}{\sigma^2_{ei}}\boldsymbol{d}_2 + 2\frac{\Lambda_i^2}{\sigma^2_{ei}}a_{ij}\boldsymbol{\beta}_i\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d} - \frac{\Lambda_i}{\sigma^2_{ei}}[(\boldsymbol{W}_{2ij} - \mu_x\boldsymbol{\beta}_i)\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d} + a_{ij}\boldsymbol{d}_2],\\ \frac{\partial P_{ij}(\omega_{ij})}{\partial \sigma^2_{ei}} &= -\frac{1}{\sigma^4_{ei}}\boldsymbol{W}_{2ij}^\top\boldsymbol{d}_2 - \frac{\Lambda_i^2}{\sigma^4_{ei}}a_{ij}\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d} + \frac{\Lambda_i}{\sigma^4_{ei}}[\boldsymbol{W}_{2ij}^\top\boldsymbol{\beta}_i\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d} + a_{ij}\boldsymbol{\beta}_i^\top\boldsymbol{d}_2],\\ \frac{\partial P_{ij}(\omega_{ij})}{\partial \sigma^2_u} &= -\frac{d_1}{\sigma^4_u}W_{1ij} - \frac{\Lambda_i^2}{\sigma^4_u}a_{ij}\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d} + \frac{\Lambda_i}{\sigma^4_u}[W_{1ij}\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d} + a_{ij}d_1],\\ \frac{\partial P_{ij}(\omega_{ij})}{\partial \mu_x} &= -\boldsymbol{\beta}_{0i}^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{d}, \quad \frac{\partial P_{ij}(\omega_{ij})}{\partial \phi_x} = -c_i^{-2}a_{ij}\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{d}, \quad \frac{\partial P_{ij}(\omega_{ij})}{\partial \lambda_x} = 0,\end{aligned}$$

where $W_{1ij} = X_{ij} - \mu_x$, $\boldsymbol{W}_{2ij} = \boldsymbol{Y}_{ij}(\omega_{ij}) - \boldsymbol{\beta}_i\mu_x$, $d_1 = 0$ and $\boldsymbol{d}_2 = \boldsymbol{S}$, with $\boldsymbol{Y}_{ij}(\omega_{ij}) = (y_{1_{ij}} + S_1\omega_{ij}, \ldots, y_{m_{ij}} + S_m\omega_{ij})^\top$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$ and $\boldsymbol{S} = (S_1, ..., S_m)^\top$. Then $Q_{ij}(\omega_{ij}) = \frac{1}{\sigma^2_{ei}}\boldsymbol{S}^\top\boldsymbol{\beta}_i$. The vector $\frac{\partial a_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ is as given in the unperturbed case and can be found in Appendix A. The $\frac{\partial A_i}{\partial \boldsymbol{\theta}}$ can also be found in the Appendix A.

4.1.3 Explanatory variables perturbation

If we are interested in investigating the sensitivity of minor perturbation in the explanatory variable, we can define for example, the following perturbation scheme for the explanatory variable, in the same way that was defined for the response variable

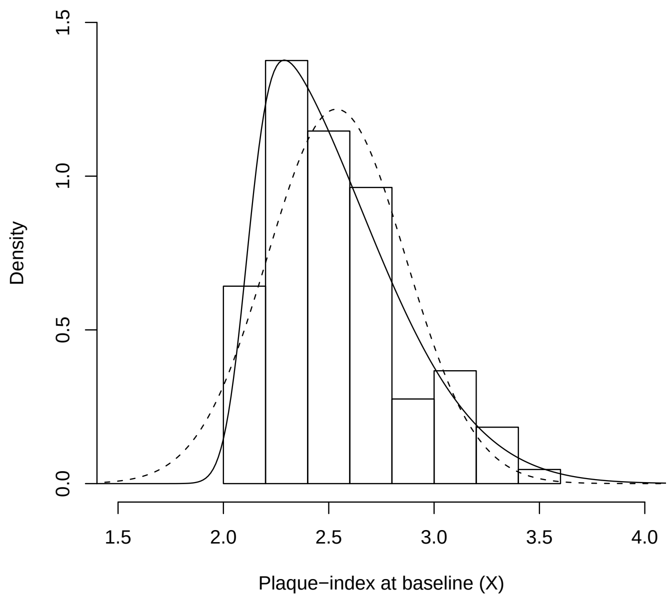
$$X_{ij}(\omega_{ij}) = X_{ij} + \omega_{ij}. \tag{40}$$

The perturbed log-likelihood follows from (38) with X_{ij} replaced by $X_{ij}(\omega_{ij})$ and $y_{k_{ij}}(\omega_{ij})$ replaced by $y_{k_{ij}}$,$i = 1, \dots, p$, $j = 1, \dots, n_i$. As in the response variables perturbation scheme, $\boldsymbol{\omega} = (\omega_{11}, \dots, \omega_{1n_1}, \dots, \omega_{p1} \dots \omega_{pn_p})^\top$, $\boldsymbol{\omega}_o = \mathbf{0}$ and the $\boldsymbol{\Delta}$ matrix is as given in (35), with $\boldsymbol{\Delta}_{ij}$ as given in (39) replacing $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}} + S_1\omega_{ij}, \dots, y_{m_{ij}} + S_m\omega_{ij})^\top$ by $\mathbf{z}_{ij}(\omega_{ij}) = (X_{ij} + \omega_{ij}, y_{1_{ij}}, \dots, y_{m_{ij}})^\top$. In this case $d_1 = 1$ and $\boldsymbol{d}_2 = \mathbf{0}_{\boldsymbol{m}}$, which leads to $Q_{ij}(\omega_{ij}) = \frac{1}{\sigma_u^2}$. The expressions for $\dfrac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ evaluated at $\boldsymbol{\omega}_o = \mathbf{0}$ are the same as given in the response variable perturbation shceme, noting that $W_{1ij} = X_{ij} + \omega_{ij} - \mu_x$, $\boldsymbol{W}_{2ij} = \boldsymbol{Y}_{ij} - \boldsymbol{\beta}_i\mu_x$ and $\boldsymbol{d} = (1, \mathbf{0}_{\boldsymbol{m}}{}^\top)^\top$.

5 Application

In this section, we apply the methodology discussed in this work to a real data set analyzed in Hadgu and Koch (1999) using generalized estimating equations. The data set and the objective of the study was described in the Introduction and Section 3.1, respectively.

Figure 1: Dental plaque index data set. Histogram of the observed covariate X (plaque-index at baseline) superimposed by the estimated densities using skew-normal (solid) and normal distribution (dashed).



The ML estimates of the parameters of the model were obtained for the SN-MEM and N-MEM and are presented in Table 1. As can be seen, the estimate of the parameters for the two models are close, except for the estimates of μ_x and σ_x^2. Moreover, clearly, the values of β_{ij} which is less than 1 indicates dental plaque reduction. Note that the estimated standard deviation for λ_x seems to be large, but AIC, BIC and HQ values shown in the bottom of the Table 1 seems to favor SN-MEM over N-MEM, supporting the contention of the departure from normality. This conclusion is also supported by the results from the likelihood ratio test for $H_0 : \lambda_x = 0$ ($\xi_{LR} = 17.8642$, p-value $\simeq 0$) and also graphically by Figure 1. Nevertheless, a nominally 95% symmetric confidence interval for λ_x, calculated using the (very large) estimated standard deviation of 6.0782 and large-sample normal approximation, was found to be $(-5.8, 18.0)$, in clear disagreement with the previously quoted results. However, as noted in simulation studies conducted by the authors, it appears to indicate that Wald type statistics based on the asymptotic covariance matrix, estimated using the observed information matrix, is typically less powerful at detecting skewness than the likelihood ratio statistic.

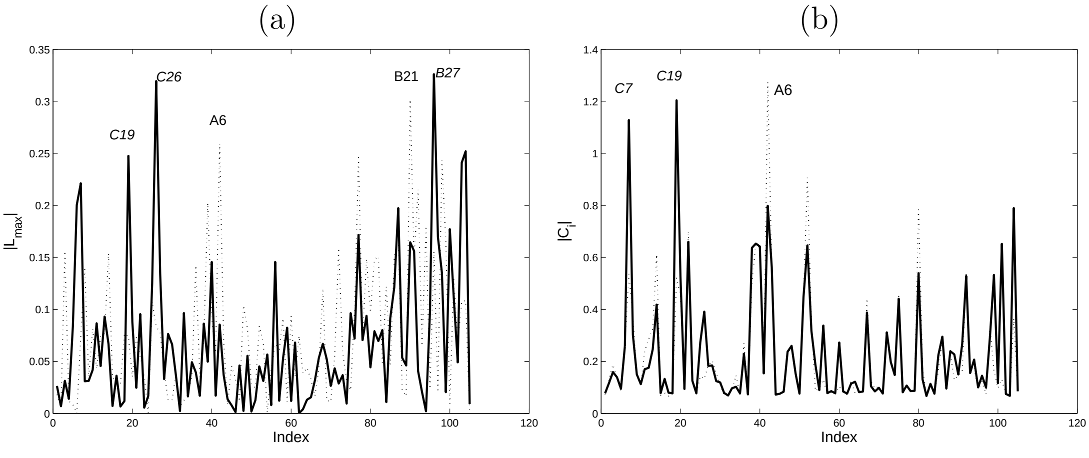
Table 1: Results of fitting SN-MEM and N-MEM to the dental plaque index data set. SE represents the estimated asymptotic standard errors based on the information matrix given in Appendix B.

	SN-MEM		N-MEM	
Parameter	Estimate	SE	Estimate	SE
β_{1_1}	0.7020	0.0339	0.7021	0.0340
β_{1_2}	0.5239	0.0441	0.5241	0.0442
β_{1_3}	0.5088	0.0317	0.5087	0.0317
β_{2_1}	0.6857	0.0339	0.6859	0.0340
β_{2_2}	0.5016	0.0441	0.5017	0.0441
β_{2_3}	0.4139	0.0317	0.4139	0.0317
$\sigma^2_{e_1}$	0.2746	0.0460	0.2739	0.0461
$\sigma^2_{e_2}$	0.4306	0.0752	0.4308	0.0751
$\sigma^2_{e_3}$	0.2257	0.0377	0.2253	0.0380
σ^2_u	0.0010	0.0154	0.0021	0.0210
μ_x	2.1082	0.0425	2.5343	0.0325
σ^2_x	0.2907	0.0550	0.1086	0.0210
λ_x	6.1291	6.0782	-	-
log-likelihood	-194.4457		-203.3778	
AIC	1.9757		2.0512	
BIC	2.1400		2.2029	
HQ	2.0422		2.1127	

Considering the hypothesis $H_{01} : \beta_{1_1} = \beta_{1_2} = \beta_{1_3}$, which corresponds to a comparison of the effects after three months, of the experimental mouth rinses A and B with that of the control mouth rinse C, the restricted ML estimates are given by $\widetilde{\beta}_{1_1} = \widetilde{\beta}_{1_2} = \widetilde{\beta}_{1_3} = 0.5811, \widetilde{\beta}_{2_1} = 0.6857, \widetilde{\beta}_{2_2} = 0.5016, \widetilde{\beta}_{2_3} = 0.4139, \widetilde{\sigma}^2_{\epsilon 1} = 0.2746, \widetilde{\sigma}^2_{\epsilon 2} = 0.4305, \widetilde{\sigma}^2_{\epsilon 3} = 0.2257, \widetilde{\sigma}^2_u = 0.0011, \widetilde{\mu}_x = 2.1082, \widetilde{\sigma}^2_x = 0.2907$ and $\widetilde{\lambda}_x = 6.1285$. The H_{01} hypothesis is emphatically rejected since $\xi_{LR} = 19.59$, $\xi_W = 19.5635$ and $\xi_{SR} = 25.0820$ corresponding to a p-values around zero. Considering the hypothesis $H_{02} : \beta_{2_1} = \beta_{2_2} = \beta_{2_3}$, which corresponds to a comparison of the effects after six months, of the experimental mouth rinses A and B with that of the control mouth rinse C, the restricted ML estimates are given by $\widetilde{\beta}_{2_1} = \widetilde{\beta}_{2_2} = \widetilde{\beta}_{2_3} = 0.5372, \widetilde{\beta}_{1_1} = 0.7020, \widetilde{\beta}_{1_2} = 0.5239, \widetilde{\beta}_{1_3} = 0.5088, \widetilde{\sigma}^2_{\epsilon 1} = 0.2746, \widetilde{\sigma}^2_{\epsilon 2} = 0.4305, \widetilde{\sigma}^2_{\epsilon 3} = 0.2257, \widetilde{\sigma}^2_u = 0.2257, \widetilde{\mu}_x = 2.1082, \widetilde{\sigma}^2_x = 0.2907$ and $\widetilde{\lambda}_x = 6.1285$. The H_{02} hypothesis is also rejected since $\xi_{LR} = 34.9505$, $\xi_W = 34.8530$ and $\xi_{SR} = 51.7213$ which corresponds to p-values around zero. If we consider the hypothesis $H_{03} : \beta_{1_2} = \beta_{2_2}$ and $H_{04} : \beta_{1_3} = \beta_{2_3}$, we obtain $\xi_{LR} = 0.1288$, $\xi_W = 0.1287$, $\xi_{SR} = 0.1294$ and $\xi_{LR} = 4.4733$, $\xi_W = 4.4730$, $\xi_{SR} = 5.0487$ respectively. Thus, we fail to reject H_{03} and reject H_{04}, which means that the experimental mouth rinse B is long lasting. If we now consider the hypothesis $H_{05} : \beta_{1_1} = \beta_{2_1}, \beta_{1_2} = \beta_{2_2}$, which corresponds to analyzing whether the control mouth rinse C and the mouth rinse A reduce dental plaque at the same rates over the entire clinical trial, we fail to reject it since $\xi_{LR} = 0.2440$, $\xi_W = 0.2439$ and $\xi_{SR} = 0.2440$, which corresponds to p-values greater than 0.1. The general conclusion is that the mouth rinse B is more effective for dental plaque reduction.

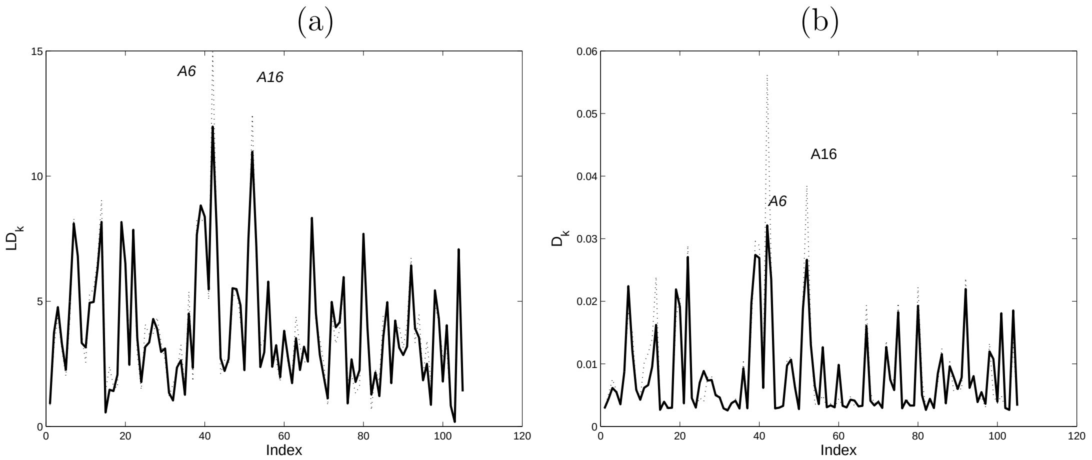
Next, we apply the diagnostic methods specified in Section 4 to the Hadgu and Koch data set. The index plots of $\mathbf{l}_{max}$ to assess the influence of the perturbation on the ML estimate of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p, \sigma^2_{e_1}, \ldots, \sigma^2_{e_p}, \sigma^2_u, \mu_x, \phi_x, \lambda_x)^T$ are presented in Figure 2 through Figure 5. Considering these graphs, the first 36 observations correspond to the observations obtained by the volunteers who used the control mouth rinse C, the observations 37 through 69 correspond to those obtained using the experimental mouth rinse A, while the last 36 observations correspond to those obtained using the experimental mouth rinse B.

Figure 2: Dental plaque index data set. Index plot of (a) $|lmax|$ and (b) C_k for the case weight perturbation scheme. (—) and (...) denotes the index plot for the SN-MEM and N-MEM, respectively.



In Figure 2 we present the index plot of $|l_{max}|$ and $|C_k|$ under the case weights perturbation. Based on this perturbation scheme we find that subjects 7, 19 and 26 of the control month rinse C and subjects 27 and 104 of B are the most influential on $\widehat{\boldsymbol{\theta}}$ under the SN-MEM. These volunteers are the ones with the smallest dental plaque index in the beginning of the study and also they presented a reasonable reduction of the dental plaque index after the use of the control mouth rinse C and experimental mouth rinse B, respectively. Notice that, if we focus on the first 36 observations which corresponds to those obtained by the use of the control mouth rinse C, these observations (7, 19 and 26) stands out more than the observations 27 and 104 stands out considering the last 36 observations, which corresponds to the values obtained by the use of the experimental mouth rinse B. This result is in accordance with the fact that the behavior of these observations considering the ones obtained using the control mouth rinse C are more atypical. Under the N-MEM the observations that corresponds to the subjects 6 of A and 21 of the month rinse B are the most influential. The observation 6 of A is the observation with the highest value of the dental plaque index in the beginning of the study. So the observations which are influential considering these two models (SN-MEM and N-MEM) are not the same.

Figure 3: Dental plaque index data set. Index plot of (a) Likelihood displacement LD_k and (b) Cook's distance D_k. (—) and (...) denotes the index plot for the SN-MEM and N-MEM, respectively.



In order to compare with the result of local influence, in Figure 3 we present some results of global influence, such as, likelihood distance (LD_k) and Cook's distance (D_k), $k = 1, \ldots, N$. Note that, (LD_k) and (D_k) reveals subjects 6 and 16 of A as the most globally influential under the N-MEM and SN-MEM. These two observations are the ones with the greatest value of the dental plaque index in the beginning of the study. Also, these observations stands out more in the N-MEM then in SN-MEM.

Under the perturbation of the response and explanatory variables we find that the $C_{lmax}(\widehat{\boldsymbol{\theta}})$ = 6.2178 and $C_{lmax}(\widehat{\boldsymbol{\theta}}) = 110.4162$, respectively. The index plot of $|\ell_{max}|$ and $|C_k|$ under the perturbation of the response variable is given in Figure 4. Notice that the observations corresponding to the control mouth rinse C stands out. When the explanatory variable perturbation is considered (Figure 5), subjects 20 of the month rinse A and 30 and 32 of the month rinse B are the most influential, which are also very different of the one under the N-MEM as expected, due to the asymmetric distribution that we have considered.

Figure 4: Dental plaque index data set. Perturbation of the responses variables. Index plot of (a) $|lmax|$ and (b) $|C_k|$. (—) and (...) denotes the index plot for the SN-MEM and N-MEM, respectively.

(a) (b)

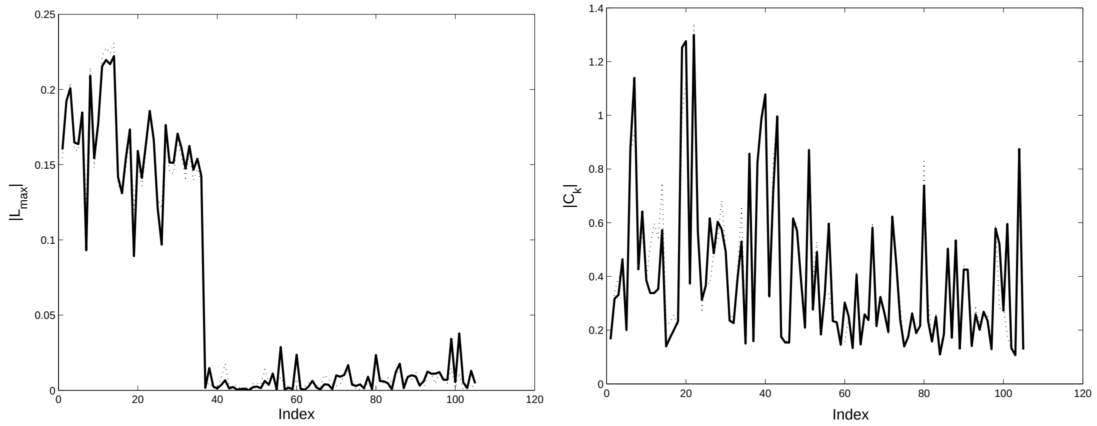
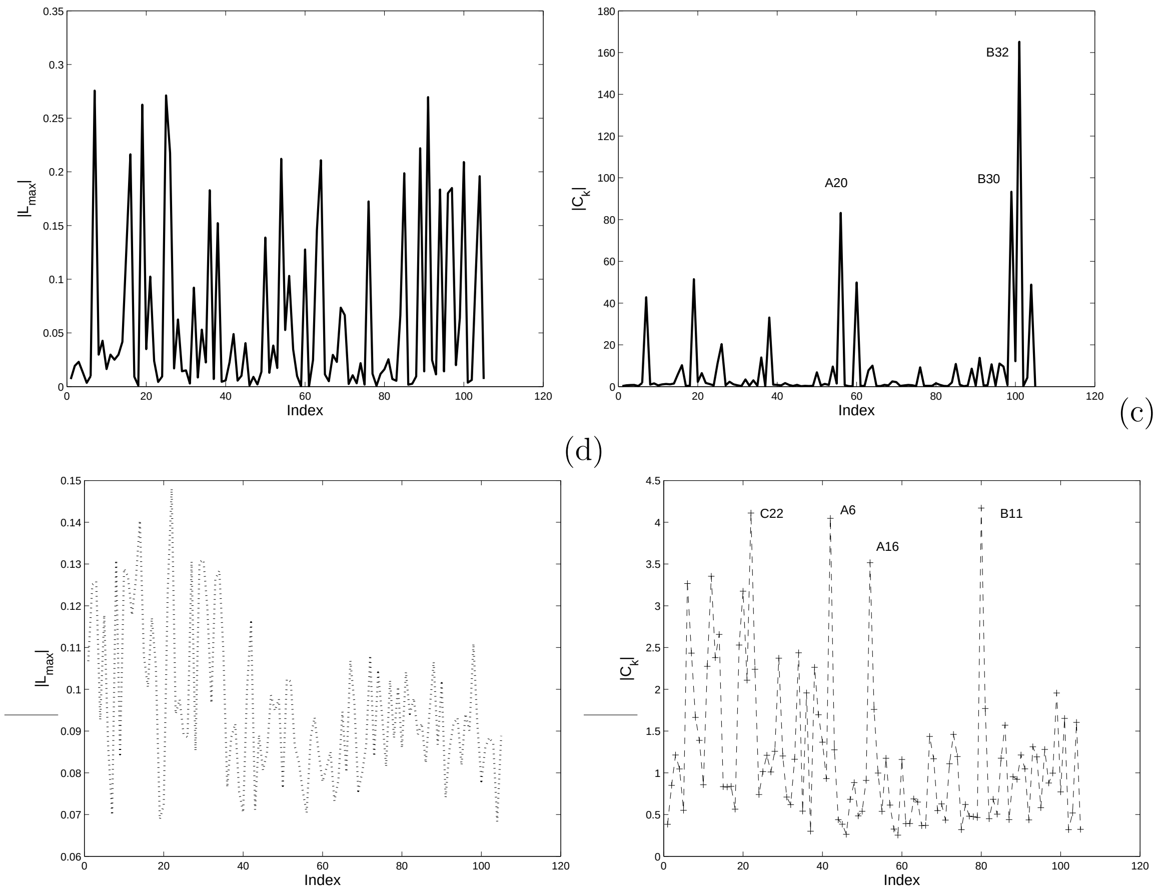


Figure 5: Dental plaque index data set. Perturbation of the of the explanatory variable. Index plot of (a) $|lmax|$ and (b) $|C_k|$. (—) and (...) denotes the index plot for the SN-MEM and N-MEM, respectively.

(a) (b)



6 Final conclusions

In this work we have treated the problem of estimation, hypothesis testing and influence diagnostics to the null intercept measurement error model under the skew-normal distribution. Parameter estimates are obtained via maximum likelihood considering the ECM algorithm, yielding closed form expression for the equations in the CM-step. Hypothesis testing is approached by using likelihood ratio, score and Wald statistics. We also have derived the local influence methods for SN-MEM in order to evaluate the effect of a small perturbation in the model or the data and different perturbation schemes were investigated. We applied the proposed methodology considering the real data set analyzed previously in Hadgu and Koch (1999). The main conclusion is that the skew-normal model presents a better fit and influent observations are different from those obtained when we consider the normal model. The conclusion of the analysis of the data set regarding the questions of interest are the same in all of the considered models. Finally, we want to mention that this work extends the early results found in Lachos et al. (2006).

Acknowledgements

The authors acknowledges the partial financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Appendix A: The score function

In this appendix we give the necessary expressions to obtain the first order partial derivatives of the log-likelihood function in (14) with respect to $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p$, $\sigma^2_{e_1},\ldots,\sigma^2_{e_p}$, σ^2_u, μ_x, ϕ_x and λ_x. Let $\mathbf{y}_{ij}=(y_{1ij},\cdots,y_{mij})^\top$, $W_{1ij}=X_{ij}-\mu_x$, $\boldsymbol{W}_{2ij}=\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x$, bij=$(\mathbf{z}_{ij}-\boldsymbol{\mu}_i)^\top\mathbf{B}_i(\mathbf{z}_{ij}-\boldsymbol{\mu}_i)$, with $\mathbf{B}_i=\mathrm{D}^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}\boldsymbol{\beta}_{0i}^\top\mathrm{D}^{-1}(\boldsymbol{\phi}_i)$ and let

$$(I)\text{ represent}\frac{\partial\log|\boldsymbol{\Sigma_i}|}{\partial\boldsymbol{\gamma}},\ (II)\text{ represent}\frac{\partial g_{ij}}{\partial\boldsymbol{\gamma}},\ (III)\text{ represent}\frac{\partial a_{ij}}{\partial\boldsymbol{\gamma}},\ (IV)\text{ represent}\frac{\partial A_i}{\partial\boldsymbol{\gamma}},$$

with $\boldsymbol{\gamma}=\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p,\sigma^2_{e_1},\ldots,\sigma^2_{e_p},\sigma^2_u,\mu_x,\phi_x,\lambda_x, i=1,\ldots,p, j=1,\ldots,n_i$, then:

a)-$\boldsymbol{\gamma}=\boldsymbol{\beta}_i$: $i=1,\ldots,p$

$$(I):\frac{2\phi_x}{\sigma^2_{e_i}c_i}\boldsymbol{\beta}_i;\quad (II):-\frac{2}{\sigma^2_{e_i}}\left[\mu_x(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)+\frac{\phi_x}{c_i}a_{ij}(\mathbf{y}_{ij}-2\boldsymbol{\beta}_i\mu_x)-\frac{\phi_x^2}{c_i^2}b_{ij}\boldsymbol{\beta}_i\right];$$

$$(III):\frac{1}{\sigma^2_{ei}}(\boldsymbol{W}_{2ij}-\mu_x\boldsymbol{\beta}_i);\quad (IV):-\frac{(2c_i+\lambda_x^2)}{\lambda_x^2\sigma^2_{ei}}A_i^3\boldsymbol{\beta}_i.$$

b)-$\boldsymbol{\gamma}=\sigma^2_{e_i}$: $i=1,\ldots,p$

$$(I):\frac{1}{\sigma^2_{e_i}}\left(p-\frac{\phi_x}{\sigma^2_{e_i}ci}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i\right);$$

$$(II): -\frac{1}{\sigma_{e_i}^4}\left[\frac{\phi_x^2}{c_i^2} b_{ij}\boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i + (\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x) - \frac{2\phi_x a_{ij}}{c_i}\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)\right];$$

$$(III): -\frac{1}{\sigma_u^4}\boldsymbol{\beta}_i^\top \boldsymbol{W}_{2ij}; \quad (IV): \frac{(2c_i + \lambda_x^2)}{2\lambda_x^2\sigma_{ei}^4} A_i^3 \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i.$$

c)-$\boldsymbol{\gamma} = \sigma_u^2$:

$$(I): \frac{c_i\sigma_u^2 - \phi_x}{c_i\sigma_u^4};$$

$$(II): -\frac{1}{\sigma_u^4}\left[\frac{\phi_x^2 b_{ij}}{c_i^2} + \left(1 - \frac{2\phi_x}{c_i\sigma_u^2}\right)(X_{ij} - \mu_x)^2 - \frac{2\phi_x}{c_i\sigma_{e_i}^2}(X_{ij} - \mu_x)[(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)^\top \boldsymbol{\beta}_i]\right];$$

$$(III): -\frac{1}{\sigma_u^4}W_{1ij}; \quad (IV): \frac{(2c_i + \lambda_x^2)}{2\lambda_x^2\sigma_u^4}A_i^3.$$

d)-$\boldsymbol{\gamma} = \mu_x$:

$$(I): 0; \quad (II): -\frac{2a_{ij}}{c_i}; \quad (III): -\frac{c_i - 1}{\phi_x}; \quad (IV): 0.$$

e)-$\boldsymbol{\gamma} = \phi_x$:

$$(I): \frac{c_i - 1}{\phi_x c_i}; \quad (II): -\frac{b_{ij}}{c_i^2}; \quad (III): 0; \quad (IV): \frac{(2c_i + \lambda_x^2 - c_i^2)}{2c_i^2\lambda_x^2\Lambda_i^2}A_i^3.$$

f)-$\boldsymbol{\gamma} = \lambda_x$:

$$(I): 0; \quad (II): 0; \quad (III): 0; \quad (IV): \frac{\phi_x}{\Lambda_i^2\lambda_x^3}A_i^3.$$

with $\boldsymbol{\beta}_{0i}$, $\boldsymbol{\beta}_i$ and $\mathbf{z}_{ij}$ as given in (6), $\boldsymbol{\phi}_i$ as in (7), $\boldsymbol{\mu}_i$ as in (11) and c_i and a_{ij} as given in Section 2.1, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$.

Appendix B: The observed information matrix

Here we derive the necessary formulas to obtain the second order partial derivatives of the log-likelihood function in (18) with respect to $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$, $\sigma_{e_1}^2, \ldots, \sigma_{e_p}^2$, σ_u^2, μ_x, ϕ_x and λ_x. Let

(I) represent $\dfrac{\partial^2 \log|\boldsymbol{\Sigma_i}|}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}}$, (II) represent $\dfrac{\partial^2 g_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}}$, (III) represent $\dfrac{\partial^2 a_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}}$ and (IV) represent $\dfrac{\partial^2 A_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}}$

for $\boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_p, \sigma_{e_1}^2, \cdots, \sigma_{e_p}^2, \sigma_u^2, \mu_x, \phi_x, \lambda_x$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, then we have:

a)-$\boldsymbol{\gamma} = \boldsymbol{\tau} = \boldsymbol{\beta}_i$, $i = 1, \ldots, p$:

$$(I) : \frac{2\phi_x}{\sigma_{e_i}^2 c_i}\left(I_m - \frac{2\phi_x}{c_i\sigma_{e_i}^2}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top\right);$$

$$(II) : \quad -\frac{2\mu_x}{\sigma_{e_i}^2}(2\phi_x\frac{a_{ij}}{c_i} - 1)\mathbf{I}_m + \frac{2\phi_x}{\sigma_{e_i}^4}\left(2\phi_x\frac{a_{ij}}{c_i^2}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)\boldsymbol{\beta}^\top - \frac{1}{c_i}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)^\top - \right.$$
$$\left. 4\phi_x^2\frac{b_{ij}}{c_i^3}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top + 2\phi_x\frac{a_{ij}}{c_i^2}\boldsymbol{\beta}_i(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)^\top\right);$$

$$(III) : -2\frac{\mu_x}{\sigma_{ei}^2}\boldsymbol{I}_m; \quad (IV) : -(4\frac{\phi_x}{\lambda_x^2}A_i^3 - \frac{3(2c_i + \lambda_x^2)^2}{\lambda_x^4}A_i^5)\frac{1}{\sigma_{ei}^4}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top - \frac{2c_i + \lambda_x^2}{\lambda_x^2\sigma_{ei}^2}A_i^3\boldsymbol{I}_m.$$

b)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \boldsymbol{\beta}_l$ for $i \neq l$, $i, l = 1, \ldots, p$:

$$(I) : \mathbf{0}; \quad (II) : \mathbf{0}; \quad (III) : \mathbf{0}; \quad (IV) : \mathbf{0}.$$

c)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \sigma_{e_i}^2$, $i = 1, \ldots, p$:

$$(I) : -\frac{2\phi_x}{c_i\sigma_{e_i}^4}\left(1 - \frac{\phi_x}{c_i\sigma_{e_i}^2}[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i]\right)\boldsymbol{\beta}_i;$$

$$(II) : \quad \frac{2}{\sigma_{e_i}^2}\left[\mu_x(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x) + \phi_x\frac{a_{ij}}{c_i}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x) - \phi_x^2\frac{b_{ij}}{c_i^2}\boldsymbol{\beta}_i + \phi_x^2\frac{b_{ij}}{2c_i}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i - \right.$$
$$\left. \frac{\phi_x}{c_i}(\mathbf{y}_{ij-}2\boldsymbol{\beta}_i\mu_x)\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x) - \phi_x^3\frac{b_{ij}}{c_i^3}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i + \phi_x\frac{a_{ij}}{c_i^2}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)\right];$$

$$(III) : \frac{1}{\sigma_{ei}^4}(\boldsymbol{W}_{2ij} - \mu_x\boldsymbol{\beta}_i); \quad (IV) : [2\frac{\phi_x}{\lambda_x^2}A_i^3 - \frac{3(2c_i + \lambda_x^2)^2}{2\lambda_x^4}A_i^5]\frac{1}{\sigma_{ei}^6}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i\boldsymbol{\beta}_i + \frac{2c_i + \lambda_x^2}{\lambda_x^2}A_i^3\frac{1}{\sigma_{ei}^4}\boldsymbol{\beta}_i.$$

d)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \sigma_{e_l}^2$ for $i \neq l$, $i, l = 1, \ldots, p$:

$$(I) : \mathbf{0}_m; \quad (II) : \mathbf{0}_m; \quad (III) : \mathbf{0}_m; \quad (IV) : \mathbf{0}_m.$$

e)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \sigma_u^2$, $i = 1, \ldots, p$:

$$(I) : \frac{2}{\sigma_{e_i}^2}\left(\frac{\phi_x}{c_i\sigma_u^2}\right)^2\boldsymbol{\beta}_i;$$

$$(II) : \quad \frac{2}{\sigma_{e_i}^2\sigma_u^4}\left[\left(\frac{\phi_x}{c_i}(X_{ij} - \mu_x) - \frac{\phi_x^2}{c_i^2}a_{ij}\right)(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x) + \left(\frac{\phi_x^3}{c_i}b_{ij} - \frac{\phi_x^2}{c_i^2}a_{ij}(X_{ij} - \mu_x)\right)\boldsymbol{\beta}_i\right];$$

$$(III): \mathbf{0}_m; \quad (IV): [2\frac{\phi_x}{\lambda_x^2}A_i^3 - \frac{3(2c_i+\lambda_x^2)^2}{2\lambda_x^4}A_i^5]\frac{1}{\sigma_u^4\sigma_{ei}^2}\boldsymbol{\beta}_i.$$

f)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \mu_x$, $i = 1, \ldots, p$:

$$(I): \mathbf{0}_m; \quad (II): -\frac{2}{\sigma_{e_i}^2}\left[\frac{1}{c_i}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x) - 2\phi_x\frac{a_{ij}}{c_i^2}\boldsymbol{\beta}_i\right]; \quad (III): -\frac{2}{\sigma_{ei}^2}\boldsymbol{\beta}_i^\top; \quad (IV): \mathbf{0}_m.$$

g)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \phi_x$, $i = 1, \ldots, p$:

$$(I): \frac{2}{\sigma_{e_i}^2 c_i^2}\boldsymbol{\beta}_i; \quad (II): -\frac{2}{\sigma_{e_i}^2 c_i^2}\left[a_{ij}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x) - \phi_x\frac{b_{ij}}{c_i}\boldsymbol{\beta}_i\right]; \quad (III): \mathbf{0}_m;$$

$$(IV): -[\frac{2(c_i-1)}{\lambda_x^2\phi_x}A_i^3 + \frac{3(2c_i+\lambda_x^2)(2c_i+\lambda_x^2-c_i^2)}{2\lambda_x^4\phi_x^2}A_i^5]\frac{1}{\sigma_{ei}^2}\boldsymbol{\beta}_i.$$

h)-$\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \lambda_x$, $i = 1, \ldots, p$:

$$(I): \mathbf{0}_m; \quad (II): \mathbf{0}_m; \quad (III): \mathbf{0}_m; \quad (IV): \frac{\phi_x A_i^3}{\lambda_x^5\Lambda_i^2}(-3A_i^2(2c_i+\lambda_x^2) + 4\lambda_x^2\Lambda_i)\frac{1}{\sigma_{ei}^2}\boldsymbol{\beta}_i.$$

i)-$\boldsymbol{\gamma} = \sigma_{e_i}^2$ and $\boldsymbol{\tau} = \sigma_{e_i}^2$, $i = 1, \ldots, p$:

$$(I): -\frac{1}{\sigma_{e_i}^4}\left[p - \frac{\phi_x}{\sigma_{e_i}^2 c_i}\left(2 - \frac{\phi_x}{\sigma_{e_i}^2 c_i}[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i]\right)[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i]\right];$$

$$(II): \frac{2}{\sigma_{e_i}^6}\left[\frac{\phi_x^2}{c_i^2}b_{ij}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i + (\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x) - \frac{2\phi_x a_{ij}}{c_i}\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)\right] -$$
$$\frac{2}{\sigma_{e_i}^8}\left[\left(\phi_x^3\frac{b_{ij}}{c_i^3}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i - \frac{\phi_x^2}{c_i^2}a_{ij}\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)\right)\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i - \right.$$
$$\left.\left(\phi_x^2\frac{a_{ij}}{c_i^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i - 2\phi_x(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i\right)\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)\right];$$

$$(III): \frac{2}{\sigma_{ei}^6}\boldsymbol{W}_{2ij}^\top\boldsymbol{\beta}_i, \quad (IV): [-\frac{\phi_x}{\lambda_x^2}A_i^3 + \frac{3(2c_i+\lambda_x^2)^2}{4\lambda_x^4}A_i^5]\frac{1}{\sigma_{ei}^8}(\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i)^2 - \frac{2c_i+\lambda_x^2}{\lambda_x^2}A_i^3\frac{1}{\sigma_{ei}^6}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i.$$

j)-$\boldsymbol{\gamma} = \sigma_{e_i}^2$ and $\boldsymbol{\tau} = \sigma_{e_l}^2$, for $i \neq l$, $i, l = 1, \ldots, p$:

$$(I): 0; \quad (II): 0; \quad (III): 0; \quad (IV): 0.$$

k)-$\boldsymbol{\gamma} = \sigma_{e_i}^2$ and $\boldsymbol{\tau} = \sigma_u^2$, $i = 1, \ldots, p$:

$$(I): -\left(\frac{\phi_x}{\sigma_u^2\sigma_{e_i}^2 c_i}\right)^2[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i];$$

$$(II): -\frac{2}{\sigma_{e_i}^4\sigma_u^4}\left[\left(\frac{\phi_x^3}{c_i}b_{ij}-\frac{\phi_x^2}{c_i^2}a_{ij}(X_{ij}-\mu_x)\right)\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i+\left(\frac{\phi_x}{c_i}(X_{ij}-\mu_x)-\frac{\phi_x^2}{c_i^2}a_{ij}\right)\boldsymbol{\beta}_i(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)\right];$$

$$(III): 0;\quad (IV): [-\frac{\phi_x}{\lambda_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)^2}{4\lambda_x^4}A_i^5]\frac{1}{\sigma_u^2\sigma_{ei}^2}\mathbf{1}_m^\top\boldsymbol{\beta}_i.$$

l)-$\boldsymbol{\gamma}=\sigma_{e_i}^2$ and $\boldsymbol{\tau}=\mu_x$, $i=1,\ldots,p$:

$$(I): 0;\quad (II): \frac{2}{c_i\sigma_{e_i}^2}\left[\frac{\phi_x}{c_i}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i-2(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i\right];\quad (III): \frac{1}{\sigma_{ei}^4}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i;\quad (IV): 0.$$

m)-$\boldsymbol{\gamma}=\sigma_{e_i}^2$ and $\boldsymbol{\tau}=\phi_x$, $i=1,\ldots,p$:

$$(I): -\frac{1}{(c_i\sigma_{e_i}^2)^2}[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i];\quad (II): -\frac{1}{c_i^2\sigma_{e_i}^2}\left[\phi_x\frac{b_{ij}}{c_i}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i-2a_{ij}\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)\right];\quad (III): 0;$$

$$(IV): [\frac{(c_i-1)}{\lambda_x^2\phi_x}A_i^3+\frac{3(2c_i+\lambda_x^2)(2c_i+\lambda_x^2-c_i^2)}{4\lambda_x^4\phi_x^2}A_i^5]\frac{1}{\sigma_{ei}^4}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i.$$

n)-$\boldsymbol{\gamma}=\sigma_{e_i}^2$ and $\boldsymbol{\tau}=\lambda_x$, $i=1,\ldots,p$:

$$(I): 0;\quad (II): 0;\quad (III): 0;\quad (IV): \frac{\phi_x A_i^3}{2\lambda_x^5\Lambda_i^2}[3A_i^2(2c_i+\lambda_x^2)-4\lambda_x^2\Lambda_i]\frac{1}{\sigma_{ei}^4}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i.$$

o)-$\boldsymbol{\gamma}=\sigma_u^2$ and $\boldsymbol{\tau}=\sigma_u^2$:

$$(I): -\frac{1}{\sigma_u^4}\left[1-\frac{\phi_x}{c_i\sigma_u^2}\left(2-\frac{\phi_x}{c_i\sigma_u^2}\right)\right];$$

$$(II): \frac{2}{\sigma_u^6}\left[\frac{\phi_x^2 b_{ij}}{c_i^2}+\left(1-\frac{2\phi_x}{c_i\sigma_u^2}\right)(X_{ij}-\mu_x)^2-\frac{2\phi_x}{c_i\sigma_{e_i}^2}(X_{ij}-\mu_x)[(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i]\right]-$$

$$-\frac{2\phi_x}{\sigma_u^8}\left[\phi_x^2\frac{b_{ij}}{c_i}-\phi_x\frac{a_{ij}}{c_i^2}(X_{ij}-\mu_x)-2(X_{ij}-\mu_x)^2(\frac{\phi_x}{c_i^2\sigma_u^2}-\frac{1}{c_1})-\frac{\phi_x}{c_i^2}(X_{ij}-\mu_x)(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i\right];$$

$$(III): \frac{2}{\sigma_u^6}W_{1ij};\quad (IV): [-\frac{\phi_x}{\lambda_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)^2}{4\lambda_x^4}A_i^5]\frac{1}{\sigma_u^8}-\frac{2c_i+\lambda_x^2}{\lambda_x^2}A_i^3\frac{1}{\sigma_u^6}.$$

p)-$\boldsymbol{\gamma}=\sigma_u^2$ and $\boldsymbol{\tau}=\mu_x$:

$$(I): 0;\quad (II): -\frac{2}{\sigma_u^4}\left[\phi_x\frac{a_{ij}}{c_i^2}-\frac{(X_{ij}-\mu_x)}{c_i}\right];\quad (III): \frac{1}{\sigma_u^4};\quad (IV): 0.$$

q)-$\boldsymbol{\gamma}=\sigma_u^2$ and $\boldsymbol{\tau}=\phi_x$:

$$(I): -\frac{1}{(c_i\sigma_u^2)^2};\quad (II): -\frac{2}{\sigma_u^4}\left[\phi_x\frac{b_{ij}}{c_i^3}+2a_{ij}\frac{(X_{ij}-\mu_x)}{c_i^2}\right];\quad (III): 0;$$

$$(IV):[\frac{(c_i-1)}{\lambda_x^2\phi_x}A_i^3+\frac{3(2c_i+\lambda_x^2)(2c_i+\lambda_x^2-c_i^2)}{4\lambda_x^4\phi_x^2}A_i^5]\frac{1}{\sigma_u^4}.$$

r)-$\boldsymbol{\gamma}=\sigma_u^2$ and $\boldsymbol{\tau}=\lambda_x$:

$$(I):0;\quad (II):0;\quad (III):0;\quad (IV):\frac{\phi_x A_i^3}{2\lambda_x^5\Lambda_i^2}[3A_i^2(2c_i+\lambda_x^2)-4\lambda_x^2\Lambda_i]\frac{1}{\sigma_u^4}.$$

s)-$\boldsymbol{\gamma}=\mu_x$ and $\boldsymbol{\tau}=\mu_x$:

$$(I):0;\quad (II):\frac{2}{c_i}(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_{e_i}^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i);\quad (III):0;\quad (IV):0.$$

t)-$\boldsymbol{\gamma}=\mu_x$ and $\boldsymbol{\tau}=\phi_x$:

$$(I):0;\quad (II):\frac{2a_{ij}}{c_i^2}(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_{e_i}^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i);\quad (III):0;\quad (IV):0.$$

u)-$\boldsymbol{\gamma}=\mu_x$ and $\boldsymbol{\tau}=\lambda_x$:

$$(I):0;\quad (II):0;\quad (III):0;\quad (IV):0.$$

v)-$\boldsymbol{\gamma}=\phi_x$ and $\boldsymbol{\tau}=\phi_x$:

$$(I):\frac{1}{\phi_x c_i}\left(\frac{1}{\phi_x}+\frac{1}{c_i}\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}\right)-\frac{1}{\phi_x^2};\quad (II):-2\frac{b_{ij}}{c_i^3}(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_{e_i}^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i);\quad (III):0;$$

$$(IV):-\frac{\lambda_x^2+1}{\lambda_x^2\phi_x^3}A_i^3+\frac{3(2c_i+\lambda_x^2-c_i^2)^2}{4\lambda_x^4\phi_x^4}A_i^5.$$

x)-$\boldsymbol{\gamma}=\phi_x$ and $\boldsymbol{\tau}=\lambda_x$:

$$(I):0;\quad (II):0;\quad (III):0;\quad (IV):\frac{c_i-2}{\lambda_x^3\Lambda_i\phi_x}A_i^3+\frac{3(2c_i+\lambda_x^2-c_i^2)}{2\lambda_x^5\Lambda_i^2\phi_x}A_i^5.$$

z)-$\boldsymbol{\gamma}=\lambda_x$ and $\boldsymbol{\tau}=\lambda_x$:

$$(I):0;\quad (II):0;\quad (III):0;\quad (IV):-\frac{3\phi_x}{\lambda_x^4\Lambda_i^2}A_i^3+\frac{3\phi_x^2}{\lambda_x^6\Lambda_i^4}A_i^5.$$

with $\mathbf{B}_i$, b_{ij}, $\boldsymbol{\beta}_{0i}$, $\boldsymbol{\beta}_i$, $\mathbf{y}_{ij}$, $\mathbf{z}_{ij}$, $\boldsymbol{\phi}_i$, $\boldsymbol{\mu}_i$, c_i, a_{ij},W_{1ij} and $\boldsymbol{W}_{2ij}$ as given in Appendix A, $i=1,\ldots,p$, $j=1,\ldots,n_i$.

References

Abdullah, M. B. (1995). Detection of influential observations in functional errors-in-variables model. *Communications in Statistics: Theory and Methods*, 24, 1585-1595.

Aoki, R., Bolfarine, H. and Singer, J. M. (2001). Null intercept measurement error regression models. *TEST*, 10, 441-454.

Aoki R., Bolfarine, H., Achcar J. A. and Leão Pinto Jr D. (2003). Bayesian Analysis of a Multivariate Null Intercept Error-in-Variables Regression Model, *Journal of Biopharmaceutical Statistics*, vol. 13, 4, 763-771.

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2005). Skew-Normal Linear Mixed Models, *Journal of Data Science*, 3, 415-438.

Arellano-Valle, R. B. and Genton, M. G. (2005). On Fundamental Skew Distributions. *Journal of Multivariate Analysis*, 96, 93-116.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics,* 12, 171-178.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika,* 83, 715-726.

Bradley, R.A. Gart, J.J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika,* 49, 205-214.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression.* Wiley, new York.

Cheng, C. L. and Ness, V. (1999) - Statistical Regression with Measurement Error, first edition, Oxford University Press.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.

Cook, R. D. (1986). Assessment of local influence. *Journal of Royal Statistical Society*, 48, 133-169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, London.

Cysneiros, F. J. A., Paula, G. A. (2004). One-sided tests in linear models with multivariate t-distribution. *Communications in Statistics: Simulation and Computation*, 33, 747-772.

Dempster, A.P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society,* B, 39, 1-22.

Escobar, E. and Meeker, W. (1992). Assessing influence in regression analysis with censured data. *Biometrics*, 48, 507-528.

Fahrmeir, L., Klinger, J. (1994). Estimating and testing generalized linear models under inequality restrictions. *Statistical Papers*, 59, 981-995.

Fuller, A. (1987) - *Measurement Error Models*, first edition, Wiley.

Hadgu, A. and Koch, G. (1999). Application of generalized estimating equations to a dental randomized clinical trial. *Journal of Biopharmaceutical Statistics*, 9, 161-178.

Henze, N. (1986). A probabilistic representation of the "skew normal" distribution. *Scandinavian Journal of Statistics*, 13, 271-275.

Johnson, N. L. Kotz, S. and Balakrishnan, N. (1994). *Continuos Univariate Distributions*, Vol. 1. New York. Wiley.

Kelly, G. (1984). The influence function in the errors in variables problem. *The Annals of Statistics*, 12, 87-100.

Kim, M. G. (2000). Outliers and influential observations in the structural errors-in-variables model. *Journal Applied Statistics*, 24, 461-473.

Labra, F., Aoki, R. and Bolfaine, H. (2005). Local influence in null intercept measurement error regression under a student-t model. *Journal Applied Statistics*, 32, 723-740.

Lachos, V. H., Montenegro, L. C. and Bolfarine, H. (2006). Inference and assessment of local influence in skew-normal null intercept measurement error model. *Journal of Statistical Computation and Simulation*. To appear.

Meng, X. and Rubin, D.B. (1993). Maximum likelihood estimation via ECM-algorithm: a general framework. *Biometrika*, 80, 2, 267-278.

Nyquist, H. (1991). Restricted estimation of generalized linear models. *Journal Applied Statistics*, 40, 133-141.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* Springer. New York.

Wellman, J. M. and Gunst, R. F. (1991). Influence diagnostics for linear measurement error models. *Biometrika*, 78, 373-380.

Zhao, Y. and Lee, A. (1998). Influence diagnostics for simultaneous equations models. *Australian and New Zealand Journal Statistics*, 40, 345-357.