Assessment of Local Influence in Elliptical Linear Models with Longitudinal Structure

Felipe Osorio [a], Gilberto A. Paula [a,1,*], Manuel Galea [b]

[a] *Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil*

[b] *Departamento de Estadística, Universidad de Valparaíso, Chile*

Abstract

The aim of this paper is to derive local influence curvatures under various perturbation schemes for elliptical linear models with longitudinal structure. The elliptical class provides a useful generalization of the normal model since it covers both light- and heavy-tailed distributions for the errors, such as Student-t, power exponential, contaminated normal, among others. It is well known that elliptical models with longer-than-normal tails may present robust parameter estimates against outlying observations. However, little has been investigated on the robustness aspects of the parameter estimates against perturbation schemes. We use appropriate derivative operators to express the normal curvatures in tractable forms for any correlation structure. Estimation procedures for the position and variance-covariance parameters are also presented. A data set previously analyzed under a normal linear mixed model is reanalyzed under elliptical models. Local influence graphics are used to select less sensitive models with respect to some perturbation schemes.

Key words: Correlated data; Likelihood displacement; Matrix differential; Outliers; Regression diagnostic; Robust estimation.

1 Introduction

We discuss in this paper the assessment of local influence in elliptical linear models for analyzing longitudinal data. This class is based on the models proposed by Lange, Little and Taylor (1989) and extends the repeated-measure

* Corresponding author.

Email addresses: osorio@ime.usp.br (Felipe Osorio), giapaula@ime.usp.br (Gilberto A. Paula), manuel.galea@uv.cl (Manuel Galea).

1 Address for correspondence: Instituto de Matemática e Estatística, USP - Caixa Postal 66281 (Ag. Cidade de São Paulo), 05311-970 São Paulo - SP - Brazil

normal linear models in the sense of covering both light- and heavy-tailed symmetric distributions for the errors, including Student-t, power exponential, contaminated normal, among others. Thus, data sets containing a larger number of outliers than can be expected under normal error may be better accommodated under elliptical models with longer-than-normal tails and since the inferences for elliptical models are similar to those for normal models, robust parameter estimates against outlying observations may be easily obtained. In addition, the methodology of local influence may be helpful for studying the robustness aspects of the maximum likelihood estimates against perturbation schemes in the model (or data).

Linear and nonlinear multivariate elliptical models have been investigated by various authors. For example, under t error, Lange, Little and Taylor (1989) present an approach for modeling multivariate t-distributions with known and unknown degrees of freedom, Welsch and Richardson (1997) describe multivariate t linear mixed models using the marginal distribution for the response, Kowalski et al. (1999) compare the classical and Bayesian inferences for multivariate t linear models, Fernández and Steel (1999) reveal some pitfalls of both Bayesian and maximum likelihood methods in multivariate t linear models with unknown degrees of freedom, Pinheiro, Liu and Wu (2001) propose a robust hierarchical linear mixed model in which the random effects and the within-subject errors have a multivariate t-distribution and Cysneiros and Paula (2004) discuss restricted methods in t linear models with longitudinal structure. Under other elliptical errors, for instance, Huggins (1993) constructs M-estimators based in symmetric multivariate distributions with applications in pedigree data, Lindsey (1999) discusses the application of the power exponential model in repeated-measurement problems, while Galea, Paula and Bolfarine (1997), Liu (2000, 2002) and Díaz-García, Galea and Leiva-Sánchez (2003) derive influence diagnostic graphics in multivariate elliptical linear models. More recently, Savalli, Paula and Cysneiros (2006) discuss the assessment of variance components in elliptical linear mixed models. A detailed description of the elliptical multivariate class is given, for instance, in Fang, Kotz and Ng (1990).

Local influence (Cook, 1986) has become a very popular tool to assess model assumptions. The approach consists in studying the effects of small perturbations in the model (or data) on same influence measure. The methodology has been largely applied in linear and nonlinear regression models. In particular, in linear mixed models, for instance, Beckman, Nachtsheim and Cook (1987) have applied the approach to detect influential observations in a normal linear mixed model with emphasis to study influence of single observations, while Lesaffre and Verbeke (1998) extend the local influence methodology to normal linear mixed models in repeated-measurement context and under the case-weight perturbation scheme. The local influence approach has also been applied in non-normal mixed models, such as generalized linear models (see, for instance, Ouwens, Tan and Berger, 2001 and Zhu and Lee, 2003). Our focus will be on studying this methodology in elliptical linear models with longitudinal structure, which includes as a particular case the mixed-effect case. However, the results are more general including various possible correlation structures for the errors as well as different perturbation schemes. Robustness aspects of the maximum likelihood estimates against some perturbation schemes are investigated for a particular data set fitted under normal and elliptical linear mixed models with heavier-tailed error distributions.

The paper is organized as follows. In Section 2 the elliptical linear class with longitudinal structure is defined, an iterative process for the parameter estimation and some inferential results are also given. Section 3 deals with some basic calculations related with local influence. Derivations of the normal curvature under different perturbation schemes are made in Section 4 in which appropriate derivative operators are used. A data set previously analyzed under normal error is reanalyzed in Section 5 under elliptical errors with heavy-tailed distributions. The last section deals with some conclusions.

2 Elliptical Linear Model

The class of elliptical distributions has received special attention in the last years, particularly due the fact of including distributions such Student-t, power exponential, contaminated normal, among others, with heavier or lighter tails than the normal ones. We say that an m-dimensional vector $\boldsymbol{Y}$ has a multivariate elliptical distribution (see, for instance, Fang, Kotz and Ng, 1990 and Arellano, 1994) with location parameter $\boldsymbol{\mu} \in \mathbb{R}^m$ and a positive definite scale matrix $\boldsymbol{\Sigma}$ if its density function assumes the form

$$f_Y(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-1/2} g[(\boldsymbol{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})], \tag{1}$$

where $g : \mathbb{R} \to [0, \infty)$ such that $\int_0^\infty u^{m/2-1} g(u)\,\mathrm{d}u < \infty$. Typically $g(\cdot)$ is known as the density generator. We will use the notation $\boldsymbol{Y} \sim EC_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. A description of multivariate elliptical distributions may be found, for instance, in Galea, Paula and Bolfarine (1997).

Consider $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$, $i = 1, \ldots, n$, independent random vectors such that $\boldsymbol{Y}_i \sim EC_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, g)$, where $\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}$, with $\boldsymbol{X}_i$ being an $m_i \times p$ model matrix for the ith individual and $\boldsymbol{\beta}$ a p-dimensional unknown parameter. In addition we will assume that $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$ can be structured by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^T$, as described, for instance, in Jennrich and Schluchter (1986). However, it is important to observe that $\boldsymbol{\Sigma}_i$ is proportional to the variance-covariance matrix of $\boldsymbol{Y}_i$ by a quantity η_i that may be obtained from the derivative of the characteristic function (see, for instance, Fang, Kotz and

Ng, 1990). In particular, for the normal and Student-t model with ν degrees of freedom one has $\eta_i = 1$ and $\eta_i = \nu/(\nu-2)$, $(\nu > 2)$, respectively. Thus, the estimates of $\alpha_1, \ldots, \alpha_k$ are not comparable among different elliptical errors.

The log-likelihood function for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ is given by $L(\boldsymbol{\theta}) = \sum_{i=1}^n L_i(\boldsymbol{\theta})$, where

$$L_i(\boldsymbol{\theta}) = -\tfrac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log g(u_i) \tag{2}$$

and $u_i = (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$ is the Mahanalobis distance, $i = 1, \ldots, n$. Assuming $g(\cdot)$ continuous and differentiable we can define the quantities

$$W_g(u) = \frac{\mathrm{d}}{\mathrm{d}u}\log g(u) = \frac{g'(u)}{g(u)} \quad \text{and} \quad W_g'(u) = \frac{\mathrm{d}}{\mathrm{d}u}W_g(u). \tag{3}$$

Examples of $W_g(u)$ and $W_g'(u)$ for some multivariate elliptical distributions are:

- *Student-t with $\nu > 0$ degrees of freedom, $t_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ (Lange, Little and Taylor, 1989).* The use of the t-distribution as an alternative to the normal distribution, has frequently been suggested in the literature, for example, Little (1988) and Lange, Little and Taylor (1989) use the Student-t distribution for robust modeling. In this case, we have

$$W_g(u) = -\frac{1}{2}\left(\frac{\nu+m}{\nu+u}\right) \quad \text{and} \quad W_g'(u) = \frac{1}{2}\frac{\nu+m}{(\nu+u)^2}.$$

- *Power exponential, $PE_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$, with the shape parameter $\lambda > 0$ (Gómez, Gómez-Villegas and Marín, 1998).* This family presents both light- $(\lambda > 1)$ and heavy-tailed $(\lambda < 1)$ distributions and includes the normal case $(\lambda = 1)$. Applications of this distribution for analyzing repeated-measure data may be found, for instance, in Lindsey (1999). Thus, for $u \neq 0$ and $\lambda \neq \frac{1}{2}$ we have

$$W_g(u) = -\tfrac{1}{2}\lambda u^{\lambda-1} \quad \text{and} \quad W_g'(u) = -\tfrac{1}{2}\lambda(\lambda-1)u^{\lambda-2}.$$

- *Contaminated normal, $CN_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta, \gamma)$, $0 < \delta < 1$ and $0 \le \gamma < 1$ (Little, 1988).* This distribution may also be applied for modeling symmetric data with outlying observations. The parameter δ represents the percentage of outliers while γ may be interpreted as a scale factor. Applications as well as discussions on this distribution are given, for example, in Little (1988) and Lange, Little and Taylor (1989). We obtain

$$W_g(u) = -\frac{1}{2}\frac{1-\delta+\delta\gamma^{m/2+1}e^{(1-\gamma)u/2}}{1-\delta+\delta\gamma^{m/2}e^{(1-\gamma)u/2}} \qquad \text{and}$$

$$W_g'(u) = -\frac{1}{2}\frac{\delta\gamma^{m/2}(1-\gamma)\{W_g(u)+\gamma/2\}e^{(1-\gamma)u/2}}{1-\delta+\delta\gamma^{m/2}e^{(1-\gamma)u/2}}.$$

The score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are, respectively, given by

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} v_i \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \qquad \text{and} \tag{4}$$

$$\boldsymbol{U}(\boldsymbol{\alpha}) = (U(\alpha_1), \ldots, U(\alpha_k))^T$$

with

$$U(\alpha_j) = -\tfrac{1}{2}\sum_{i=1}^{n}\{\operatorname{tr} \boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(j) - v_i \boldsymbol{r}_i^T \boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(j)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i\}, \tag{5}$$

for $j = 1, \ldots, k$, $\dot{\boldsymbol{\Sigma}}_i(j) = \partial \boldsymbol{\Sigma}_i / \partial \alpha_j$, $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$ and $v_i = v_i(\boldsymbol{\theta}) = -2W_g(u_i)$, $i = 1, \ldots, n$.

A joint iterative procedure to obtain the maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{\beta}^{(r+1)} = \Big(\sum_{i=1}^{n} v_i(\boldsymbol{\theta}^{(r)})\, \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-(r)} \boldsymbol{X}_i\Big)^{-1} \sum_{i=1}^{n} v_i(\boldsymbol{\theta}^{(r)})\, \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-(r)} \boldsymbol{Y}_i \quad \text{and} \tag{6}$$

$$\boldsymbol{\alpha}^{(r+1)} = \arg\max_{\alpha}\{L(\boldsymbol{\beta}^{(r+1)}, \boldsymbol{\alpha})\} \tag{7}$$

for $r = 0, 1, \ldots$. To perform the maximization in (7) we consider a multivariate secant method (see, for instance, Dennis and Schnabel, 1996) where the score functions are given in (5). Starting values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\alpha}^{(0)}$ are required for (6)-(7). The quantity $v_i(\boldsymbol{\theta}) = -2W_g(u_i)$ that appears in the equations (4)-(5) may be interpreted as a weight and since $g(u_i)$ is in general a positive decreasing function one has that $v_i(\boldsymbol{\theta}) > 0$ for the majority of the elliptical models. In particular, for the Student-t and power exponential ($\lambda < 1$) distributions $v_i(\boldsymbol{\theta})$ is inversely proportional to the Mahalanobis distance. Thus, the larger u_i is smaller $v_i(\boldsymbol{\theta})$ is, and the estimation procedure (6)-(7) tends to give smaller weight to outlying observations in the sense of the Mahalanobis distance.

The Fisher information matrix for $\boldsymbol{\theta}$ assumes the following block diagonal form:

$$\boldsymbol{K}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{K}(\boldsymbol{\beta}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}(\boldsymbol{\alpha}) \end{pmatrix}, \tag{8}$$

where

$$\boldsymbol{K}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{4d_{gi}}{m_i} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i \quad \text{and} \quad \boldsymbol{K}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \boldsymbol{K}_i(\boldsymbol{\alpha}).$$

The (r, s)th element of $\boldsymbol{K}_i(\boldsymbol{\alpha})$ is given by

$$K_{i,rs}(\boldsymbol{\alpha}) = \frac{b_{rsi}}{4}\Big(\frac{4f_{gi}}{m_i(m_i+2)} - 1\Big) + \frac{2f_{gi}}{m_i(m_i+2)} \operatorname{tr}\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(s)\},$$

where $d_{gi} = E\{W_g^2(U_i)U_i\}$, $f_{gi} = E\{W_g^2(U_i)U_i^2\}$ with $U_i = ||\boldsymbol{Z}_i||^2$, $\boldsymbol{Z}_i \sim EC_{m_i}(\boldsymbol{0}, \boldsymbol{I}_{m_i}, g)$ and $b_{rsi} = \operatorname{tr}\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\} \operatorname{tr}\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(s)\}$. It is possible to obtain closed-form expressions for d_{gi} and f_{gi} for some multivariate elliptical distributions. In particular, for the Student-t distribution with ν degrees of freedom one has $d_{gi} = \frac{m_i}{4}\left(\frac{\nu+m_i}{\nu+m_i+2}\right)$ and $f_{gi} = \frac{m_i(m_i+2)}{4}\left(\frac{\nu+m_i}{\nu+m_i+2}\right)$, and for the power exponential distribution with shape parameter λ we find $d_{gi} = \frac{\lambda^2}{2^{1/\lambda}}\Gamma\left(\frac{m_i+2}{2\lambda}+2\right)/\Gamma\left(\frac{m_i}{2\lambda}\right)$ and $f_{gi} = \frac{m_i(m_i+2\lambda)}{4}$, where $\Gamma(\cdot)$ denotes the gamma function. In this work the asymptotic variance-covariance matrices for the maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$ are estimated, respectively, through $\boldsymbol{K}^{-1}(\boldsymbol{\beta})$ and $\boldsymbol{K}^{-1}(\boldsymbol{\alpha})$.

3 Local Influence

The interest of the local influence method is to investigate the behavior of some influence measure when perturbations are made in the model (or data). Let $\boldsymbol{\omega}$ be a q-dimensional vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^q$, the perturbed log-likelihood function is denoted by $L(\boldsymbol{\theta}|\boldsymbol{\omega})$. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$. To assess the influence of minor perturbations on the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$, one may consider the likelihood displacement $LD(\boldsymbol{\omega}) = 2\{L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_\omega)\}$, where $\widehat{\boldsymbol{\theta}}_\omega$ denotes the maximum likelihood estimate under $L(\boldsymbol{\theta}|\boldsymbol{\omega})$.

Cook (1986) suggests to study the local behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. Based on this measure he shows that the normal curvature in the direction $\boldsymbol{\ell}$ is given by $C_\ell(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^T\boldsymbol{\Delta}^T\ddot{\boldsymbol{L}}^{-1}\boldsymbol{\Delta}\boldsymbol{\ell}|$, where $||\boldsymbol{\ell}|| = 1$, $\boldsymbol{\Delta} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T$ and $-\ddot{\boldsymbol{L}}$ is the observed information matrix, both $\boldsymbol{\Delta}$ and $\ddot{\boldsymbol{L}}$ are evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. In order to have a curvature invariant under a uniform change of scale Poon and Poon (1999) proposed the conformal normal curvature $B_\ell(\boldsymbol{\theta}) = C_\ell(\boldsymbol{\theta})/||2\boldsymbol{\Delta}^T\ddot{\boldsymbol{L}}^{-1}\boldsymbol{\Delta}||_F$, where $||\cdot||_F$ denotes the Frobenius norm defined as $||\boldsymbol{A}||_F = \{\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A})\}^{1/2}$ with $\boldsymbol{A}$ being a $r \times s$ matrix. An interesting property of the conformal normal curvature is that for any unitary direction $\boldsymbol{\ell}$ one has $0 \leq B_\ell(\boldsymbol{\theta}) \leq 1$. It allows comparison of curvatures among different elliptical error models.

A suggestion is to consider the direction $\boldsymbol{\ell}_{\max}$ corresponding to the largest curvature $B_{\ell_{\max}}(\boldsymbol{\theta})$. However, specific directions may also be evaluated; for example the index plot of $B_i = B_{\ell_i}(\boldsymbol{\theta})$ (see, for instance, Lesaffre and Verbeke, 1998), where $\boldsymbol{\ell}_i$ is a $q\times 1$ vector with one at the ith position and zeros elsewhere. From Cook (1986) and by using the quantities above we can also calculate the conformal normal curvatures $B_\ell(\boldsymbol{\beta})$ and $B_\ell(\boldsymbol{\alpha})$.

4 Curvature Derivation

In this section we derive the observed information matrix $-\ddot{\boldsymbol{L}}$ as well as $\boldsymbol{\Delta}$ for different perturbation schemes. These matrices are obtained using results of matrix differentiation (see, for instance, Magnus and Neudecker, 1988). The differentials $\mathrm{d}^2_\theta L(\boldsymbol{\theta})$ for the postulated model and $\mathrm{d}^2_{\theta\omega} L(\boldsymbol{\theta}|\boldsymbol{\omega})$ for the perturbed model are given in Appendix A.

4.1 Observed Information Matrix

Considering $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$, the log-likelihood function for the postulated model is given in (2). Thus, the observed information matrix evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ becomes given by $-\ddot{\boldsymbol{L}}(\widehat{\boldsymbol{\theta}}) = -\sum_{i=1}^n \ddot{\boldsymbol{L}}_i(\widehat{\boldsymbol{\theta}})$, with $\ddot{\boldsymbol{L}}_i$ having the partitioned form

$$\ddot{\boldsymbol{L}}_i(\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\Big|_{\theta=\widehat{\theta}} = \begin{pmatrix} \ddot{\boldsymbol{L}}_{11,i} & \ddot{\boldsymbol{L}}_{12,i} \\ \ddot{\boldsymbol{L}}_{12,i}^T & \ddot{\boldsymbol{L}}_{22,i} \end{pmatrix}, \tag{9}$$

where

$$\ddot{\boldsymbol{L}}_{11,i} = 2\boldsymbol{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1}\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + 2W_g'(\widehat{u}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\}\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i,$$

$$\ddot{\boldsymbol{L}}_{12,i} = \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}^T}\Big|_{\theta=\widehat{\theta}},$$

with

$$\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\alpha_r}\Big|_{\theta=\widehat{\theta}} = 2\boldsymbol{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1}\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + W_g'(\widehat{u}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\}\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i,$$

for $r = 1, \dots, k$, and

$$\ddot{\boldsymbol{L}}_{22,i} = \frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T}\Big|_{\theta=\widehat{\theta}}$$

whose (r, s)th element is given by

$$\begin{aligned}\frac{\partial^2 L_i(\boldsymbol{\theta})}{\partial\alpha_r\partial\alpha_s}\Big|_{\theta=\widehat{\theta}} &= \tfrac{1}{2}\operatorname{tr}\widehat{\boldsymbol{\Sigma}}_i^{-1}\{\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(s) - \ddot{\boldsymbol{\Sigma}}_i(r,s)\} \\ &\quad + \widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\{W_g'(\widehat{u}_i)\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r) - W_g(\widehat{u}_i)\ddot{\boldsymbol{\Sigma}}_i(r,s) \\ &\quad + W_g(\widehat{u}_i)\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(s) + W_g(\widehat{u}_i)\dot{\boldsymbol{\Sigma}}_i(s)\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i,\end{aligned}$$

for $r, s = 1, \dots, k$. Here, $\widehat{u}_i = \widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i$, $\widehat{\boldsymbol{r}}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i(\widehat{\boldsymbol{\alpha}})$, for $i = 1, \dots, n$, and $\dot{\boldsymbol{\Sigma}}_i(r) = \partial\boldsymbol{\Sigma}_i/\partial\alpha_r|_{\theta=\widehat{\theta}}$, $\ddot{\boldsymbol{\Sigma}}_i(r,s) = \partial^2\boldsymbol{\Sigma}_i/\partial\alpha_r\partial\alpha_s|_{\theta=\widehat{\theta}}$, $r, s = 1, \dots, k$. Note that, when $m_i = 1, \forall i$ (univariate case), the expressions above reduce to the ones given in Galea, Paula and Uribe-Opazo (2003).

4.2 Perturbation Schemes

We will consider the following perturbation schemes in the model defined in Section 2: case-weight, scale matrix, explanatory variable and response perturbations. The sensitivity of some model assumptions may be checked through appropriate perturbation schemes. In addition, this analysis can produce valuable information for the modeling process.

The $\boldsymbol{\Delta}$ matrix for each perturbation scheme assumes the form

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 \end{pmatrix},$$

where $\boldsymbol{\Delta}_1 = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^T|_{\theta=\widehat{\theta},\omega=\omega_0}$, $\boldsymbol{\Delta}_2 = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\omega}^T|_{\theta=\widehat{\theta},\omega=\omega_0}$ and q is the dimension of the perturbation vector $\boldsymbol{\omega}$ for the scheme under consideration.

4.2.1 Case-weight Perturbation

First consider the following arbitrary attribution of weights for the observations in the log-likelihood function:

$$L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i L_i(\boldsymbol{\theta}), \tag{10}$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ are the weights, which satisfies $0 \leq \omega_i \leq 1$, for $i = 1, \dots, n$, and $L_i(\boldsymbol{\theta})$ is defined as in (2). It is possible to note that, for $\omega_i = 0$ and $\omega_j = 1$, $j \neq i$, we perform the exclusion of the ith subject from the log-likelihood function. For this scheme the no perturbation vector is given by $\boldsymbol{\omega}_0 = (1, \dots, 1)^T \in \mathbb{R}^n$. Using the differentiation method we find

$$\begin{aligned}
\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\omega_i}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} &= \widehat{v}_i \boldsymbol{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) \text{ and} \\
\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\alpha_r\partial\omega_i}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} &= -\tfrac{1}{2}\{\operatorname{tr} \widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r) - \widehat{v}_i \widehat{\boldsymbol{r}}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i\},
\end{aligned}$$

for $j = 1, \dots, k$ and $i = 1, \dots, n$. Here $\widehat{v}_i = v_i(\widehat{\boldsymbol{\theta}})$, $i = 1, \dots, n$.

This perturbation scheme allows to identify those observations that exercise notable influence on the estimation process and consequently on the parameter estimates.

Note that, when $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$ has a random effect structure (see, for instance, Jennrich and Schluchter, 1986), the expressions for $\ddot{\boldsymbol{L}}$ and $\boldsymbol{\Delta}$ reduce to the ones obtained by Lesaffre and Verbeke (1998) for the normal case ($v_i(\boldsymbol{\theta}) = 1$). However, the models defined in this work extend the ones discussed by Lesaffre and Verbeke in the sense of allowing structures such as $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) = \boldsymbol{Z}_i\boldsymbol{D}(\boldsymbol{\alpha})\boldsymbol{Z}_i^T + \boldsymbol{\Psi}_i(\boldsymbol{\alpha})$ as well as to permit a larger range of error distributions; the members of the elliptical class are manipulated through $v_i = -2W_g(u_i)$, $\forall i$.

4.2.2 Scale Matrix Perturbation

The perturbation scheme is introduced by considering the model

$$\boldsymbol{Y}_i \sim EC_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \omega_i^{-1}\boldsymbol{\Sigma}_i, g), \qquad i = 1, \ldots, n, \tag{11}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T \in \mathbb{R}^n - \{\mathbf{0}\}$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^T$ such that $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$ given in (2). Taking differentials of $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ with respect to $\boldsymbol{\theta}$ and ω_i, we find

$$\left.\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\omega_i}\right|_{\theta=\widehat{\theta},\omega=\omega_0} = -2\{W_g(\widehat{u}_i) + \widehat{u}_i W_g'(\widehat{u}_i)\}\boldsymbol{X}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i \text{ and}$$

$$\left.\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\alpha_r\partial\omega_i}\right|_{\theta=\widehat{\theta},\omega=\omega_0} = -\{W_g(\widehat{u}_i) + \widehat{u}_i W_g'(\widehat{u}_i)\}\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i,$$

for $j = 1, \ldots, k$ and $i = 1, \ldots, n$, with $W_g(u_i)$ and $W_g'(u_i)$ being evaluated at $\widehat{u}_i = \widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i$.

This perturbation scheme may reveal those individuals that are most influential, in the sense, of the likelihood displacement on the scale structure and consequently on the $\boldsymbol{\alpha}$ estimate.

4.2.3 Explanatory Variable Perturbation

Here the interest is on perturbing a particular continuous explanatory variable, namely $\mathbf{x}_{it\omega} = \mathbf{x}_{it} + \boldsymbol{\omega}_i s_i$, where $\mathbf{x}_{it} \in \mathbb{R}^{m_i}$ is the tth column of the matrix $\boldsymbol{X}_i$, $\boldsymbol{\omega}_i$ denotes the $m_i \times 1$ perturbation vector and s_i is a scale factor. In this case the perturbed log-likelihood function equals $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$, where

$$L_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\tfrac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log g(u_{i\omega}), \tag{12}$$

$u_{i\omega} = \boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_{i\omega}$, $\boldsymbol{r}_{i\omega} = \boldsymbol{Y}_i - \boldsymbol{X}_{i\omega}\boldsymbol{\beta}$ and $\boldsymbol{X}_{i\omega} = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it} + \boldsymbol{\omega}_i s_i, \ldots, \mathbf{x}_{ip})$ for $i = 1, \ldots, n$. Let $N = \sum_{i=1}^n m_i$, thus the no perturbation vector is $\boldsymbol{\omega}_0 =$

$\mathbf{0} \in \mathbb{R}^N$. Taking differentials of $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\omega}_i$, we obtain

$$\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} = -\,2W_g(\widehat{u}_i)s_i(\widehat{\beta}_t \boldsymbol{X}_i^T - \boldsymbol{c}_t \widehat{\boldsymbol{r}}_i^T)\widehat{\boldsymbol{\Sigma}}_i^{-1} + 4W_g'(\widehat{u}_i)s_i\widehat{\beta}_t \boldsymbol{X}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1} \text{ and}$$

$$\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} = 2s_i\widehat{\beta}_t\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + W_g'(\widehat{u}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\}\widehat{\boldsymbol{\Sigma}}_i^{-1},$$

for $j = 1, \ldots, k$ and $i = 1, \ldots, n$. Here $\boldsymbol{c}_t$ denotes a $p \times 1$ vector with 1 at the tth position and zero elsewhere and $\widehat{\beta}_t$ denotes the tth element of $\widehat{\boldsymbol{\beta}}$.

This perturbation scheme allows to identify values of continuous explanatory variables that are very sensitive in the sense of the likelihood displacement. In particular, we may identify ill conditioning of the matrices $\boldsymbol{X}_i$, $i = 1, \ldots, n$.

4.2.4 Response Perturbation

A perturbation of the observed response $(\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T)^T$ is introduced by replacing $\boldsymbol{Y}_i$ by $\boldsymbol{Y}_{i\omega} = \boldsymbol{Y}_i + \boldsymbol{\omega}_i$, where $\boldsymbol{\omega}_i$ is an $m_i \times 1$ perturbation vector; in this case one has $\boldsymbol{\omega}_0 = \mathbf{0}$ with $N = \sum_{i=1}^n m_i$. The perturbed log-likelihood function is also given by $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}|\boldsymbol{\omega})$, where

$$L_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\tfrac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log g(u_{i\omega}), \quad (13)$$

$u_{i\omega} = \boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_{i\omega}$ and $\boldsymbol{r}_{i\omega} = \boldsymbol{Y}_{i\omega} - \boldsymbol{X}_i\boldsymbol{\beta}$, $i = 1, \ldots, n$. We obtain

$$\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_i^T}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} = -2\boldsymbol{X}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + 2W_g'(\widehat{u}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\}\widehat{\boldsymbol{\Sigma}}_i^{-1} \text{ and}$$

$$\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha_r \partial \boldsymbol{\omega}_i^T}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} = -2\widehat{\boldsymbol{r}}_i^T\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(r)\widehat{\boldsymbol{\Sigma}}_i^{-1}\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + W_g'(\widehat{u}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\}\widehat{\boldsymbol{\Sigma}}_i^{-1},$$

for $r = 1, \ldots, k$ and $i = 1, \ldots, n$.

An interesting aspect of this perturbation scheme is the connection with generalized leverage (Wei, Hu e Fung, 1998) as showed in Appendix B. For instance, if $\boldsymbol{\alpha}$ is fixed or known the index plot of B_i may reveal those observations with high influence on their fitted values.

5 Application

To illustrate the methodology developed in the previous sections, we will consider the orthodontic data set introduced by Potthoff and Roy (1964), where dental measurements were made on 11 girls and 16 boys at ages 8, 10, 12 and

14. The response variable was the distance, in millimeters, from the center of pituitary to the pterygomaxillary fissure. Figure 1 displays the individual profiles of girls and boys. Various authors have analyzed this data set. For instance, Pendergast and Broffitt (1985) consider robust procedures for growth curve models while Pinheiro, Liu and Wu (2001) used a linear mixed model with the Student-t distribution in a hierarchical setting. On the other hand, Pan, Fang and von Rosen (1997) and Pan and Bai (2003) present sensitivity studies using the local influence method and Pan (2002) considers elimination diagnostic procedures. More recently, Savalli, Paula and Cysneiros (2006) apply a score-type test for assessing the variance components. In all these works outlying and influential observations were detected under normal error, indicating that robustness methods could be used in order to reduce the influence of such observations on the parameter estimates.

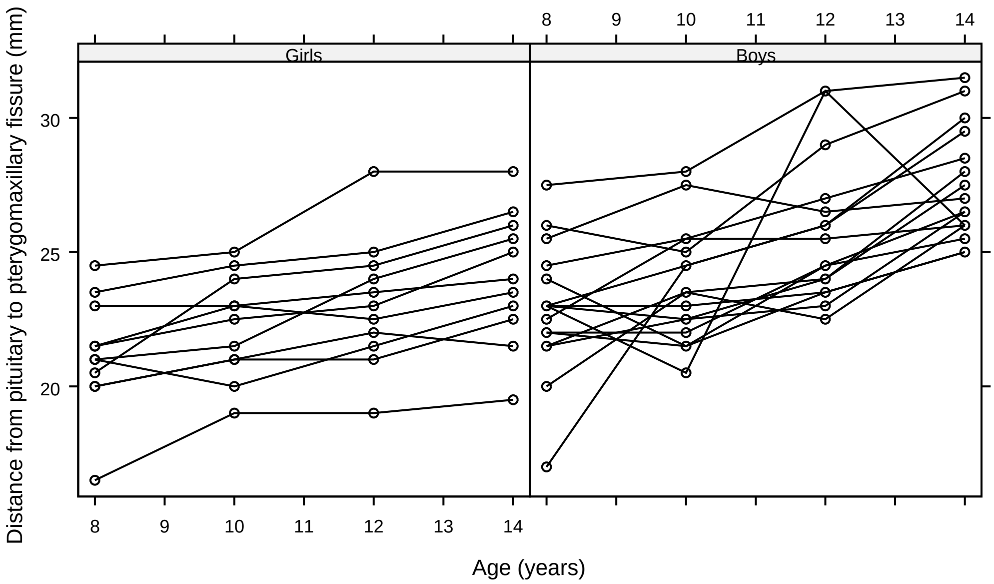

Fig. 1. Individual profiles for the orthodontic data set.

Thus, we suggest for analyzing this data set the following random intercept-slope elliptical model:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, 27,$$

where $\boldsymbol{Y}_i$ is a 4-dimensional random vector of responses from the ith observation, $\boldsymbol{X}_i$ is an 4×4 design matrix such that,

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{pmatrix}$$

for the girls' group, and

$$\boldsymbol{X}_i = \begin{pmatrix} 0 & 0 & 1 & 8 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 1 & 12 \\ 0 & 0 & 1 & 14 \end{pmatrix}$$

for the boys' group, $\boldsymbol{\beta} = (\alpha_1, \alpha_2, \beta_1, \beta_2)^T$ is the fixed parameter vector, where α_1 and β_1 represent the intercept and the slope for the girls' group while α_2 and β_2 correspond to the intercept and slope for the boys' group, respectively, and $\boldsymbol{Z}_i$ is a 4×2 design matrix of the random effects $\boldsymbol{b}_i$ given by

$$\boldsymbol{Z}_i^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix}, \quad i = 1, \ldots, 27.$$

In addition, we will assume the joint distribution of $(\boldsymbol{Y}_i^T, \boldsymbol{b}_i^T)^T$ as

$$\begin{pmatrix} \boldsymbol{Y}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim EC_6\left(\begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_4 & \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{Z}_i^T & \boldsymbol{D} \end{pmatrix}, g \right),$$

where $\boldsymbol{D}$ is assumed to be an unstructured symmetric 2×2 matrix with elements d_{11}, d_{12} and d_{22}. The inferences will be based on the marginal model given by $\boldsymbol{Y}_i \sim EC_4(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i(\boldsymbol{\alpha}), g)$, where $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_4$ with $\boldsymbol{\alpha} = (d_{11}, d_{12}, d_{22}, \sigma^2)^T$. Note that this model takes the same form described in Section 2.

5.1 Analyses of the Fitted Models

In our analysis we will assume Student-t, power exponential and normal errors for comparative purposes. As suggested by Lange, Little and Taylor (1989) the Schwarz information criterion was used for choosing among some values of the degrees of freedom ν for the Student-t model and of the shape parameter λ

in the case of the power exponential distribution; we found $\nu = 5$ and $\lambda = \frac{2}{3}$, respectively. Therefore, for both models heavy-tailed distributions will be assumed for the errors. The maximum likelihood estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ for the three fitted models are given in Table 1.

Table 1
Parameter estimates of the three models fitted on the orthodontic data.

	Normal		Student-t		Power Exponential	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
α_1	17.373	(1.182)	17.610	(0.992)	17.568	(1.095)
β_1	0.480	(0.100)	0.459	(0.084)	0.462	(0.093)
α_2	16.341	(0.980)	16.948	(0.823)	16.699	(0.908)
β_2	0.784	(0.083)	0.716	(0.070)	0.744	(0.077)
d_{11}	4.557	(4.672)	3.270	(2.950)	1.185	(1.100)
d_{12}	-0.198	(0.379)	-0.133	(0.233)	-0.053	(0.088)
d_{22}	0.024	(0.034)	0.020	(0.022)	0.007	(0.008)
σ^2	1.716	(0.330)	0.887	(0.223)	0.358	(0.079)

We can notice from Table 1 that the intercept and slope estimates are similar among the three fitted models, however the standard errors of the Student-t and power exponential models are smaller than the ones of the normal model, indicating that the two models with longer-than-normal tails seem to produce more accurate maximum likelihood estimates. The inferences for the variance components are similar for the three fitted models, but the estimates are not comparable since they are in different scales.

In order to detect outlying observations in multivariate elliptical models the Mahalanobis distance $u_i = (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$, $i = 1, \ldots, n$ (Little, 1988, Lange, Little and Taylor, 1989 and Copt and Victoria-Feser, 2006) has been considered. For the normal case, one has that u_i follows a chi-squared distribution with m_i degrees of freedom. Thus, we can use as cutoff points the quantiles $\chi^2_{m_i}(\xi)$, where $0 < \xi < 1$, to identify outliers. The modified Mahalanobis distances $F_i = u_i/m_i \sim F_{m_i,\nu}$ and $T_i = U_i^{\lambda} \sim \text{Gamma}(\frac{1}{2}, \frac{m_i}{2\lambda})$ may be also considered for the Student-t distribution with ν degrees of freedom and power exponential distribution with shape parameter λ, respectively. Figure 2 displays such distances for the three fitted models. The cutoff lines corresponds to the quantile $\xi = 0.975$.

We can see from these figures that observations 20 and 24, which correspond to the measures of the 9th and 13th boys, appear as possible outliers. The estimated weights for these two boys are the smallest ones for both fitted models

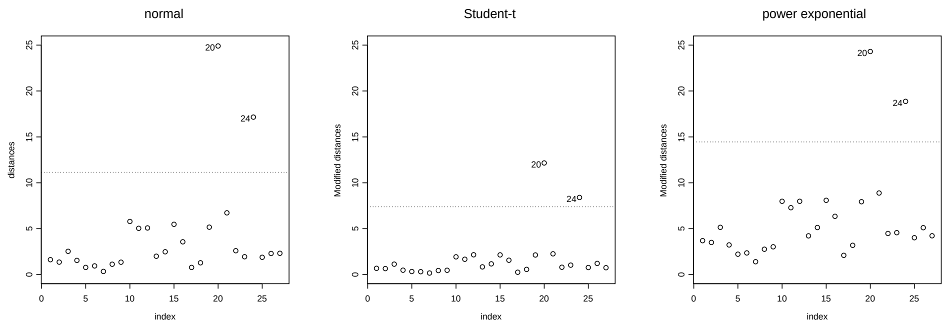

Fig. 2. Index plots of the Mahalanobis distances for the three fitted models.

with heavy-tailed error distributions (see Table 2), confirming the robustness aspects of the maximum likelihood estimates against outlying observations. For the normal case one has ($v_i(\boldsymbol{\theta}) = 1, \forall i$).

Table 2
Estimated weights for the Student-t and power exponential models.

Subject	1	2	3	4	5	6	7	8	9
Student-t	1.17	1.18	0.94	1.30	1.43	1.44	1.59	1.33	1.31
power exponential	0.35	0.36	0.29	0.37	0.45	0.43	0.56	0.40	0.38
Subject	10	11	12	13	14	15	16	17	18
Student-t	0.71	0.77	0.66	1.08	0.93	0.66	0.80	1.49	1.24
power exponential	0.24	0.25	0.24	0.32	0.29	0.23	0.26	0.46	0.37
Subject	19	20	21	22	23	24	25	26	27
Student-t	0.67	**0.17**	0.64	1.10	0.99	**0.23**	1.12	0.92	1.13
power exponential	0.24	**0.14**	0.22	0.32	0.31	**0.15**	0.33	0.30	0.32

For the purpose of identifying influential observations in the models fitted on the orthodontic data, some index plots of B_i will be performed in the sequel under three perturbation schemes discussed in the Section 4.

Case-weight perturbation. Based on this perturbation scheme, the index plots of $B_i(\boldsymbol{\beta})$ and $B_i(\boldsymbol{\alpha})$ for the three fitted models are displayed in Figures 3-4, respectively.

We can notice under normal error that subject 24 is the most influential on $\widehat{\boldsymbol{\beta}}$ as well as with smaller influence subjects 10, 11, 15 and 21. No one observation appears with outstanding influence under Student-t and power exponential errors. These results agree with the ones reported by Pan, Fang and von Rosen (1997), Pan (2002) and Pan and Bai (2003), and also with the comments given in Pinheiro, Liu and Wu (2001) and Savalli, Paula and Cysneiros (2006). Looking at Figure 4 we can observe the large influence of observation 20 on

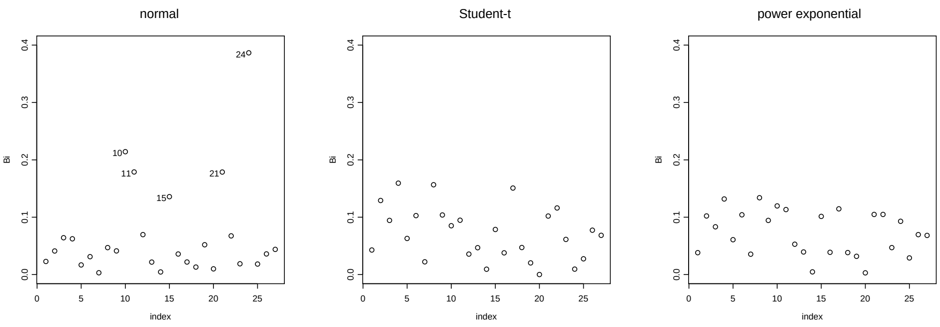

Fig. 3. Index plots of B_i for $\widehat{\boldsymbol{\beta}}$ under case-weight perturbation.

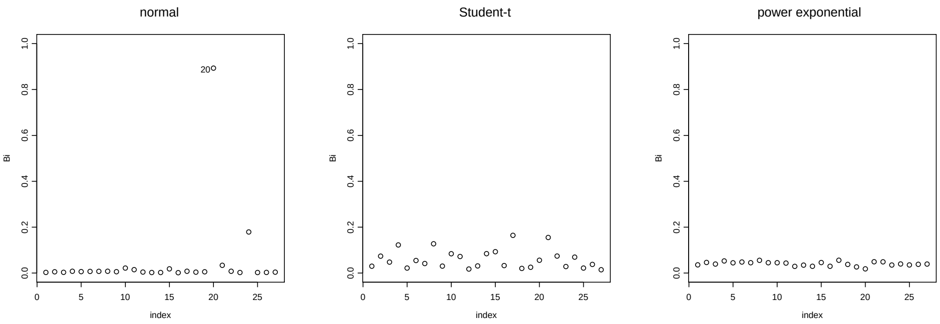

Fig. 4. Index plots of B_i for $\widehat{\boldsymbol{\alpha}}$ under case-weight perturbation.

$\widehat{\boldsymbol{\alpha}}$ under normal error. The index plots of $B_i(\boldsymbol{\theta})$ (omitted here) are similar to the ones given in Figure 4.

Scale matrix perturbation. Figures 5 and 6 show the index plots of $B_i(\boldsymbol{\beta})$ and $B_i(\boldsymbol{\alpha})$, for the normal, Student-t and power exponential models. The index plots of $B_i(\boldsymbol{\theta})$ are not presented here because they are very similar to the ones given in Figure 6.

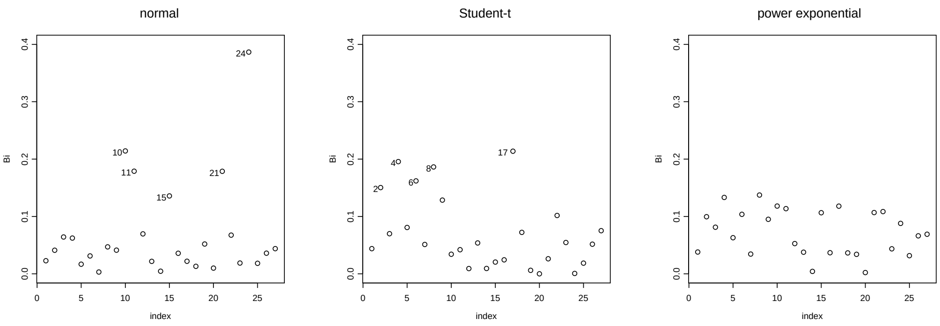

Fig. 5. Index plots of B_i for $\widehat{\boldsymbol{\beta}}$ under scale matrix perturbation.

Figures 5 and 6 are similar to Figures 3 and 4, except that under Student-t error five observations appear with moderate influence on $\widehat{\boldsymbol{\beta}}$ while under power exponential error one observation is pointed out with some influence on $\widehat{\boldsymbol{\alpha}}$. These results agree with the ones reported by Pan, Fang and von Rosen

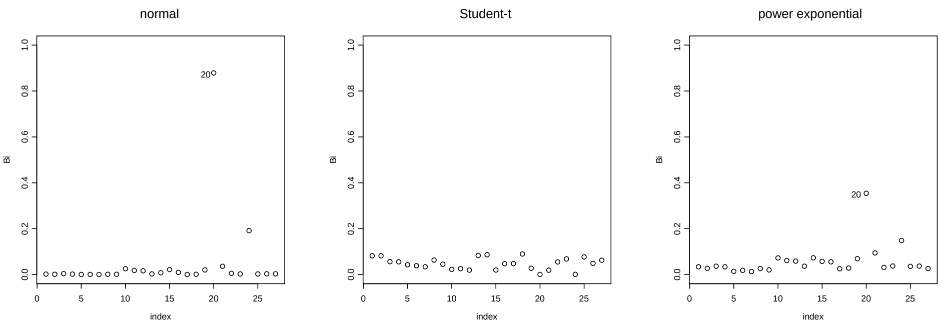

Fig. 6. Index plots of B_i for $\widehat{\boldsymbol{\alpha}}$ under scale matrix perturbation.

(1997), Pan (2002) and Pan and Bai (2003) under normal error.

Response perturbation. Here the response perturbation scheme is considered and the index plots of $B_i(\widehat{\boldsymbol{\theta}})$ are given in Figure 7. It is interesting to note in this case that we can also extract the influence of each individual measurement. From Figure 7 we can notice some influence under normal error, particularly when the responses of the individuals 20 and 24 are perturbed.

Thus, our main conclusion for this example is that the maximum likelihood estimates from the elliptical models with heavy-tailed error distributions seem to be more robust against outlying observations, as expected, but also against the three perturbation schemes considered for illustration than the estimates from the normal model.

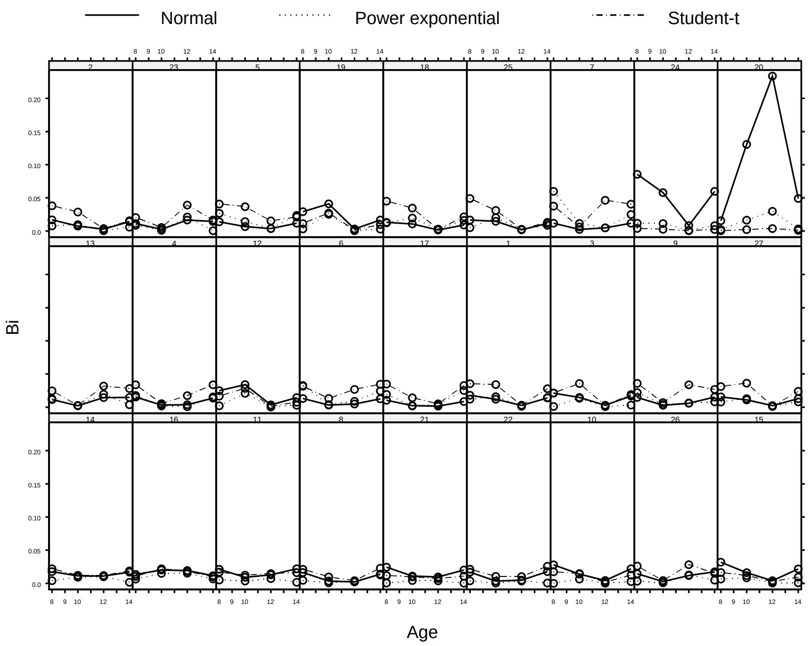

Fig. 7. Index plots of B_i for $\widehat{\boldsymbol{\theta}}$ under response perturbation.

6 Concluding Remarks

The present work generalizes the results given in Lesaffre and Verbeke (1998) for a large class of linear models with longitudinal structure that includes both light- and heavy-tailed error distributions. We allow the inclusion of structured forms for the scale matrix $\boldsymbol{\Sigma}_i$. Many structures typically used for modeling repeated-measurement data, such that, compound symmetric, heterogeneous, Toeplitz and unstructured are known as linear scale structures. In these cases, we can have important simplifications in the differential expressions $\mathrm{d}^2_\theta L(\boldsymbol{\theta})$ and $\mathrm{d}^2_{\theta\omega} L(\boldsymbol{\theta}|\boldsymbol{\omega})$ for different perturbation schemes using the fact that $\mathrm{d}\boldsymbol{\alpha} = \boldsymbol{D}_i\,\mathrm{dvec}\boldsymbol{\Sigma}_i$, where $\boldsymbol{D}_i$ is the duplication matrix given by Nel (1980, Sec. 6). Then, $\mathrm{d}\boldsymbol{\alpha} = \boldsymbol{D}_i\,\mathrm{dvec}\boldsymbol{\Sigma}_i$ and $\mathrm{d}^2\boldsymbol{\alpha} = \boldsymbol{0}$. However, this property does not follow, for instance, for structures such as ARMA (see, for instance, Jennrich and Schluchter, 1986). Due to this consideration our derivations were made for the general case.

Further work on influence diagnostics for linear mixed models considering a subclass of the elliptically contoured distributions, known as scale mixtures of normal distributions, is being developed by the authors and will be the subject of an incoming paper. Such models are one natural extension of the hierarchical model described in Pinheiro, Liu and Wu (2001).

Acknowledgements

This work was supported by the National Scholarship Program of Chile, MIDEPLAN-Chile and CNPq-Brazil and by the project FONDECYT 1030588, Chile. The authors thank the Associate Editor and two anonymous referees for valuable suggestions.

A Differential Calculations

In this Appendix we derive the matrices $\ddot{\boldsymbol{L}}^{-1}$ and $\boldsymbol{\Delta}$ for different perturbation schemes under the elliptical linear model with longitudinal structure. These measures are obtained efficiently using the differentiation method described in Magnus and Neudecker (1988).

A.1 Observed Information Matrix

Using results of matrix differentiation we obtain $\mathrm{d}^2_\theta L(\boldsymbol{\theta}) = \sum_{i=1}^n \mathrm{d}^2_\theta L_i(\boldsymbol{\theta})$, with

$$\mathrm{d}_\beta L_i(\boldsymbol{\theta}) = -2W_g(u_i)\boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\,\mathrm{d}\boldsymbol{\beta} \text{ and} \tag{A.1}$$

$$\mathrm{d}^2_\beta L_i(\boldsymbol{\theta}) = 2(\mathrm{d}\boldsymbol{\beta})^T\boldsymbol{X}_i^T\boldsymbol{\Sigma}_i^{-1}\{W_g(u_i)\boldsymbol{\Sigma}_i + 2W_g'(u_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\,\mathrm{d}\boldsymbol{\beta}, \tag{A.2}$$

where $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$. From (A.1) it follows that

$$\mathrm{d}^2_{\alpha\beta} L_i(\boldsymbol{\theta}) = 2\boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\{W_g(u_i)\boldsymbol{\Sigma}_i + W_g'(u_i)\boldsymbol{r}_i\boldsymbol{r}_i^T\}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\,\mathrm{d}\boldsymbol{\beta}. \tag{A.3}$$

Finally, the first and second differentials of $L_i(\boldsymbol{\theta})$ with respect to $\boldsymbol{\alpha}$ are given by the expressions

$$\mathrm{d}_\alpha L_i(\boldsymbol{\theta}) = -\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\Sigma}_i - W_g(u_i)\boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i \text{ and} \tag{A.4}$$

$$\begin{aligned}\mathrm{d}^2_\alpha L_i(\boldsymbol{\theta}) &= \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\Sigma}_i - \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}^2\boldsymbol{\Sigma}_i \\ &\quad + \boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}\{W_g'(u_i)(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i) - W_g(u_i)\,\mathrm{d}^2\boldsymbol{\Sigma}_i \\ &\quad + W_g(u_i)(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\Sigma}_i + W_g(u_i)(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\Sigma}_i\}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i.\end{aligned} \tag{A.5}$$

Evaluating (A.1)-(A.5) at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ we obtain the observed information matrix given in equation (9).

A.2 Perturbation Schemes

For each perturbation scheme we derive the differential $\mathrm{d}^2_{\theta\omega} L(\boldsymbol{\theta}|\boldsymbol{\omega})$. We obtain the matrix $\boldsymbol{\Delta} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T|_{\theta=\widehat{\theta},\omega=\omega_0}$ by evaluating $\mathrm{d}^2_{\theta\omega} L(\boldsymbol{\theta}|\boldsymbol{\omega})$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

A.2.1 Case-weights Perturbation

For this perturbation scheme we have

$$\begin{aligned}\mathrm{d}^2_{\beta\omega_i} L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= v_i(\mathrm{d}\boldsymbol{\beta})^T\boldsymbol{X}_i^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\,\mathrm{d}\omega_i \text{ and} \\ \mathrm{d}^2_{\alpha\omega_i} L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -\tfrac{1}{2}\{\operatorname{tr}\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i) - v_i\boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i\}\,\mathrm{d}\omega_i,\end{aligned}$$

with $v_i = v_i(\boldsymbol{\theta})$ and $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$, for $i = 1, \ldots, n$.

A.2.2 Scale Matrix Perturbation

Taking differentials of $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ with respect to $\boldsymbol{\theta}$ and ω_i, we obtain

$$\begin{aligned}
\mathrm{d}^2_{\beta\omega_i}L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2\{W_g(u_{i\omega}) + \omega_i u_i W_g'(u_{i\omega})\}(\mathrm{d}\boldsymbol{\beta})^T\boldsymbol{X}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i\,\mathrm{d}\omega_i \text{ and}\\
\mathrm{d}^2_{\alpha\omega_i}L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -\{W_g(u_{i\omega}) + \omega_i u_i W_g'(u_{i\omega})\}\boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i\,\mathrm{d}\omega_i,
\end{aligned}$$

where $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$ and $u_{i\omega} = \omega_i u_i$ with $u_i = \boldsymbol{r}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i$, $i = 1, \ldots, n$.

A.2.3 Explanatory Variable Perturbation

Using the differentiation method, we have that

$$\begin{aligned}
\mathrm{d}^2_{\beta\omega_i}L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -s_i 2W_g(u_{i\omega})(\mathrm{d}\boldsymbol{\beta})^T(\boldsymbol{c}_t^T\boldsymbol{\beta}\boldsymbol{X}_{i\omega} - \boldsymbol{c}_t\boldsymbol{r}_{i\omega}^T)\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\omega}_i\\
&\quad + s_i 4W_g'(u_{i\omega})\boldsymbol{c}_t^T\boldsymbol{\beta}(\mathrm{d}\boldsymbol{\beta})^T\boldsymbol{X}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_{i\omega}\boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\omega}_i \text{ and}\\
\mathrm{d}^2_{\alpha\omega_i}L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= 2s_i\boldsymbol{c}_t^T\boldsymbol{\beta}\boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\{W_g(u_{i\omega})\boldsymbol{\Sigma}_i + W_g'(u_{i\omega})\boldsymbol{r}_{i\omega}\boldsymbol{r}_{i\omega}^T\}\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\omega}_i,
\end{aligned}$$

where $\boldsymbol{X}_{i\omega} = \boldsymbol{X}_i + s_i\boldsymbol{\omega}_i\boldsymbol{c}_t^T$, with s_i being a scale factor, $\boldsymbol{\omega}_i$ an $m_i \times 1$ perturbation vector and $\boldsymbol{c}_t$ a p-dimensional vector with 1 at tth position and zero elsewhere. Here $\boldsymbol{r}_{i\omega} = \boldsymbol{r}_i - s_i\boldsymbol{c}_t^T\boldsymbol{\beta}\boldsymbol{\omega}_i$ and $u_{i\omega} = \boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_{i\omega}$, with $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$, for $i = 1, \ldots, n$.

A.2.4 Response Perturbation

For this scheme it is possible to show that

$$\begin{aligned}
\mathrm{d}^2_{\beta\omega_i}L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2(\mathrm{d}\boldsymbol{\beta})^T\boldsymbol{X}_i^T\boldsymbol{\Sigma}_i^{-1}\{W_g(u_{i\omega})\boldsymbol{\Sigma}_i + 2W_g'(u_{i\omega})\boldsymbol{r}_{i\omega}\boldsymbol{r}_{i\omega}^T\}\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\omega}_i \text{ and}\\
\mathrm{d}^2_{\alpha\omega_i}L(\boldsymbol{\theta}|\boldsymbol{\omega}) &= -2\boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}(\mathrm{d}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\{W_g(u_{i\omega})\boldsymbol{\Sigma}_i + W_g'(u_{i\omega})\boldsymbol{r}_{i\omega}\boldsymbol{r}_{i\omega}^T\}\boldsymbol{\Sigma}_i^{-1}\,\mathrm{d}\boldsymbol{\omega}_i,
\end{aligned}$$

where $u_{i\omega} = \boldsymbol{r}_{i\omega}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_{i\omega}$ and $\boldsymbol{r}_{i\omega} = \boldsymbol{r}_i + \boldsymbol{\omega}_i$, with $\boldsymbol{\omega}_i$ denoting an $m_i \times 1$ perturbation vector and $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$, $i = 1, \ldots, n$.

B Connection between Local Influence and Generalized Leverage

In order to examine the relationship between local influence and generalized leverage under additive perturbations in the response values we will assume that $\boldsymbol{\alpha}$ is fixed. Thus, it follows from Wei, Hu e Fung (1998) that the generalized leverage matrix takes the form

$$\boldsymbol{GL}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{X}\{-\ddot{\boldsymbol{L}}(\widehat{\boldsymbol{\beta}})\}^{-1}\boldsymbol{X}^T\boldsymbol{A} = \boldsymbol{X}\Big(\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{A}_i\boldsymbol{X}_i\Big)^{-1}\boldsymbol{X}^T\boldsymbol{A},$$

where $\boldsymbol{X}^T = (\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_n^T)$ and $\boldsymbol{A} = \text{blc diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$ with $\boldsymbol{A}_i = -2\widehat{\boldsymbol{\Sigma}}_i^{-1}$ $\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + 2W_g(\widehat{u}_i)\widehat{\boldsymbol{r}}_i\widehat{\boldsymbol{r}}_i^T\}\widehat{\boldsymbol{\Sigma}}_i^{-1}$, for $i = 1, \ldots, n$. On the other hand, the normal curvature in this case is given by $C_\ell(\boldsymbol{\beta}) = 2|\boldsymbol{\ell}^T\boldsymbol{B}\boldsymbol{\ell}|$, where $\boldsymbol{B} = \boldsymbol{\Delta}_1^T\{\ddot{\boldsymbol{L}}(\widehat{\boldsymbol{\beta}})\}^{-1}\boldsymbol{\Delta}_1$. Since $\boldsymbol{\Delta}_1 = \boldsymbol{X}^T\boldsymbol{A}$, we obtain the relationship $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{G}\boldsymbol{L}(\widehat{\boldsymbol{\beta}})$.

References

Arellano, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Beckman, R.J., Nachtsheim, C.J. and Cook, R.D., 1987. Diagnostics for mixed-model analysis of variance. Technometrics 29, 413-426.

Cook, R.D., 1986. Assessment of local influence (with discussion). Journal of the Royal Statistical Society, Ser. B 48, 133-169.

Copt, S. and Victoria-Feser, M., 2006. High breakdown inference in the mixed linear model. Journal of the American Statistical Association (to appear).

Cysneiros, F.J.A. and Paula, G.A., 2004. One-sided test in linear models with multivariate t-distribution. Communications in Statistics - Simulation & Computation 33, 747-771.

Dennis, J.E. and Schnabel, R.B., 1996. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM Publications, Philadelphia.

Díaz-García, J.A., Galea, M. and Leiva-Sánchez, V., 2003. Influence diagnostics for elliptical multivariate linear regression models. Communications in Statistics - Theory & Methods 32, 625-641.

Fang, K.T., Kotz, S. and Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London.

Fernández, C. and Steel, M.F.J., 1999. Multivariate Student-t regression models: pitfalls and inference. Biometrika 86, 153-167.

Galea, M., Paula, G.A. and Bolfarine, H., 1997. Local influence in elliptical linear regression models. The Statistician 46, 71-79.

Galea, M., Paula, G.A. and Uribe-Opazo, M., 2003. On influence diagnostic in univariate elliptical linear regression models. Statistical Papers 44, 23-45.

Gómez, E., Gómez-Villegas, M.A. and Marín, J.M., 1998. A multivariate generalization of the power exponential family of distributions. Communications in Statistics - Theory & Methods 27, 589-600.

Huggins, R.M., 1993. A robust approach to the analysis of repeated measures. Biometrics 49, 715-720.

Kowalski, J., Mendoza-Blanco, J.R., Tu, X. M. and Gleser, L. J., 1999. On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions. Communications in Statistics - Theory & Methods 28, 2119-2140.

Jennrich, R.I. and Schluchter, M.D., 1986. Unbalanced repeated-measures models with structured covariance matrices. Biometrics 42, 805-820.

Lange, K.L., Little, R.J.A. and Taylor, J.M.G., 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84, 881-896.

Lesaffre, E. and Verbeke, G., 1998. Local influence in linear mixed models. Biometrics 54, 570-582.

Lindsey, J.K., 1999. Multivariate elliptically contoured distributions for repeated measurements. Biometrics 55, 1277-1280.

Little, R.J.A., 1988. Robust estimation of the mean and covariance matrix from data with missing values. Applied Statistics 37, 23-38.

Liu, S., 2000. On local influence for elliptical linear models. Statistical Papers 41, 211-224.

Liu, S., 2002. Local influence in multivariate elliptical linear regression models. Linear Algebra and its Applications 354, 159-174.

Magnus J.R. and Neudecker, H., 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.

Nel, D.G., 1980. On matrix differentiation in statistics. South African Statistical Journal 14, 137-193.

Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F., 2001. Local influence to detect influential data structures for generalized linear mixed models. Biometrics 57, 1166-1172.

Pan, J., 2002. Influential observation identification in the growth curve model with Rao's simple covariance structure. Communications in Statistics - Theory & Methods 31, 813-831.

Pan, J., Fang, K.T. and von Rosen, D., 1997. Local influence assessment in the growth curve model with unstructured covariance. Journal of Statistical Planning and Inference 62, 263-278.

Pan, J. and Bai, P., 2003. Local influence analysis in the growth curve model with Rao's simple covariance structure. Journal of Applied Statistics 7, 771-781.

Pendergast, J.F. and Broffitt, J.D., 1985. Robust estimation in growth curve models Communications in Statistics - Theory & Methods 14, 1919-1939.

Pinheiro, J.C., Liu, C. and Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. Journal of Computational and Graphical Statistics 10, 249-276.

Poon, W. and Poon, Y.S., 1999. Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society, Series B 61, 51-61.

Potthoff, R.F. and Roy, S.N., 1964. A generalized multivariate analysis of a variance model useful especially for growth curve problems. Biometrika 51, 313-326.

Savalli, C., Paula, G.A. and Cysneiros, F.J.A., 2006. Assessment of variance components in elliptical linear mixed models. Statistical Modelling, Vol. 6 (to appear).

Wei, B.C., Hu, Y.Q. and Fung, W.K., 1998. Generalized leverage and its applications. Scandinavian Journal of Statistics 25, 25-37.

Welsh, A.H. and Richardson, A.M., 1997. Approaches to the robust estimation of mixed models. In: G.S. Maddala and C.R. Rao (Eds.), Handbook of Statistics, Vol. 15, Elsevier Science, Amsterdam, 343-384.

Zhu, H. and Lee, S., 2003. Local influence for generalized linear mixed models. The Canadian Journal of Statistics 31, 293-309.