Brazilian Journal of Probability and Statistics
2011, Vol. 25, No. 2, 145–170
DOI: 10.1214/09-BJPS114

Skew-normal distribution in the multivariate null intercept measurement error model

F. V. Labra[a], **R. Aoki**[b], **V. Garibay**[b] **and V. H. Lachos**[a]

[a]*IMECC/UNICAMP*
[b]*ICMC/USP-São Carlos*

Abstract. In this paper we discuss inferential aspects and the local influence analysis of the multivariate null intercept measurement error model where the unobserved value of the covariate (latent variable) follows a skew-normal distribution. In order to develop the hypotheses testing of interest and the local influence diagnostics, closed-form expressions of the marginal likelihood, the score function and the observed information matrix are presented. Additionally, an EM-type algorithm for evaluating the unrestricted and restricted maximum likelihood estimates of the parameters under equality constraints on the regression coefficients is examined. Also, 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 [*Journal of Biopharmaceutical Statistic* **9** (1999) 161–178].

1 Introduction

Error-in-variables regression models constitute an attractive alternative to modeling many practical experimental problems, especially when the same responses are observed on the same units under different experimental conditions. A wide bibliography can be found, for instance, in Cheng and Van-Ness (1999). Consider a bivariate random variable (η_i, ξ_i) satisfying the linear relation $\eta_i = \alpha + \beta\xi_i$, $i = 1, \dots, n$. Suppose that η_i and ξ_i can not be observed directly, but instead we observe

$$x_i = \xi_i + \delta_i, \tag{1.1}$$

$$y_i = \eta_i + \varepsilon_i, \tag{1.2}$$

where y_i and x_i, $i = 1, \dots, n$, respectively denote the observed values of the response and explanatory variables, ξ_i, $i = 1, \dots, n$, indicate the true values of the latter, α represents the (unknown) intercept, β stands for the (unknown) slope and the errors δ_i and ε_i have zero means and unknown variances σ_δ^2 and σ^2, respectively and are independently distributed, $i = 1, \dots, n$. It is common to assume

Key words and phrases. EM algorithm, hypothesis testing, maximum likelihood, measurement error, local influence, skewness.
Received December 2008; accepted December 2009.

that all the random variables in the error-in-variables regression model are jointly normal. In this case it is well known that such a model is not identifiable and to bypass this inconvenience, we must make an extra assumption about the parameters. Among the available alternatives, we identify (i) σ^2 or (and) σ_δ^2 known, (ii) $\lambda = \sigma^2/\sigma_\delta^2$ known, (iii) $k_x = \sigma_x^2/(\sigma_x^2 + \sigma_\delta^2)$ known or (iv) α known, where σ_x^2 is the variance of ξ_i, $i = 1, \dots, n$. In the normal structural model with no side conditions, $E(\xi_i) = \mu_x$ is the only parameter that is identifiable. Chan and Mak (1979) studied the maximum likelihood estimation of the parameters when the intercept is known and presented the information matrix explicitly. Patefield (1985) considered the case where the intercept and the ratio of the variances are known. Aoki, Bolfarine and Singer (2001) discussed the model with known intercept and repeated measurement data. The first development of the class of skew-normal measurement error models is given in Arellano-Valle et al. (2005), and Lachos, Montenegro and Bolfarine (2008) considered the model defined in (1.1) and (1.2) with known intercept where the true value of the covariate follows a skew-normal distribution and discussed the maximum likelihood estimation, as well as the influence diagnostics analysis. In this paper we extend the model defined in Lachos, Montenegro and Bolfarine (2008) for a multivariate context.

The skew-normal distribution represents a superset of the normal family and has a shape parameter that defines the direction of the asymmetry of the distribution. Motivation for using such general structures include easiness of interpretation, as well as estimation efficiency and the most important fact is that there are many real datasets presenting clear indication of skewness in diverse areas, such as engineering, medicine, psychology and agriculture, among others (Genton, 2004). Although the idea of modeling a parametric class of asymmetric distributions which are analytically tractable and that can accommodate practical values of skewness and kurtosis, including the normal distribution, was proposed in the literature for a long time, it was Azzalini (1985) who thoroughly set the foundations for the univariate skew-normal distribution. An extension to the multivariate setting was proposed by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) emphasized the statistical applications of the multivariate skew-normal distribution. Generalizations of these ideas have been proposed by many authors. For example, skew-t distributions (Azzalini and Capitanio, 2003); skew-elliptical distributions (Branco and Dey, 2001); and fundamental skew distributions (Arellano-Valle and Genton, 2005). The book edited by Genton (2004) presents a recent overview of the skewed distributions including many real problems.

Despite the interesting properties of the skew-normal distribution, it is well known that when the "direct parameterizations"—as defined in Azzalini (1985) and Azzalini and Capitanio (1999)—is used, the Fisher information matrix is singular for $\lambda = 0$. Pewsey (2000) discuss the problem of maximization of the direct parameterizations of the model defined in Azzalini (1985) and Azzalini and Capitanio (1999). Sartori (2006) proposed the use of a modified score function as an estimating equation for the shape parameter to avoid the problem of finding the maximum likelihood (ML) of the shape parameter to be infinity with positive probability for moderate sample sizes. Liseo and Loperfido (2006) proposed different methods to deal with the problem of the shape parameter from a Bayesian perspective. Recently, Loperfido (2010) deals with this problem in the multivariate setting and proposed a canonical transformation to circumvent the problem, being based on the maximizing sample skewness. However, this methodology is hard to be applied in complicated models as is the case of our proposed model. We would like to call attention to the fact that this problem is only confined to the shape parameter and this rarely appears in practice in a multivariate context.

Hadgu and Koch (1999) analyzed a clinical dataset where 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, after three months and after six months from the baseline with the use of the corresponding mouth rinses (A, B or control C). The dental plaque was scored by the Turesky modification of the Quingley–Hein index (Turesky, Gilmore and Glickman, 1970), a continuous measurement. As the measurements are subject to measurement error, Aoki et al. (2003) proposed the use of the measurement error model. In addition, hence the plaque index was collected at baseline and two followup times, to incorporate a possible within subjects correlation structure, the structural model where the true unobserved value of the covariate follows a normal distribution was considered. No intercept were included in the proposed model, since null dental plaque index at baseline imply null expected post test values after the use of each mouth rinse, that is, the dental plaque index should not increase after the use of each mouth rinse. The proposed model was given by

$$\mathbf{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \tag{1.3}$$

$$\mathbf{y}_{1,i} = \beta_{1,i}\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_{1,i}, \tag{1.4}$$

$$\mathbf{y}_{2,i} = \beta_{2,i}\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_{2,i}, \qquad i = 1, \dots, p, \tag{1.5}$$

where $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{in_i})$, $\mathbf{y}_{k,i}^\top = (y_{k,i1}, \dots, y_{k,in_i})$, $\boldsymbol{\xi}_i^\top = (\xi_{i1}, \dots, \xi_{in_i})$, $\boldsymbol{\delta}_i^\top = (\delta_{i1}, \dots, \delta_{in_i})$, $\boldsymbol{\varepsilon}_{k,i}^\top = (\varepsilon_{k,i1}, \dots, \varepsilon_{k,in_i})$, with $\delta_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_\delta^2)$, $\varepsilon_{k,ij} \overset{\text{ind}}{\sim} N(0, \sigma_i^2)$, δ_{ij} and $\varepsilon_{k,ij}$ not correlated and independent of $\xi_{ij} \overset{\text{iid}}{\sim} N(\mu_x, \sigma_x^2)$, $k = 1, 2$, $i = 1, \dots, p$, $j = 1, \dots, n_i$. Considering the clinical trial, we have $i = 1, 2, 3$ representing the mouth rinse type: control, A and B, respectively, $k = 1$ (2) represents the plaque index after three (six) months from the baseline, so that $y_{1,ij}$ ($k = 1$) and $y_{2,ij}$ ($k = 2$) represents, respectively, the observed plaque index after three months and after six months from the beginning of the study, with the use of the mouth rinse i for the subject j and $\beta_{k,i}$ stands for the (unknown) slopes, while x_{ij} represents the observed plaque index at baseline for the jth subject that used the ith mouth rinse and ξ_{ij} the corresponding unobserved real plaque index at baseline. The parameters of interest are $\beta_{k,i}$, $k = 1, 2$, $i = 1, 2, 3$, as it represents the dental plaque reduction after three ($k = 1$) and after six ($k = 2$) months from the baseline, with the use of the ith mouth rinse.

A generalization of that proposed model can be considered with ξ_{ij} following a skew normal distribution, which is a class of distributions which includes the normal distribution as a special case. The asymmetry parameter of the skew normal distribution incorporates skewness in the latent variable ξ_{ij} and consequently in the observed quantities. If this parameter is set to 0, then the asymmetric model reduces to the normal model. So we are introducing more flexibility to the model to make it capable of adapting as closely as possible to the real data. Let $\mathbf{y}_i = (\mathbf{y}_{1,i}^\top, \mathbf{y}_{2,i}^\top)^\top$ represent the observed values of the response variables, here we extend it by considering that $\mathbf{y}_i$ is given by $\mathbf{y}_i = (\mathbf{y}_{1,i}^\top, \dots, \mathbf{y}_{m,i}^\top)^\top$, rather than $\mathbf{y}_i = (\mathbf{y}_{1,i}^\top, \mathbf{y}_{2,i}^\top)^\top$ as was defined in Aoki et al. (2003). Inspired by the work of Lachos, Montenegro and Bolfarine (2008), who considered the case in which $m = 1$, in this paper we consider the study of inference and the local influence analysis in the multivariate null intercept measurement error regression model with the assumption that the unknown quantity ξ_{ij} (latent variable) follows a univariate skew-normal distribution, implying that the observed vector $\mathbf{z}_i = (\mathbf{x}_i^\top, \mathbf{y}_{1,i}^\top, \dots, \mathbf{y}_{m,i}^\top)^\top$, $i = 1, \dots, p$, follows a multivariate skew-normal distribution (Arellano-Valle et al., 2005).

On the other hand, influence diagnostics is an important step in the analysis of a dataset, 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 to assess 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. Aoki, Singer and Bolfarine (2007) considered the local influence diagnostic for the null intercept measurement error model defined in Aoki, Bolfarine and Singer (2001). More recently, Lachos, Montenegro and Bolfarine (2008) applied the local influence method in the skew-normal null intercept measurement error model without a longitudinal structure, that is, $m = 1$ (the univariate case). Here, we extend those results to a model that allows the longitudinal structure (the multivariate case).

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). The EM-algorithm to obtain the maximum likelihood estimates of the parameters and an EM-type algorithm for evaluating the restricted maximum likelihood estimates under equality constraints on the regression coefficients are presented. The latter will be used to obtain the score and the likelihood ratio test statistics to test the hypotheses of interest. Section 3 contains the main concepts of local influence and the related concepts of diagnostics. Considering various perturbation schemes and the model proposed in Section 2, we derive the appropriate matrices to assess the normal curvature and to construct influence graphs that provide us an indication of bad model fitting or of influential observations. Finally, in Section 4 applications of the results and methods are illustrated with a numerical example and in Section 5 some final conclusions are discussed.

2 The skew-normal multivariate null intercept measurement error model and the hypotheses testing

To better motivate our proposed methodology, we give a brief introduction of the multivariate SN distribution. We say that a $k \times 1$ random vector $\mathbf{Y}$ follows a SN-distribution with $k \times 1$ location vector $\boldsymbol{\mu}$, $k \times k$ positive definite dispersion matrix $\boldsymbol{\Psi}$ and $k \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, and write $\mathbf{Y} \sim SN_k(\boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, if its probability density function (pdf) is given by

$$f(\mathbf{y}) = 2\phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Psi})\Phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Psi}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\big), \tag{2.1}$$

where $\phi_k(\cdot; \boldsymbol{\mu}, \boldsymbol{\Psi})$ stands for the pdf of the k-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Psi}$, $N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$ say, and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard univariate normal and $\boldsymbol{\Psi}^{1/2}$ satisfies $\boldsymbol{\Psi}^{1/2}\boldsymbol{\Psi}^{1/2} = \boldsymbol{\Psi}$. Note that for $\boldsymbol{\lambda} = \mathbf{0}$ (2.1) reduces to the symmetric $N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$-pdf, while for nonzero values of $\boldsymbol{\lambda}$, it produces a perturbed (asymmetric) family of $N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$-pdf's. Except for a straightforward difference in the parametrization considered in (2.1), this model corresponds to that introduced by Azzalini and Dalla Valle (1996), with properties extensively studied in Arellano-Valle and Genton (2005) and Arellano-Valle, Bolfarine and Lachos (2005). An interesting result, due to Arellano-Valle and Genton (2005), is the marginal stochastic representation of a SN random vector with pdf (2.1), which is given by

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}\big(\boldsymbol{\delta}|T_0| + (\mathbf{I}_k - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1\big), \qquad \text{with } \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \tag{2.2}$$

where $T_0 \sim N_1(0, 1)$ and $\mathbf{T}_1 \sim N_k(\mathbf{0}, \mathbf{I}_k)$ are independent, with $\mathbf{I}_k$ denoting an identity matrix of order k and "$\stackrel{d}{=}$" meaning "distributed as."

2.1 The model

A generalization of the model defined in (1.3), (1.4) and (1.5) may be obtained by considering

$$x_{ij} = \xi_{ij} + \delta_{ij}, \tag{2.3}$$

$$y_{k,ij} = \beta_{k,i}\xi_{ij} + \varepsilon_{k,ij}, \tag{2.4}$$

$k = 1, \dots, m$, $i = 1, \dots, p$, $j = 1, \dots, n_i$, where for the unobserved random variables ξ_{ij}, δ_{ij} and $\varepsilon_{k,ij}$ it is assumed that they are independent for all i, j, k, with

$$\xi_{ij} \stackrel{\text{iid}}{\sim} SN(\mu_x, \sigma_x^2, \lambda_x), \qquad \delta_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_\delta^2) \quad \text{and} \quad \varepsilon_{k,ij} \stackrel{\text{ind}}{\sim} N(0, \sigma_i^2). \tag{2.5}$$

The above model considers, for instance, that in the case of the Hadgu and Koch (1999) dataset the dental plaque index may not be symmetrically distributed in the population. On the other hand, the errors are related to measurement errors so that it is expected to be normally distributed. 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 nonnormality.

Let $\mathbf{z}_{ij} = (x_{ij}, y_{1,ij}, \dots, y_{m,ij})^\top$ be the vector of observations for the jth subject in the ith group, then from Arellano-Valle et al. (2005)—see also Lachos, Montenegro and Bolfarine (2008)—it follows that

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

$$f_{\mathbf{Z}_{ij}}(\mathbf{z}_{ij}) = 2\phi_{m+1}(\mathbf{z}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\Phi_1\big(\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)\big), \tag{2.6}$$

where $\boldsymbol{\mu}_i = \boldsymbol{\beta}_{0i}\mu_x$, $\boldsymbol{\Sigma}_i = \mathrm{D}(\boldsymbol{\phi}_i) + \sigma_x^2 \boldsymbol{\beta}_{0i}\boldsymbol{\beta}_{0i}^\top$, $\bar{\boldsymbol{\lambda}}_i = \lambda_x \sigma_x^2 \boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\beta}_{0i}/\sqrt{\sigma_x^2 + \lambda_x^2 \Lambda_i}$, with $\boldsymbol{\beta}_{0i} = (1, \boldsymbol{\beta}_i^\top)^\top$, $\boldsymbol{\beta}_i = (\beta_{1,i}, \dots, \beta_{m,i})^\top$, $\boldsymbol{\phi}_i = (\sigma_\delta^2, \sigma_i^2, \dots, \sigma_i^2)^\top$ and $\Lambda_i = \sigma_x^2/(1 + \sigma_x^2 \boldsymbol{\beta}_{0i}^\top \mathrm{D}^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i})$.

It follows that the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, {\boldsymbol{\sigma}^2}^\top, \sigma_\delta^2, \mu_x, \sigma_x^2, \lambda_x)^\top \in \mathbb{R}^{(m+1)p+4}$, with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_p^\top)^\top$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_p^2)^\top$, given the observed sample $\mathbf{z} = (\mathbf{z}_{11}^\top, \dots, \mathbf{z}_{1n_1}^\top, \dots, \mathbf{z}_{p1}^\top, \dots, \mathbf{z}_{pn_p}^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{p}\sum_{j=1}^{n_i} \ell_{ij}(\boldsymbol{\theta}), \tag{2.7}$$

$$\ell_{ij}(\boldsymbol{\theta}) = \log(2) - \frac{(m+1)}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}g_{ij} + \log(K_{ij}), \tag{2.8}$$

with $g_{ij} = (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)$, $K_{ij} = \Phi_1(\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{z}_{ij} - \boldsymbol{\mu}_i))$ and $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i$ as in (2.6). The score function and the observed information matrix are given in Appendix A and Appendix B, respectively. These derivatives can be used to obtain the asymptotic confidence intervals and also to test the hypotheses of interest. Note that algorithms such as Newton–Raphson (NR) can be implemented using these results. An oft-voiced complaint of the NR algorithm is that it may not converge unless good starting values are used. The EM algorithm (Dempster, Laird and Rubin, 1977), which takes advantage of being insensitive to the starting values is a powerful computational tool that requires the construction of unobserved data. It has been well developed and has become a broadly applicable approach to the iterative computation of ML estimates. One of the major reasons for the popularity of the EM algorithm is that the M-step involves only the complete data ML estimation, which is often computationally simple. Moreover, the EM algorithm is stable and straightforward to implement since the iterations converge monotonically and no second derivatives are required. In the next subsections we discuss the unrestricted and the restricted estimation of the parameters based on the EM algorithm, as well as the hypotheses testing of interest. First, note from (2.2) that

$$\mathbf{z}_{ij}|\xi_{ij} \overset{\text{ind}}{\sim} N_{m+1}(\boldsymbol{\beta}_{0i}\xi_{ij}, \mathrm{D}(\boldsymbol{\phi}_i)), \tag{2.9}$$

$$\xi_{ij}|T_{ij} = t_{ij} \overset{\text{ind}}{\sim} N_1(\mu_x + \sigma_x\delta_x t_{ij}, \sigma_x^2(1-\delta_x^2)), \tag{2.10}$$

$$T_{ij} \overset{\text{iid}}{\sim} HN_1(0,1), \tag{2.11}$$

$i = 1,\ldots,p,\ j = 1,\ldots,n_i$, all independent, where $HN_1(0,1)$ denote the standardized univariate half-normal distribution and $\delta_x = \lambda_x/(1+\lambda_x^2)^{1/2}$.

2.2 Maximum likelihood estimation

In this subsection we summarize the E-step and the M-step of the EM algorithm to obtain the maximum likelihood estimates of the parameters. From Arellano-Valle et al. (2005)—see also Lachos, Montenegro and Bolfarine (2008)—and considering the hierarchical representation (2.9)–(2.11) with $\upsilon^2 = \sigma_x^2(1-\delta_x^2)$ and $\tau = \phi_x^{1/2}\delta_x$, we can obtain the E-step and the M-step as follows.

E-step. Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ compute

$$\widehat{t}_{ij} = E[T_{ij}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}] = \widehat{\mu}_{T_{ij}} + W_{\Phi_1}\left(\frac{\widehat{\mu}_{T_{ij}}}{\widehat{N}_{Ti}}\right)\widehat{N}_{Ti}, \tag{2.12}$$

$$\widehat{t^2_{ij}} = E[T_{ij}^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}] = \widehat{\mu}^2_{T_{ij}} + \widehat{N}^2_{Ti} + W_{\Phi_1}\left(\frac{\widehat{\mu}_{T_{ij}}}{\widehat{N}_{Ti}}\right)\widehat{N}_{Ti}\widehat{\mu}_{T_{ij}},$$

$$\widehat{\xi}_{ij} = E[\xi_{ij}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}] = \widehat{c}_{ij} + \widehat{d}_i\widehat{t}_{ij},$$

$$\widehat{\xi^2_{ij}} = E[\xi_{ij}^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}] = \widehat{M}_i^2 + \widehat{c}^2_{ij} + 2\widehat{c}_{ij}\widehat{d}_i\widehat{t}_{ij} + \widehat{d}_i^2\widehat{t^2_{ij}} \quad \text{and}$$

$$\widehat{t\xi}_{ij} = E_{t_{ij},\xi_{ij}}[T_{ij}\xi_{ij}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{z}_{ij}] = \widehat{c}_{ij}\widehat{t}_{ij} + \widehat{d}_i\widehat{t^2_{ij}},$$

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

M-step. Update $\widehat{\boldsymbol{\theta}}$

$$\widehat{\beta}_{k,i} = \frac{\sum_{j=1}^{n_i} y_{k,ij}\widehat{\xi}_{ij}}{\sum_{j=1}^{n_i}\widehat{\xi^2_{ij}}}, \qquad k = 1,\ldots,m, i = 1,\ldots,p,$$

$$\widehat{\sigma}_i^2 = \frac{1}{mn_i}\left(\sum_{k=1}^{m}\sum_{j=1}^{n_i} y_{k,ij}^2 - \frac{\sum_{k=1}^{m}(\sum_{j=1}^{n_i} y_{k,ij}\widehat{\xi}_{ij})^2}{\sum_{j=1}^{n_i}\widehat{\xi}_{ij}^2}\right), \qquad i = 1, \ldots, p,$$

$$\widehat{\sigma}_\delta^2 = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}^2 - 2x_{ij}\widehat{\xi}_{ij} + \widehat{\xi}_{ij}^2), \qquad \widehat{\tau} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{t\xi}_{ij} - \mu_x\widehat{t}_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t}_{ij}^2},$$

$$\widehat{\mu}_x = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{\xi}_{ij}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t}_{ij}^2 - \sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t\xi}_{ij}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t}_{ij}}{N\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t}_{ij}^2 - (\sum_{i=1}^{p}\sum_{j=1}^{n_i}\widehat{t}_{ij})^2} \quad \text{and}$$

$$\widehat{\upsilon}^2 = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\widehat{\xi}_{ij}^2 + \mu_x^2 + \tau^2\widehat{t}_{ij}^2 - 2\mu_x\widehat{\xi}_{ij} - 2\tau\widehat{t\xi}_{ij} + 2\tau\mu_x\widehat{t}_{ij}). \tag{2.13}$$

In our case the obtention of the maximum likelihood estimate considering the EM algorithm is straightforward, as we have closed-form expressions for all the parameters. The shape and scale parameters of the latent variable, ξ, can be estimated by noting that $\lambda_x = \tau/\upsilon$ and $\sigma_x^2 = \tau^2 + \upsilon^2$. Starting values are often chosen to be the corresponding estimates under a normal assumption, with $\lambda_x = 3$ (or $\lambda_x = -3$) if the data present positive (negative) skewness, which can be depicted by looking at a data histogram. As recommended in the literature, it is useful to run the EM-algorithm several times with different starting values. Inspection of information criteria such as Akaike Information Criterion (AIC, $-\ell(\widehat{\boldsymbol{\theta}}) + P$), Schwarz'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. Next, we discuss the hypotheses testing of interest and the EM-algorithm for evaluating the restricted MLE in the SN-MEM with especial emphasis in the slope parameters.

2.3 Hypotheses testing and restricted estimation

Let $\mathbf{I} = [\sum_{i=1}^{p}\frac{n_i}{N}E(\mathbf{J}_{ij}(\boldsymbol{\theta}))]$, where $N = \sum_{i=1}^{p} n_i$ and $\mathbf{J}_{ij}(\boldsymbol{\theta})$ as presented in (Appendix B, equation (A.2)). Then under regularity conditions, Bradley and Gart (1962) show that $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(\mathbf{0}, \mathbf{I}^{-1})$, with $\boldsymbol{\theta}_0$ denoting the true parameter vector. However, in complex models as in the case here, the use of the observed information matrix $\widehat{\mathbf{J}} = \frac{1}{N}[\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\mathbf{J}_{ij}(\boldsymbol{\theta}))\,|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}]$ in place of $\mathbf{I}$, with $\hat{\boldsymbol{\theta}}$ denoting the MLE of $\boldsymbol{\theta}$, is preferable (Pawitan, 2001). In the context of dental clinical trial the main interest was to test if the experimental mouth rinses 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 hypotheses: $H_{01}: \beta_{k,1} = \beta_{k,2}$ and $H_{02}: \beta_{k,1} = \beta_{k,3}, k = 1, 2$, since $\beta_{k,i}$ represents the dental plaque reduction rate after three ($k = 1$) and after six ($k = 2$) months from the beginning of the study with the use of the ith mouth rinse (control $i = 1$, A ($i = 2$) or B ($i = 3$)).

Another question of interest was to know if the experimental mouth rinses A and B were long lasting, that is, $H_{03}: \beta_{1,2} = \beta_{2,2}$ and $H_{04}: \beta_{1,3} = \beta_{2,3}$, respectively. Next, we conduct inference regarding the regression coefficients $\boldsymbol{\beta}$.

To test the hypothesis of interest, H_{01}, H_{02}, H_{03} and H_{04}, we may consider the likelihood ratio (LR), score (SR) or Wald (W) test statistics. These 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 $\operatorname{rank}(\mathbf{C}) = q \le mp$ and $\mathbf{d}$ is a $q \times 1$, known vector. Thus, the statistics LR, SR and W (Sen and Singer, 1993; Cox, 2006) can be written as

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

and

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

where $\ell(\boldsymbol{\theta})$ is the log-likelihood function, $\mathbf{U}_\beta$ and $\mathbf{J}^{-1}_{\beta\beta}$ corresponds to the partition of $U(\boldsymbol{\theta})$ and $\mathbf{J}(\boldsymbol{\theta})$ (see Appendix A and Appendix B, respectively) as

$$U(\boldsymbol{\theta}) = (U^\top_{\boldsymbol{\beta}}, U^\top_{\boldsymbol{\theta}-\boldsymbol{\beta}})^\top \quad \text{and} \quad \mathbf{J}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{J}^{-1}_{\boldsymbol{\beta},\boldsymbol{\theta}-\boldsymbol{\beta}} \\ \mathbf{J}^{-1}_{\boldsymbol{\theta}-\boldsymbol{\beta},\boldsymbol{\beta}} & \mathbf{J}^{-1}_{\boldsymbol{\theta}-\boldsymbol{\beta},\boldsymbol{\theta}-\boldsymbol{\beta}} \end{bmatrix},$$

with $\boldsymbol{\theta} - \boldsymbol{\beta} = (\sigma_1^2, \dots, \sigma_p^2, \sigma_\delta^2, \mu_x, \sigma_x^2, \lambda_x)^\top$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_p^\top)^\top$. Under the null hypothesis, the three statistics follow asymptotically a chi-square distribution with q degrees of freedom (χ^2_q).

The EM algorithm for estimating the parameters of the model (2.3), (2.4) and (2.5) under the restriction $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, denoted by $\widetilde{\boldsymbol{\theta}}_c$, follows the same procedures given in (2.12)–(2.13), replacing $\widehat{\boldsymbol{\beta}}$ by $\widetilde{\boldsymbol{\beta}}_c$ in the M-step of the algorithm.

$$\begin{aligned} \widetilde{\boldsymbol{\beta}}_c^{(r+1)} &= (\boldsymbol{\Delta}_x^{(r)})^{-1}\boldsymbol{\delta}_{xy}^{(r)} + (\boldsymbol{\Delta}_x^{(r)})^{-1}\mathbf{C}^\top[\mathbf{C}(\boldsymbol{\Delta}_x^{(r)})^{-1}\mathbf{C}^\top]^{-1}[\mathbf{d} - \mathbf{C}(\boldsymbol{\Delta}_x^{(r)})^{-1}\boldsymbol{\delta}_{xy}^{(r)}] \\ &= \widehat{\boldsymbol{\beta}}^{(r)} + (\boldsymbol{\Delta}_x^{(r)})^{-1}\mathbf{C}^\top[\mathbf{C}\boldsymbol{\Delta}_x^{(r)-1}\mathbf{C}^\top]^{-1}[\mathbf{d} - \mathbf{C}\widehat{\boldsymbol{\beta}}^{(r)}] \end{aligned} \tag{2.14}$$

for $r = 0, 1, \dots,$ where $\widehat{\boldsymbol{\beta}}^{(r)}$, $\boldsymbol{\delta}_{xy}^{(r)}$ and $\boldsymbol{\Delta}_x^{(r)}$ are obtained using the M-step given in the unrestricted case. Note from (2.14) that the problem of testing linear inequality hypotheses of the form $H_0: \mathbf{C}\boldsymbol{\beta} - \mathbf{d} \ge \mathbf{0}$ 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.

3 Local influence

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 dataset. The work of Cook (1986), laid the foundation for assessing local influence of a group of observations when a minor perturbation is made in the statistical model or in the dataset. Based on his proposal many papers have been written on the subject. In his seminal paper, Cook (1986) shows that the normal curvature for $\boldsymbol{\theta} \in \mathbb{R}^{mp+p+4}$ in the direction of $\mathbf{d} \in \mathbb{R}^q$, $\|\mathbf{d}\| = 1$ is given by $C_d = 2|\mathbf{d}^\top \boldsymbol{\Delta}^\top \mathbf{J}^{-1} \boldsymbol{\Delta} \mathbf{d}|$, where $\mathbf{J}$ is the observed information matrix and $\boldsymbol{\Delta}$ is the $(mp + p + 4) \times q$ matrix with elements $\boldsymbol{\Delta}_{rs} = \partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\theta_r\, \partial\omega_s$, both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_o$ (postulated model), with $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ denoting the log-likelihood function of the perturbed model, $r = 1, \ldots, (mp + p + 4)$, $s = 1, \ldots, q$. The suggestion is to pick the direction $d_{\max}$ corresponding to the largest curvature $C_{\max}$. The index plot of $d_{\max}$ may reveal how to perturb the model (or data) to obtain large changes in the estimate of $\boldsymbol{\theta}$. For a more detailed information, we refer the reader to the work of Lachos, Montenegro and Bolfarine (2008) and the references therein.

Another important direction, according to Escobar and Meeker (1992), is $\mathbf{l} = \mathbf{e}_k$, a $N \times 1$ vector of zeros with a one in the kth position. In that case, the normal curvature called the total local influence of subject k, is given by $C_{e_k} = 2|\mathbf{e}_k^\top \mathbf{B} \mathbf{e}_k| = 2|b_{kk}|$, where b_{kk} is the kth diagonal element of $\mathbf{B} = \boldsymbol{\Delta}^\top \mathbf{J}^{-1} \boldsymbol{\Delta}$, $k = 1, \ldots, N$. 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

$$
\begin{aligned}
D_{ij} &= (\widehat{\boldsymbol{\theta}}_{(ij)} - \widehat{\boldsymbol{\theta}})^\top \mathbf{J} (\widehat{\boldsymbol{\theta}}_{(ij)} - \widehat{\boldsymbol{\theta}})/(mp + p + 4), \\
LD_{ij} &= 2[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{(ij)})],
\end{aligned}
$$

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

3.1 Curvature derivation for SN-MEM

In order to obtain the normal curvature we derive the $\boldsymbol{\Delta}$ matrix in closed-form expressions for different perturbation schemes.

Case weight perturbation. The logarithm of the likelihood function is given by (2.7), 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

$$
\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \omega_{ij} \ell_{ij}(\boldsymbol{\theta}), \tag{3.1}
$$

where, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \sigma_\delta^2, \mu_x, \sigma_x^2, \lambda_x)^\top$ 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 and $\boldsymbol{\omega}_0 = \mathbf{1}_N = (1, \ldots, 1)^\top$ (no perturbation vector), with

$N = \sum_{i=1}^{p} n_i$. This perturbation scheme is intended to evaluate whether the contribution of the observations with differing weights affect the maximum likelihood estimate of $\boldsymbol{\theta}$ and it is the most commonly used method to evaluate the influence of a small modification in the model. Thus, using (3.1) it follows after some algebraic manipulation that the delta matrix is given by

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{11}(\boldsymbol{\theta}), \ldots, \boldsymbol{\Delta}_{1n_1}(\boldsymbol{\theta}), \ldots, \boldsymbol{\Delta}_{p1}(\boldsymbol{\theta}), \ldots, \boldsymbol{\Delta}_{pn_p}(\boldsymbol{\theta})), \tag{3.2}$$

where, $\boldsymbol{\Delta}_{ij} = \frac{\partial \ell_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, with individual elements given in Appendix A. The above $\boldsymbol{\Delta}$ matrix is to be evaluated at $\hat{\boldsymbol{\theta}}$.

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

$$y_{k,ij}(\omega_{ij}) = y_{k,ij} + S_k \omega_{ij},$$

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 $\boldsymbol{\omega}_0 = \mathbf{0}$ and the perturbed log-likelihood function can be obtained from (2.7) with $y_{k,ij}$ replaced by $y_{k,ij}(\omega_{ij})$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. Then

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \ell_{ij}(\boldsymbol{\theta}|\omega_{ij}), \tag{3.3}$$

where $\ell_{ij}(\boldsymbol{\theta}|\omega_{ij}) \propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2} g_{ij}(\omega_{ij}) + \log(K_{ij}(\omega_{ij}))$ 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 $K_{ij}(\omega_{ij}) = \Phi_1(A_i a_{ij}(\omega_{ij}))$ 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 (3.2), where

$$\begin{aligned}\boldsymbol{\Delta}_{ij}(\boldsymbol{\theta}) = {} & -\frac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}} + W_{\Phi_1}(A_i a_{ij}(\omega_{ij}))\left[\frac{\partial A_i}{\partial \boldsymbol{\theta}} Q_{ij}(\omega_{ij}) + A_i \frac{\partial Q_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}\right] \\ & + A_i W'_{\Phi_1}(A_i a_{ij}(\omega_j)) Q_{ij}(\omega_{ij}) \left[A_i \frac{\partial a_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}} + a_{ij}(\omega_{ij}) \frac{\partial A_i}{\partial \boldsymbol{\theta}}\right],\end{aligned} \tag{3.4}$$

with $P_{ij}(\omega_{ij}) = (\mathbf{z}_{ij}(\omega_{ij}) - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{d}$, $Q_{ij}(\omega_{ij}) = \mathbf{d}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$, $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, $\mathbf{d} = \frac{\partial \mathbf{z}_{ij}(\omega_{ij})}{\partial \omega_{ij}} = (d_1, \mathbf{d}_2^\top)^\top$ a $(m+1) \times 1$ vector and $\frac{\partial a_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ 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 $\frac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ evaluated at $\boldsymbol{\omega}_o = \mathbf{0}$ is given by

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

where $W_{1ij} = x_{ij} - \mu_x$, $\mathbf{W}_{2ij} = \mathbf{Y}_{ij}(\omega_{ij}) - \boldsymbol{\beta}_i\mu_x$, $d_1 = 0$ and $\mathbf{d}_2 = \mathbf{S}$, with $\mathbf{Y}_{ij}(\omega_{ij}) = (y_{1,ij} + S_1\omega_{ij}, \dots, y_{m,ij} + S_m\omega_{ij})^\top$, $i = 1, \dots, p$, $j = 1, \dots, n_i$ and $\mathbf{S} = (S_1, \dots, S_m)^\top$. Then $Q_{ij}(\omega_{ij}) = \frac{1}{\sigma_i^2}\mathbf{S}^\top\boldsymbol{\beta}_i$. 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.

Explanatory variable 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

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

The perturbed log-likelihood follows from (3.3) 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 (3.2), with $\boldsymbol{\Delta}_{ij}$ as given in (3.4) replacing $\mathbf{z}_{ij}(\omega_{ij}) = (x_{ij}, y_{1,ij} + S_1\omega_{ij}, \dots, y_{m,ij} + S_m\omega_{ij})^\top$ 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 $\mathbf{d}_2 = \mathbf{0_m}$, which leads to

$Q_{ij}(\omega_{ij}) = \frac{1}{\sigma_\delta^2}$. The expressions for $\frac{\partial P_{ij}(\omega_{ij})}{\partial \boldsymbol{\theta}}$ evaluated at $\boldsymbol{\omega}_o = \mathbf{0}$ are the same as given in the response variables perturbation scheme, noting that $W_{1ij} = x_{ij} + \omega_{ij} - \mu_x$, $\mathbf{W}_{2ij} = \mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x$ and $\mathbf{d} = (1, \mathbf{0}_\mathbf{m}^\top)^\top$.

4 Application

In this section, we apply the methodology discussed in this work to a real dataset analyzed in Hadgu and Koch (1999) using generalized estimating equations. The dataset and the objective of the study was respectively described in the Introduction and Section 2.3. In Lachos, Montenegro and Bolfarine (2008) a part of this dataset was analyzed considering the skew-normal distribution and in Aoki et al. (2003), this dataset was analyzed considering the normal model from a Bayesian perspective. To compare the symmetric and the asymmetric model to fit this dataset, the ML estimates of the parameters of the model were obtained for the SN-MEM and N-MEM. The results 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

Table 1 *Results of fitting SN-MEM and N-MEM to the dental plaque index dataset. CI represents the* 95% *confidence interval based on the normal approximation of the ML estimates. PCI is the* 95% *confidence interval based on the profile likelihood. SE represents the estimated asymptotic standard errors based on the observed information matrix given in Appendix B*

	SN-MEM				N-MEM	
Parameter	Estimate	SE	CI	PCI	Estimate	SE
$\beta_{1,1}$	0.7020	0.0339	[0.635; 0.768]	[0.643; 0.769]	0.7021	0.0340
$\beta_{1,2}$	0.5239	0.0441	[0.437; 0.610]	[0.434; 0.615]	0.5241	0.0442
$\beta_{1,3}$	0.5088	0.0317	[0.447; 0.571]	[0.443; 0.574]	0.5087	0.0317
$\beta_{2,1}$	0.6857	0.0339	[0.619; 0.752]	[0.616; 0.755]	0.6859	0.0340
$\beta_{2,2}$	0.5016	0.0441	[0.415; 0.588]	[0.411; 0.589]	0.5017	0.0441
$\beta_{2,3}$	0.4139	0.0317	[0.352; 0.476]	[0.349; 0.479]	0.4139	0.0317
σ_1^2	0.2746	0.0460	[0.184; 0.365]	[0.200; 0.393]	0.2739	0.0461
σ_2^2	0.4306	0.0752	[0.283; 0.578]	[0.309; 0.627]	0.4308	0.0751
σ_3^2	0.2257	0.0377	[0.152; 0.300]	[0.164; 0.324]	0.2253	0.0380
σ_δ^2	0.0010	0.0154	[−0.029; 0.031]	[0.001; 0.012]	0.0021	0.0210
μ_x	2.1082	0.0425	[2.025; 2.192]	[2.010; 2.199]	2.5343	0.0325
σ_x^2	0.2907	0.0550	[0.183; 0.398]	[0.208; 0.462]	0.1086	0.0210
λ_x	6.1291	6.0782	[−5.784; 18.042]	[3.431; 11.517]	–	–
log-likelihood			−194.4457		−203.3778	
AIC			1.9757		2.0512	
BIC			2.1400		2.2029	
HQ			2.0422		2.1127	

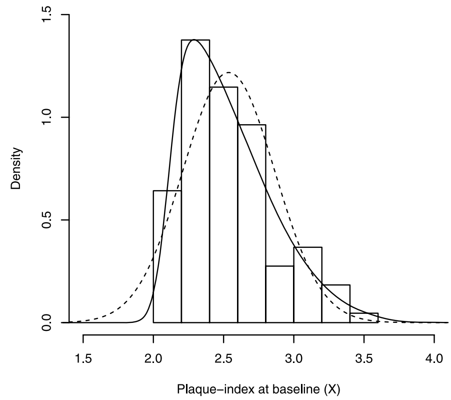

Figure 1 *Dental plaque index dataset. Histogram of the observed covariate x (plaque-index at baseline) superimposed by the estimated densities using skew-normal (solid) and normal distribution (dashed).*

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 seem 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 $H_0 : \lambda_x = 0$ ($LR = 17.8642$, 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. To overcome this problem we have also constructed the profile confidence interval (PCI) as suggested by Meeker and Escobar (1995) and Pawitan (2001), among others. This interval was found to be $(3.45, 11, 55)$, indicating the presence of asymmetry (Table 1 and Figure 2). Alternatively, in Aoki, Pinto and Achcar (2006) it was considered a model where the true value of the covariate follows a mixture of two normal distributions in order to model the asymmetry displayed by the data.

The primary purpose of the experiment was to compare the efficiency of the two experimental mouth rinses, A and B, with the control mouth rinse C, namely, we are interested in comparing the slope parameters $\beta_{k,2}$ and $\beta_{k,3}$ with respect to $\beta_{k,1}$, $k = 1, 2$. Considering the hypothesi of interest $H_{01} : \beta_{1,1} = \beta_{1,2}$ ($H_{01} : \beta_{1,1} = \beta_{1,3}$), the value of the test statistics were given by $LR = 10.7836$, $W = 10.2624$ and $SR = 12.6282$ ($LR = 17.4729$, $W = 17.2691$ and $SR = 21.8584$) which corresponds to a p-values around zero and for the hypothesis $H_{02} : \beta_{2,1} = \beta_{2,2}$ ($H_{02} : \beta_{2,1} = \beta_{2,3}$), the value of the test statistics were given by $LR = 11.5355$, $W = 10.9701$ and $SR = 13.6127$ ($LR = 34.5931$, $W = 34.2282$ and $SR = 51.8797$). Considering

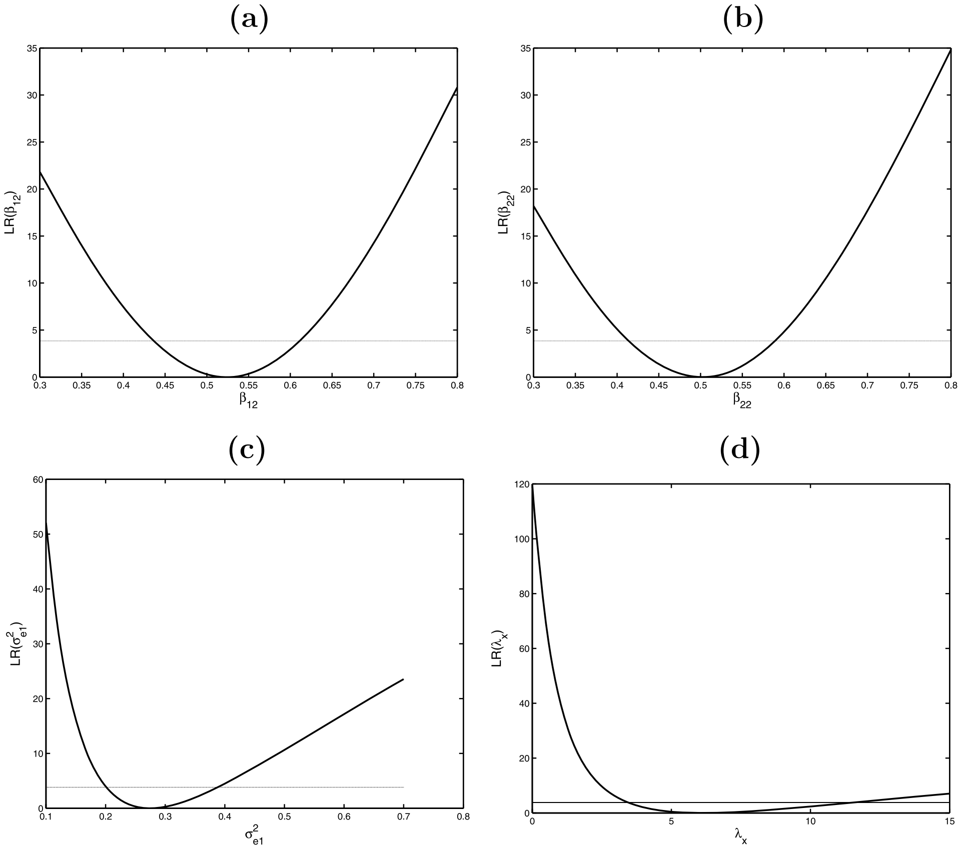

Figure 2 *Dental plaque index dataset. Likelihood ratio (LR) based on the profile likelihood:* (a) $\beta_{1,2}$; (b) $\beta_{2,2}$; (c) σ_1^2 *and* (d) λ_x. *The line in each graphic is the* 95% *limit* ($\chi_1^2(0.95) = 3.84$).

these results, we conclude that both experimental mouth rinses were more efficient then the control mouth rinse in reducing the plaque index after three and after six months. If we consider the hypotheses $H_{03}: \beta_{1,2} = \beta_{2,2}$ and $H_{04}: \beta_{1,3} = \beta_{2,3}$, we obtain $LR = 0.1288$, $W = 0.1287$, $SR = 0.1294$ and $LR = 4.4733$, $W = 4.4730$, $SR = 5.0487$, respectively. Thus, we fail to reject H_{03} and reject H_{04}, which means that the experimental mouth rinse B is long lasting. These conclusions are the same as those obtained in Hadgu and Koch (1999) and Aoki et al. (2003). 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 $LR = 0.2440$, $W = 0.2439$ and $SR = 0.2440$, 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 3 to the Hadgu and Koch dataset. The index plots of $\mathbf{l}_{\max}$ to assess the influence of the perturbation on the ML estimate of the parameter vector $\boldsymbol{\theta}$ are presented in Figures 3–5. 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.

In Figure 3 we present the index plot of $|l_{\max}|$ and $|C_k|$ under the case weight perturbation. Based on this perturbation scheme and plot (a), we find that subjects 7, 19 and 26 of the control month rinse C and subjects 27 and 35 of B are the

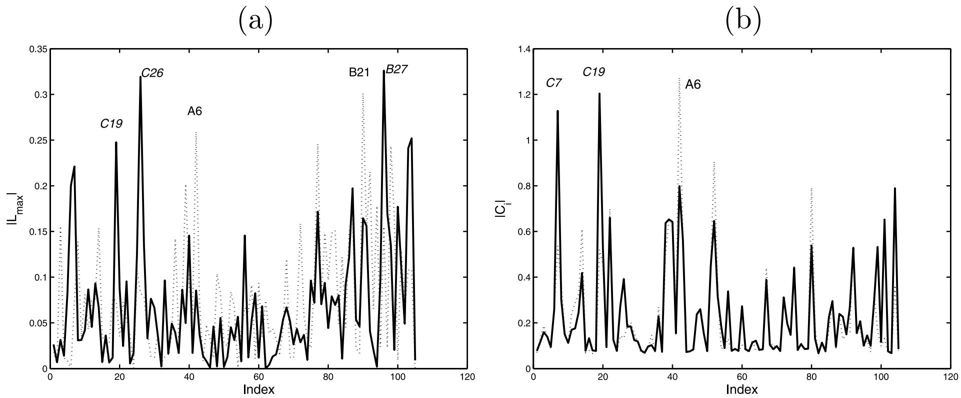

Figure 3 *Dental plaque index dataset. Index plot of* (a) $|l_{\max}|$ *and* (b) C_k *for the case weight perturbation scheme. (—) and (···) denotes the index plot for the SN-MEM and N-MEM, respectively.*

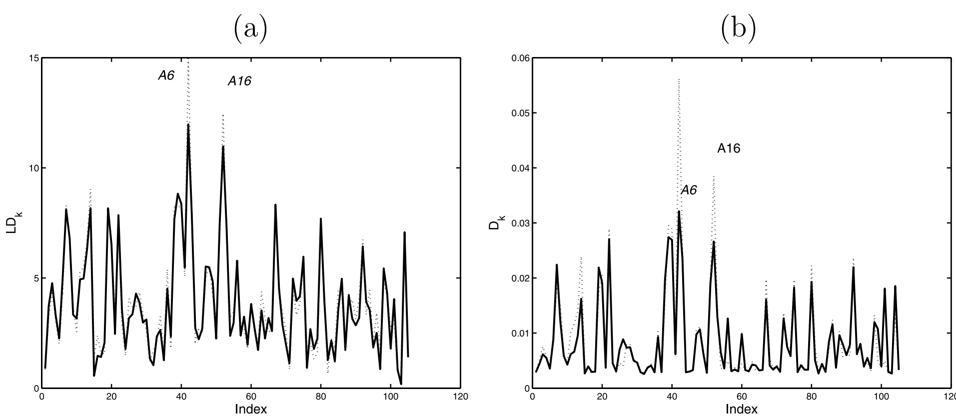

Figure 4 *Dental plaque index dataset. Index plot of* (a) *likelihood displacement* LD_k *and* (b) *Cook's distance* D_k*. (—) and (···) denotes the index plot for the SN-MEM and N-MEM, respectively.*

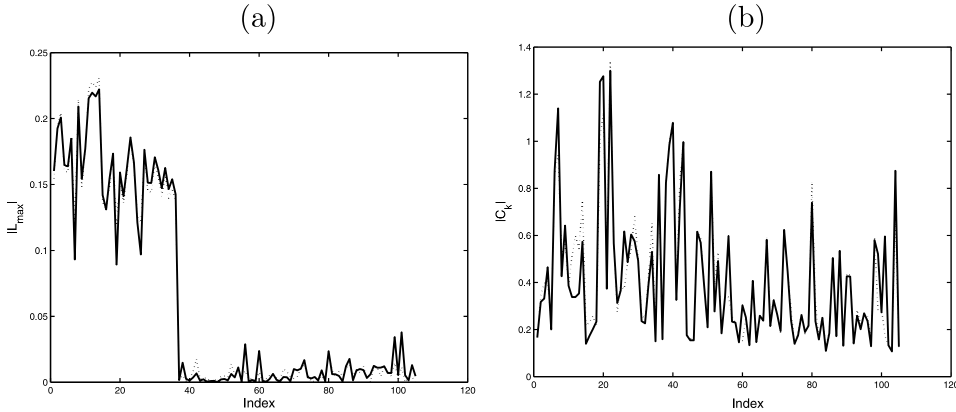

Figure 5 *Dental plaque index dataset. Perturbation of the responses variables. Index plot of* (a) $|l_{\max}|$ *and* (b) $|C_k|$. *(—) and (···) denotes the index plot for the SN-MEM and N-MEM, respectively.*

group of observations that may exert influence 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. Considering the plot (b), we notice that observations 7 and 19 of the control mouth rinse are the most influential ones. In the plot (a) these observations stands out less than the observation 26. However, in the plot (a) we are looking for the group of observations that exert influence in the parameter estimates, while in the plot (b) we are looking for the individuals that exert more influence alone. Under the N-MEM the observations that corresponds to the subjects 6 of A and 21 of the month rinse B are jointly 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 and it is also the one that stands out in the plot (b). Notice that the observations which are influential considering these two models (SN-MEM and N-MEM) are not the same.

In order to compare with the result of local influence, in Figure 4 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 $C_{l_{\max}}(\widehat{\boldsymbol{\theta}}) = 6.2178$ and $C_{l_{\max}}(\widehat{\boldsymbol{\theta}}) = 110.4162$, respectively. Notice in Figure 5 that the observations corresponding to the control mouth rinse C stands out in plot (a), which mean that they are jointly influential, however if we look at the plot (b), we observe that these observations are not individually influent. When the explanatory variable perturbation is considered none observations are jointly influential. Considering the plot of $|C_k|$ we observe that 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 (22 of the control mouth rinse, 6 of the mouth rinse A and 11 of the mouth rinse B) as expected, due to the asymmetric distribution that we have considered.

5 Final conclusions

In this work we have treated the problem of estimation, hypotheses testing and influence diagnostics to the multivariate null intercept measurement error model under the skew-normal distribution. Parameter estimates are obtained via maximum likelihood considering the EM algorithm, yielding closed-form expressions for the equations in the M-step. Hypotheses testing is approached by using likelihood ratio, score and Wald statistics. Also, we derive the appropriate matrices to evaluate the effect of a small perturbation in the model or the data and different perturbation schemes are investigated. We applied the proposed methodology considering the real dataset 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 conclusions of the analysis of the dataset regarding the questions of interest are the same in all of the considered models. In addition, as the dental plaque index is positive an alternative approach may be developed considering the paper of Chen, Gupta and Troskie (2003), where the latent variable is considered to be positive. Finally, we want to mention that this work extends the early results found in Lachos, Montenegro and Bolfarine (2008) and Aoki et al. (2003).

Although the SN-MEM model considered in this article has shown great flexibility in regulating skewness, its robustness against outliers could be seriously affected by thick-tailed observations. Lachos, Ghosh and Arellano-Valle (2010) recently proposed a remedy to accommodate skewness and heavy-tailedness simultaneously using scale mixtures of skew-normal (SMSN) distributions. We conjecture that the methodology presented in this article can be undertaken under a multivariate setting of SMSN distributions and should yield satisfactory results in certain situations, at the expense of additional complexity in its implementation. Nevertheless, a deeper investigation of those modifications is beyond the scope of the present article, but provides interesting topics for further research.

Appendix A: The score function

The score function is given by

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

with

$$U_{ij}(\boldsymbol{\theta}_1) = \frac{\partial \ell_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} = (U_{ij}(\boldsymbol{\beta}_1), \ldots, U_{ij}(\boldsymbol{\beta}_p), U_{ij}(\sigma_1^2), \ldots, U_{ij}(\sigma_p^2))^\top,$$

$$U_{ij}(\boldsymbol{\theta}_2) = \frac{\partial \ell_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} = (U_{ij}(\sigma_\delta^2), U_{ij}(\mu_x), U_{ij}(\sigma_x^2), U_{ij}(\lambda_x))^\top$$

and

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

$\boldsymbol{\gamma} = \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p, \sigma_1^2, \ldots, \sigma_p^2, \sigma_\delta^2, \mu_x, \sigma_x^2, \lambda_x, i = 1, \ldots, p, j = 1, \ldots, n_i$. From (2.6), we have after some algebraic manipulations that the expression of K_{ij} given in (2.7) can be written as $K_{ij} = \Phi_1(A_i a_{ij})$, with $A_i = \frac{\lambda_x \Lambda_i}{\sqrt{\sigma_x^2 + \lambda_x^2 \Lambda_i}}$, $a_{ij} = (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$, $\Lambda_i = \frac{\sigma_x^2}{c_i}$ and $c_i = 1 + \sigma_x^2 \boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}$. Thus, $\frac{\partial \log K_{ij}}{\partial \boldsymbol{\gamma}} = W_{\Phi_1}(A_i a_{ij})[A_i \frac{\partial a_{ij}}{\partial \boldsymbol{\gamma}} + a_{ij}\frac{\partial A_i}{\partial \boldsymbol{\gamma}}]$, with $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$, $u \in \mathbb{R}$. Let $\mathbf{y}_{ij} = (y_{1,ij}, \ldots, y_{m,ij})^\top$, $W_{1ij} = x_{ij} - \mu_x$, $\mathbf{W}_{2ij} = \mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x$, $b_{ij} = (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^\top \mathbf{B}_i(\mathbf{z}_{ij} - \boldsymbol{\mu}_i)$, with $\mathbf{B}_i = \mathrm{D}^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}\boldsymbol{\beta}_{0i}^\top \mathrm{D}^{-1}(\boldsymbol{\phi}_i)$, then

(I) $\frac{\partial \log|\boldsymbol{\Sigma}_i|}{\partial \boldsymbol{\gamma}}$ equals zero for $\boldsymbol{\gamma} = \mu_x$ and $\boldsymbol{\gamma} = \lambda_x$ and

$$\frac{\partial \log|\boldsymbol{\Sigma}_i|}{\partial \boldsymbol{\beta}_i} = \frac{2\sigma_x^2}{\sigma_i^2 c_i}\boldsymbol{\beta}_i, \qquad \frac{\partial \log|\boldsymbol{\Sigma}_i|}{\partial \sigma_i^2} = \frac{1}{\sigma_i^2}\left(p - \frac{\sigma_x^2}{\sigma_i^2 c_i}\boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i\right),$$

$$\frac{\partial \log|\boldsymbol{\Sigma}_i|}{\partial \sigma_\delta^2} = \frac{c_i\sigma_\delta^2 - \sigma_x^2}{c_i \sigma_\delta^4} \quad \text{and} \quad \frac{\partial \log|\boldsymbol{\Sigma}_i|}{\partial \sigma_x^2} = \frac{c_i - 1}{\sigma_x^2 c_i}, \quad i = 1, \ldots, p.$$

(II) $\frac{\partial a_{ij}}{\partial \boldsymbol{\gamma}}$ equals zero for $\boldsymbol{\gamma} = \sigma_x^2$ and $\boldsymbol{\gamma} = \lambda_x$ and

$$\frac{\partial a_{ij}}{\partial \boldsymbol{\beta}_i} = \frac{1}{\sigma_i^2}(\mathbf{W}_{2ij} - \mu_x \boldsymbol{\beta}_i), \qquad \frac{\partial a_{ij}}{\partial \sigma_i^2} = -\frac{1}{\sigma_\delta^4}\boldsymbol{\beta}_i^\top \mathbf{W}_{2ij},$$

$$\frac{\partial a_{ij}}{\partial \sigma_\delta^2} = -\frac{1}{\sigma_\delta^4}W_{1ij} \quad \text{and} \quad \frac{\partial a_{ij}}{\partial \mu_x} = -\frac{c_i - 1}{\sigma_x^2}, \quad i = 1, \ldots, p, j = 1, \ldots, n_i.$$

(III) $\frac{\partial A_i}{\partial \boldsymbol{\gamma}}$ equals zero for $\boldsymbol{\gamma} = \mu_x$ and

$$\frac{\partial A_i}{\partial \boldsymbol{\beta}_i} = -\frac{(2c_i + \lambda_x^2)}{\lambda_x^2 \sigma_i^2}A_i^3 \boldsymbol{\beta}_i, \qquad \frac{\partial A_i}{\partial \sigma_i^2} = \frac{(2c_i + \lambda_x^2)}{2\lambda_x^2 \sigma_i^4}A_i^3 \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i,$$

$$\frac{\partial A_i}{\partial \sigma_\delta^2} = \frac{(2c_i + \lambda_x^2)}{2\lambda_x^2 \sigma_\delta^4}A_i^3, \qquad \frac{\partial A_i}{\partial \sigma_x^2} = \frac{(2c_i + \lambda_x^2 - c_i^2)}{2c_i^2 \lambda_x^2 \Lambda_i^2}A_i^3 \quad \text{and}$$

$$\frac{\partial A_i}{\partial \lambda_x} = \frac{\sigma_x^2}{\Lambda_i^2 \lambda_x^3}A_i^3, \qquad i = 1, \ldots, p.$$

(IV) $\frac{\partial g_{ij}}{\partial \boldsymbol{\gamma}}$ equals zero for $\boldsymbol{\gamma} = \lambda_x$ and

$$\frac{\partial g_{ij}}{\partial \boldsymbol{\beta}_i} = -\frac{2}{\sigma_i^2}\Big[\mu_x(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x) + \frac{\sigma_x^2}{c_i}a_{ij}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x) - \frac{\sigma_x^4}{c_i^2}b_{ij}\boldsymbol{\beta}_i\Big],$$

$$\frac{\partial g_{ij}}{\partial \sigma_i^2} = -\frac{1}{\sigma_i^4}\Big[\frac{\sigma_x^4}{c_i^2}b_{ij}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i + (\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x) - \frac{2\sigma_x^2 a_{ij}}{c_i}\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)\Big],$$

$$\frac{\partial g_{ij}}{\partial \sigma_\delta^2} = -\frac{1}{\sigma_\delta^4}\Big[\frac{\sigma_x^4 b_{ij}}{c_i^2} + \Big(1 - \frac{2\sigma_x^2}{c_i\sigma_\delta^2}\Big)(x_{ij} - \mu_x)^2 - \frac{2\sigma_x^2}{c_i\sigma_i^2}(x_{ij} - \mu_x)[(\mathbf{y}_{ij} - \boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i]\Big],$$

$$\frac{\partial g_{ij}}{\partial \mu_x} = -\frac{2a_{ij}}{c_i} \quad \text{and} \quad \frac{\partial g_{ij}}{\partial \sigma_x^2} = -\frac{b_{ij}}{c_i^2}, \qquad i = 1, \ldots, p,\ j = 1, \ldots, n_i.$$

Appendix B: The observed information matrix

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

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

with $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top, \sigma_1^2, \ldots, \sigma_p^2)^\top$ and $\boldsymbol{\theta}_2 = (\sigma_\delta^2, \mu_x, \sigma_x^2, \lambda_x)$. From (A.1) it follows that the observed, per element, information matrix is given by

$$\mathbf{J}_{ij}(\boldsymbol{\theta}) = -\Big[\frac{\partial^2\ell_{ij}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\tau}^\top}\Big], \tag{A.2}$$

where

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

$$\frac{\partial^2\log K_{ij}}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\tau}^\top} = W_{\Phi_1}(A_i a_{ij})\Big[\frac{\partial A_i}{\partial\boldsymbol{\gamma}}\frac{\partial a_{ij}}{\partial\boldsymbol{\tau}^\top} + A_i\frac{\partial^2 a_{ij}}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\tau}^\top} + \frac{\partial a_{ij}}{\partial\boldsymbol{\gamma}}\frac{\partial A_i}{\partial\boldsymbol{\tau}^\top} + a_{ij}\frac{\partial^2 A_i}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\tau}^\top}\Big]$$

$$+\,\Delta_{\Phi_1}(A_i a_{ij})\left[A_i\frac{\partial a_{ij}}{\partial\boldsymbol{\gamma}}+a_{ij}\frac{\partial A_i}{\partial\boldsymbol{\gamma}}\right]\left[A_i\frac{\partial a_{ij}}{\partial\boldsymbol{\tau}^\top}+a_{ij}\frac{\partial A_i}{\partial\boldsymbol{\tau}^\top}\right],$$

$\Delta_{\Phi_1}(u) = W'_{\Phi_1}(u) = -W_{\Phi_1}(u)(u + W_{\Phi_1}(u))$, $u \in \mathbb{R}$, $\boldsymbol{\gamma},\boldsymbol{\tau} = \boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_p,\sigma_1^2,\ldots,\sigma_p^2,\sigma_\delta^2,\mu_x,\sigma_x^2,\lambda_x$. We also have that $\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}} = \frac{\partial^2 g_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}} = \frac{\partial^2 a_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}} = \frac{\partial^2 A_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}} = 0$ for $\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \boldsymbol{\beta}_l$ or $\boldsymbol{\tau} = \sigma_l^2$, $i \neq l$, $i,l = 1,\ldots,p$, $\boldsymbol{\gamma} = \sigma_i^2$ and $\boldsymbol{\tau} = \sigma_l^2$, $i \neq l$, $i,l = 1,\ldots,p$ and for $\boldsymbol{\gamma} = \mu_x$ and $\boldsymbol{\tau} = \lambda_x$. Also,

(I) $\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\tau}}$ equals zero for $\boldsymbol{\gamma} = \mu_x$ and $\boldsymbol{\tau} = \boldsymbol{\beta}_i$ or $\boldsymbol{\tau} = \sigma_i^2$ or $\boldsymbol{\tau} = \sigma_\delta^2$ or $\boldsymbol{\tau} = \mu_x$ or $\boldsymbol{\tau} = \sigma_x^2$, $\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \sigma_i^2$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \sigma_\delta^2$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \sigma_x^2$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \lambda_x$ and $\boldsymbol{\tau} = \lambda_x$, $i = 1,\ldots,p$,

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\beta}_i\,\partial\boldsymbol{\beta}_i} = \frac{2\sigma_x^2}{\sigma_i^2 c_i}\left(I_m - \frac{2\sigma_x^2}{c_i\sigma_i^2}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top\right),$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\beta}_i\,\partial\sigma_i^2} = -\frac{2\sigma_x^2}{c_i\sigma_i^4}\left(1 - \frac{\sigma_x^2}{c_i\sigma_i^2}[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i]\right)\boldsymbol{\beta}_i,$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\beta}_i\,\partial\sigma_\delta^2} = \frac{2}{\sigma_i^2}\left(\frac{\sigma_x^2}{c_i\sigma_\delta^2}\right)^2\boldsymbol{\beta}_i, \qquad \frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\beta}_i\,\partial\sigma_x^2} = \frac{2}{\sigma_i^2 c_i^2}\boldsymbol{\beta}_i,$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\sigma_i^2\,\partial\sigma_\delta^2} = -\left(\frac{\sigma_x^2}{\sigma_\delta^2\,\sigma_i^2 c_i}\right)^2[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i],$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\sigma_i^2\,\partial\sigma_i^2} = -\frac{1}{\sigma_i^4}\left[p - \frac{\sigma_x^2}{\sigma_i^2 c_i}\left(2 - \frac{\sigma_x^2}{\sigma_i^2 c_i}[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i]\right)[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i]\right],$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\sigma_i^2\,\partial\sigma_x^2} = -\frac{1}{(c_i\sigma_i^2)^2}[\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i], \qquad \frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\sigma_\delta^2\,\partial\sigma_\delta^2} = -\frac{1}{\sigma_\delta^4}\left[1 - \frac{\sigma_x^2}{c_i\sigma_\delta^2}\left(2 - \frac{\sigma_x^2}{c_i\sigma_\delta^2}\right)\right],$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\sigma_\delta^2\,\partial\sigma_x^2} = -\frac{1}{(c_i\sigma_\delta^2)^2},$$

$$\frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\sigma_x^2\,\partial\sigma_x^2} = \frac{1}{\sigma_x^2 c_i}\left(\frac{1}{\sigma_x^2} + \frac{1}{c_i}\boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i)\boldsymbol{\beta}_{0i}\right) - \frac{1}{\sigma_x^4}, \qquad i = 1,\ldots,p.$$

(II) $\frac{\partial^2 g_{ij}}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\tau}}$ equals zero for $\boldsymbol{\gamma} = \boldsymbol{\beta}_i$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \sigma_i^2$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \sigma_\delta^2$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \sigma_x^2$ and $\boldsymbol{\tau} = \lambda_x$, $\boldsymbol{\gamma} = \lambda_x$ and $\boldsymbol{\tau} = \lambda_x$, $i = 1,\ldots,p$.

$$\frac{\partial^2 g_{ij}}{\partial\boldsymbol{\beta}_i\,\partial\boldsymbol{\beta}_i} = -\frac{2\mu_x}{\sigma_i^2}\left(2\sigma_x^2\frac{a_{ij}}{c_i} - 1\right)\mathbf{I}_m$$

$$+\,\frac{2\sigma_x^2}{\sigma_i^4}\left(2\sigma_x^2\frac{a_{ij}}{c_i^2}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)\boldsymbol{\beta}^\top - \frac{1}{c_i}(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)(\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i\mu_x)^\top\right.$$

$$- 4\sigma_x^4 \frac{b_{ij}}{c_i^3} \boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top + 2\sigma_x^2 \frac{a_{ij}}{c_i^2} \boldsymbol{\beta}_i (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x)^\top \bigg),$$

$$\frac{\partial^2 g_{ij}}{\partial \boldsymbol{\beta}_i\, \partial \sigma_i^2} = \frac{2}{\sigma_i^2} \bigg[\mu_x (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x) + \sigma_x^2 \frac{a_{ij}}{c_i} (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x) - \sigma_x^4 \frac{b_{ij}}{c_i^2} \boldsymbol{\beta}_i$$

$$+ \sigma_x^4 \frac{b_{ij}}{2c_i} (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x) \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i - \frac{\sigma_x^2}{c_i} (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x) \boldsymbol{\beta}_i^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x)$$

$$- \sigma_x^6 \frac{b_{ij}}{c_i^3} \boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i + \sigma_x^2 \frac{a_{ij}}{c_i^2} \boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x) \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial \boldsymbol{\beta}_i\, \partial \sigma_\delta^2} = \frac{2}{\sigma_i^2 \sigma_\delta^4} \bigg[\bigg(\frac{\sigma_x^2}{c_i} (x_{ij} - \mu_x) - \frac{\sigma_x^4}{c_i^2} a_{ij} \bigg) (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x)$$

$$+ \bigg(\frac{\sigma_x^6}{c_i} b_{ij} - \frac{\phi_x^2}{c_i^2} a_{ij} (x_{ij} - \mu_x) \bigg) \boldsymbol{\beta}_i \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial \boldsymbol{\beta}_i\, \partial \mu_x} = -\frac{2}{\sigma_i^2} \bigg[\frac{1}{c_i} (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x) - 2\sigma_x^2 \frac{a_{ij}}{c_i^2} \boldsymbol{\beta}_i \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial \boldsymbol{\beta}_i\, \partial \sigma_x^2} = -\frac{2}{\sigma_i^2 c_i^2} \bigg[a_{ij} (\mathbf{y}_{ij} - 2\boldsymbol{\beta}_i \mu_x) - \sigma_x^2 \frac{b_{ij}}{c_i} \boldsymbol{\beta}_i \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial \sigma_i^2\, \partial \sigma_i^2} = \frac{2}{\sigma_i^6} \bigg[\frac{\sigma_x^4}{c_i^2} b_{ij} \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i + (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x)^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x)$$

$$- \frac{2\sigma_x^2 a_{ij}}{c_i} \boldsymbol{\beta}_i^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x) \bigg]$$

$$- \frac{2}{\sigma_i^8} \bigg[\bigg(\sigma_x^6 \frac{b_{ij}}{c_i^3} \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i - \frac{\sigma_x^4}{c_i^2} a_{ij} \boldsymbol{\beta}_i^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x) \bigg) \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i$$

$$- \bigg(\sigma_x^4 \frac{a_{ij}}{c_i^2} \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i - 2\sigma_x^2 (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x)^\top \boldsymbol{\beta}_i \bigg) \boldsymbol{\beta}_i^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x) \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial \sigma_i^2\, \partial \sigma_\delta^2} = -\frac{2}{\sigma_i^4 \sigma_\delta^4} \bigg[\bigg(\frac{\sigma_x^6}{c_i} b_{ij} - \frac{\sigma_x^4}{c_i^2} a_{ij} (x_{ij} - \mu_x) \bigg) \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i$$

$$+ \bigg(\frac{\sigma_x^2}{c_i} (x_{ij} - \mu_x) - \frac{\sigma_x^4}{c_i^2} a_{ij} \bigg) \boldsymbol{\beta}_i (\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x) \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial \sigma_i^2\, \partial \mu_x} = \frac{2}{c_i \sigma_i^2} \bigg[\frac{\sigma_x^2}{c_i} \boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i - 2(\mathbf{y}_{ij} - \boldsymbol{\beta}_i \mu_x)^\top \boldsymbol{\beta}_i \bigg],$$

$$\frac{\partial^2 g_{ij}}{\partial\sigma_i^2\,\partial\sigma_x^2} = -\frac{1}{c_i^2\sigma_i^2}\Big[\sigma_x^2\frac{b_{ij}}{c_i}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i - 2a_{ij}\boldsymbol{\beta}_i^\top(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)\Big],$$

$$\begin{aligned}\frac{\partial^2 g_{ij}}{\partial\sigma_\delta^2\,\partial\sigma_\delta^2} = \frac{2}{\sigma_\delta^6}\Big[&\frac{\sigma_x^4 b_{ij}}{c_i^2} + \Big(1-\frac{2\sigma_x^2}{c_i\sigma_\delta^2}\Big)(x_{ij}-\mu_x)^2 \\ &- \frac{2\sigma_x^2}{c_i\sigma_i^2}(x_{ij}-\mu_x)[(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i]\Big] \\ &- \frac{2\sigma_x^2}{\sigma_\delta^8}\Big[\sigma_x^4\frac{b_{ij}}{c_i} - \sigma_x^2\frac{a_{ij}}{c_i^2}(x_{ij}-\mu_x) - 2(x_{ij}-\mu_x)^2\Big(\frac{\sigma_x^2}{c_i^2\sigma_\delta^2}-\frac{1}{c_1}\Big) \\ &- \frac{\sigma_x^2}{c_i^2}(x_{ij}-\mu_x)(\mathbf{y}_{ij}-\boldsymbol{\beta}_i\mu_x)^\top\boldsymbol{\beta}_i\Big],\end{aligned}$$

$$\frac{\partial^2 g_{ij}}{\partial\sigma_\delta^2\,\partial\mu_x} = -\frac{2}{\sigma_\delta^4}\Big[\sigma_x^2\frac{a_{ij}}{c_i^2} - \frac{(x_{ij}-\mu_x)}{c_i}\Big],$$

$$\frac{\partial^2 g_{ij}}{\partial\sigma_\delta^2\,\partial\sigma_x^2} = -\frac{2}{\sigma_\delta^4}\Big[\sigma_x^2\frac{b_{ij}}{c_i^3} + 2a_{ij}\frac{(x_{ij}-\mu_x)}{c_i^2}\Big],$$

$$\frac{\partial^2 g_{ij}}{\partial\mu_x\,\partial\mu_x} = \frac{2}{c_i}\Big(\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_i^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i\Big), \qquad \frac{\partial^2\log|\boldsymbol{\Sigma}_i|}{\partial\mu_x\,\partial\sigma_x^2} = \frac{2a_{ij}}{c_i^2}\Big(\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_i^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i\Big),$$

$$\frac{\partial^2 g_{ij}}{\partial\sigma_x^2\,\partial\sigma_x^2} = -2\frac{b_{ij}}{c_i^3}\Big(\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_i^2}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i\Big).$$

(III) $\frac{\partial^2 a_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}}$ equals zero for $\boldsymbol{\gamma}=\mu_x$ and $\boldsymbol{\tau}=\mu_x$ or $\boldsymbol{\tau}=\sigma_x^2$, $\boldsymbol{\gamma}=\boldsymbol{\beta}_i$ and $\boldsymbol{\tau}=\sigma_\delta^2$, $\boldsymbol{\gamma}=\boldsymbol{\beta}_i$ and $\boldsymbol{\tau}=\sigma_x^2$, $\boldsymbol{\gamma}=\boldsymbol{\beta}_i$ and $\boldsymbol{\tau}=\lambda_x$, $\boldsymbol{\gamma}=\sigma_i^2$ and $\boldsymbol{\tau}=\sigma_\delta^2$, $\boldsymbol{\gamma}=\sigma_i^2$ and $\boldsymbol{\tau}=\sigma_x^2$, $\boldsymbol{\gamma}=\sigma_i^2$ and $\boldsymbol{\tau}=\lambda_x$, $\boldsymbol{\gamma}=\sigma_\delta^2$ and $\boldsymbol{\tau}=\sigma_x^2$, $\boldsymbol{\gamma}=\sigma_\delta^2$ and $\boldsymbol{\tau}=\lambda_x$, $\boldsymbol{\gamma}=\sigma_x^2$ and $\boldsymbol{\tau}=\sigma_x^2$, $\boldsymbol{\gamma}=\sigma_x^2$ and $\boldsymbol{\tau}=\lambda_x$, $\boldsymbol{\gamma}=\lambda_x$ and $\boldsymbol{\tau}=\lambda_x$, $i=1,\ldots,p$,

$$\frac{\partial^2 a_{ij}}{\partial\boldsymbol{\beta}_i\,\partial\boldsymbol{\beta}_i} = -2\frac{\mu_x}{\sigma_i^2}\mathbf{I}_m, \qquad \frac{\partial^2 a_{ij}}{\partial\boldsymbol{\beta}_i\,\partial\sigma_i^2} = \frac{1}{\sigma_i^4}(\mathbf{W}_{2ij}-\mu_x\boldsymbol{\beta}_i),$$

$$\frac{\partial^2 a_{ij}}{\partial\boldsymbol{\beta}_i\,\partial\mu_x} = -\frac{2}{\sigma_i^2}\boldsymbol{\beta}_i^\top, \qquad \frac{\partial^2 a_{ij}}{\partial\sigma_i^2\,\partial\sigma_i^2} = \frac{2}{\sigma_i^6}\mathbf{W}_{2ij}^\top\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 a_{ij}}{\partial\sigma_i^2\,\partial\mu_x} = \frac{1}{\sigma_i^4}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i, \qquad \frac{\partial^2 a_{ij}}{\partial\sigma_\delta^2\,\partial\sigma_\delta^2} = \frac{2}{\sigma_\delta^6}W_{1ij}, \qquad \frac{\partial^2 a_{ij}}{\partial\sigma_\delta^2\,\partial\mu_x} = \frac{1}{\sigma_\delta^4}.$$

(IV) $\frac{\partial^2 A_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}}$ equals zero for $\boldsymbol{\gamma}=\mu_x$ and $\boldsymbol{\tau}=\boldsymbol{\beta}_i$ or $\boldsymbol{\tau}=\sigma_i^2$ or $\boldsymbol{\tau}=\sigma_\delta^2$ or $\boldsymbol{\tau}=\mu_x$ or $\boldsymbol{\tau}=\sigma_x^2$,

$$\frac{\partial^2 A_i}{\partial\boldsymbol{\beta}_i\,\partial\boldsymbol{\beta}_i}=-\Big(4\frac{\sigma_x^2}{\lambda_x^2}A_i^3-\frac{3(2c_i+\lambda_x^2)^2}{\lambda_x^4}A_i^5\Big)\frac{1}{\sigma_i^4}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top-\frac{2c_i+\lambda_x^2}{\lambda_x^2\sigma_i^2}A_i^3\mathbf{I}_m,$$

$$\frac{\partial^2 A_i}{\partial\boldsymbol{\beta}_i\,\partial\sigma_i^2}=\Big[2\frac{\sigma_x^2}{\lambda_x^2}A_i^3-\frac{3(2c_i+\lambda_x^2)^2}{2\lambda_x^4}A_i^5\Big]\frac{1}{\sigma_i^6}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i\boldsymbol{\beta}_i+\frac{2c_i+\lambda_x^2}{\lambda_x^2}A_i^3\frac{1}{\sigma_i^4}\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\boldsymbol{\beta}_i\,\partial\sigma_\delta^2}=\Big[2\frac{\sigma_x^2}{\lambda_x^2}A_i^3-\frac{3(2c_i+\lambda_x^2)^2}{2\lambda_x^4}A_i^5\Big]\frac{1}{\sigma_\delta^4\sigma_i^2}\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\boldsymbol{\beta}_i\,\partial\lambda_x}=\frac{\sigma_x^2A_i^3}{\lambda_x^5\Lambda_i^2}(-3A_i^2(2c_i+\lambda_x^2)+4\lambda_x^2\Lambda_i)\frac{1}{\sigma_i^2}\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\boldsymbol{\beta}_i\,\partial\sigma_x^2}=-\Big[\frac{2(c_i-1)}{\lambda_x^2\sigma_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)(2c_i+\lambda_x^2-c_i^2)}{2\lambda_x^4\phi_x^2}A_i^5\Big]\frac{1}{\sigma_i^2}\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\sigma_i^2\,\partial\sigma_i^2}=\Big[-\frac{\sigma_x^2}{\lambda_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)^2}{4\lambda_x^4}A_i^5\Big]\frac{1}{\sigma_i^8}(\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i)^2-\frac{2c_i+\lambda_x^2}{\lambda_x^2}A_i^3\frac{1}{\sigma_i^6}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\sigma_i^2\,\partial\sigma_\delta^2}=\Big[-\frac{\sigma_x^2}{\lambda_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)^2}{4\lambda_x^4}A_i^5\Big]\frac{1}{\sigma_\delta^2\sigma_i^2}\mathbf{1}_m^\top\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\sigma_i^2\,\partial\lambda_x}=\frac{\sigma_x^2A_i^3}{2\lambda_x^5\Lambda_i^2}[3A_i^2(2c_i+\lambda_x^2)-4\lambda_x^2\Lambda_i]\frac{1}{\sigma_i^4}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\sigma_i^2\,\partial\sigma_x^2}=\Big[\frac{(c_i-1)}{\lambda_x^2\sigma_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)(2c_i+\lambda_x^2-c_i^2)}{4\lambda_x^4\sigma_x^4}A_i^5\Big]\frac{1}{\sigma_i^4}\boldsymbol{\beta}_i^\top\boldsymbol{\beta}_i,$$

$$\frac{\partial^2 A_i}{\partial\sigma_\delta^2\,\partial\sigma_\delta^2}=\Big[-\frac{\sigma_x^2}{\lambda_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)^2}{4\lambda_x^4}A_i^5\Big]\frac{1}{\sigma_\delta^8}-\frac{2c_i+\lambda_x^2}{\lambda_x^2}A_i^3\frac{1}{\sigma_\delta^6},$$

$$\frac{\partial^2 A_i}{\partial\sigma_\delta^2\,\partial\sigma_x^2}=\Big[\frac{(c_i-1)}{\lambda_x^2\sigma_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2)(2c_i+\lambda_x^2-c_i^2)}{4\lambda_x^4\sigma_x^4}A_i^5\Big]\frac{1}{\sigma_\delta^4},$$

$$\frac{\partial^2 A_i}{\partial\sigma_\delta^2\,\partial\lambda_x}=\frac{\sigma_x^2A_i^3}{2\lambda_x^5\Lambda_i^2}[3A_i^2(2c_i+\lambda_x^2)-4\lambda_x^2\Lambda_i]\frac{1}{\sigma_\delta^4},$$

$$\frac{\partial^2 A_i}{\partial\sigma_x^2\,\partial\sigma_x^2}=-\frac{\lambda_x^2+1}{\lambda_x^2\sigma_x^6}A_i^3+\frac{3(2c_i+\lambda_x^2-c_i^2)^2}{4\lambda_x^4\sigma_x^8}A_i^5,$$

$$\frac{\partial^2 A_i}{\partial\sigma_x^2\,\partial\lambda_x}=\frac{c_i-2}{\lambda_x^3\Lambda_i\sigma_x^2}A_i^3+\frac{3(2c_i+\lambda_x^2-c_i^2)}{2\lambda_x^5\Lambda_i^2\sigma_x^2}A_i^5,$$

$$\frac{\partial^2 A_i}{\partial\lambda_x\,\partial\lambda_x} = -\frac{3\sigma_x^2}{\lambda_x^4\Lambda_i^2}A_i^3 + \frac{3\sigma_x^4}{\lambda_x^6\Lambda_i^4}A_i^5,$$

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 $\mathbf{W}_{2ij}$ as given in Appendix A, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$.

Acknowledgments

We would like to thank two anonymous referees for the commentaries and suggestions that improved the quality of the paper. The authors acknowledges the partial financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and CNPq—Brazil.

References

Aoki, R., Bolfarine, H. and Singer, J. M. (2001). Null intercept measurement error regression models. *Test* **10**, 441–454. MR1881150

Aoki, R., Bolfarine, H., Achcar, J. A. and Pinto, D. L. (2003). Bayesian analysis of a multivariate null intercept error-in-variables regression model. *Journal of Biopharmaceutical Statistics* **13**, 767–775.

Aoki, R., Pinto, J. D. L. and Achcar, J. A. (2006). Mixture of normal distributions in multivariate null intercept measurement error model. *Journal of Biopharmaceutical Statistics* **16**, 785–802. MR2297405

Aoki, R., Singer, J. M. and Bolfarine, H. (2007). Local influence for measurement error regression models for the analysis of pretest/posttest data. *Journal of Applied Statistics* **15**, 317–330. MR2395937

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. MR2202402

Arellano-Valle, R. B., Ozan, S., Bolfarine, H. and Lachos, V. H. (2005). Skew normal measurement error models. *Journal of Multivariate Analysis* **98**, 265–281. MR2204978

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society* **61**, 579–602. MR1707862

Azzalini, A. and Capitanio, A. (2003). Distributions generated and perturbation of symmetry with emphasis on the multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B* **61**, 367–389.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726. MR1440039

Bradley, R. A. and Gart, J. J. (1962). The asymptotic properties of ml estimators when sampling from associated populations. *Biometrika* **49**, 205–214. MR0137259

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113. MR1867257

Chan, L. K. and Mak, T. K. (1979). On the maximun likelihood estimation of a linear structural relationship when the intercept is known. *Journal of Multivariate Analyis* **9**, 304–313. MR0538410

Chen, J. T., Gupta, A. K. and Troskie, C. G. (2003). The distribution of stock returns when the market is up. *Communication in Statistics—Theory and Methods* **32**, 1541–1558. MR1996794

Cheng, C. L. and Van-Ness, J. W. (1999). *Statistical Regression with Measurement Error*. London: Arnold. MR1719513

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics* **19**, 15–18. MR0436478

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B* **48**, 133–169. MR0867994

Cox, D. R. (2006). *Principles of Statistical Inference*. New York: Cambridge Univ. Press. MR2278763

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, Series B* **39**, 1–22. MR0501537

Escobar, E. and Meeker, W. (1992). Assessing influence in regression analysis with censured data. *Biometrics* **48**, 507–528. MR1173494

Fahrmeir, L. and Klinger, J. (1994). Estimating and testing generalized linear models under inequality restrictions. *Statistical Papers* **35**, 211–229.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications. A Journey Beyond Normality*. New York: Chapman & Hall/CRC. MR2156754

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

Lachos, V. H., Ghosh, P. and Arellano-Valle, R. B. (2010). Likelihood based inference for skew–normal/independent linear mixed models. *Statistica Sinica* **20**, 303–322.

Lachos, V. H., Montenegro, L. C. and Bolfarine, H. (2008). Inference and assessment of local influence in skew-normal null intercept measurement error model. *Journal of Statistical Computation and Simulation* **78**, 395–419. MR2432253

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* **136**, 373–389. MR2211345

Loperfido, N. (2010). Canonical transformation of skew-normal variates. *Test* **10**, 146–165.

Meeker, W. Q. and Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician* **49**, 48–53.

Patefield, W. M. (1985). Information from the maximized likelihood function. *Biometrics* **72**, 664–668. MR0817581

Pawitan, Y. (2001). *In All Likelihood: Statistical Modellling and Inference Using Likelihood*. Oxford: Clarendon Press.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics* **27**, 859–870.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew-t distributions. *Journal of Statistical Planning and Inference* **136**, 4259–4275. MR2323415

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. New York: Chapman & Hall. MR1293125

Turesky, S., Gilmore, N. D. and Glickman, J. (1970). Reduced plaque formation by the chloromethyl analogue of victamine c. *Journal of Periodontology* **41**, 41–43.

F. V. Labra
V. H. Lachos
Departamento de Estatística
IMECC/UNICAMP
Caixa Postal 6065, 13083-970
Campinas, SP
Brazil

R. Aoki
V. Garibay
Departamento de Matemática Aplicada e Estatística
ICMC/USP-São Carlos
Caixa Postal 668, 13560-970
São Carlos, SP
Brazil