Multivariate elliptical models with general parameterization

Artur J. Lemonte, Alexandre G. Patriota

Department of Statistics, University of São Paulo, São Paulo/SP, Brazil

Abstract

In this paper we introduce a general elliptical multivariate regression model in which the mean vector and the scale matrix have parameters (or/and covariates) in common. This approach unifies several important elliptical models, such as nonlinear regressions, mixed-effects model with nonlinear fixed effects, errors-in-variables models, and so forth. We discuss maximum likelihood estimation of the model parameters and obtain the information matrix, both observed and expected. Additionally, we derived the generalized leverage as well as the normal curvatures of local influence under some perturbation schemes. An empirical application is presented for illustrative purposes.

Key words: Elliptical distributions; Generalized leverage; Local influence; Maximum likelihood estimation; Multivariate models.

1 Introduction

It is well known that the normality assumption is not always tenable and alternative distributions (or methodologies) should be considered in such situations. One choice is the elliptical family of distributions which includes the normal one. This class of distributions has received an increasing attention in the statistical literature, particularly due to the fact of including important distributions as, for example, Student-t, power exponential, contaminated normal, among others, with heavier or lighter tails than the normal one.

We say that a $d \times 1$ random vector $\boldsymbol{Y}$ has a multivariate elliptical distribution with location parameter $\boldsymbol{\mu}$ $(d \times 1)$ and a positive definite scale matrix $\boldsymbol{\Sigma}$ $(d \times d)$ if its density function exists, it is given by (Fang et al., 1990)

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-1/2} g\big[(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\big], \quad \boldsymbol{y} \in \mathbb{R}^d, \tag{1}$$

where $g : \mathbb{R} \to [0, \infty)$ is such that $\int_0^\infty u^{\frac{d}{2}-1} g(u) du < \infty$. The function $g(\cdot)$ is known as the density generator. We will denote $\boldsymbol{Y} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, or, simply, $\boldsymbol{Y} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. When $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_d$, where $\boldsymbol{I}_d$ is a $d \times d$ identity matrix, we obtain the spherical family of densities. A detailed description of the elliptical multivariate class given in (1) can be found in Fang et al. (1990). Table 1, taken from Galea et al. (2000), reports examples of distributions in the elliptical family.

Table 1: Multivariate elliptical distributions.[†]

Distribution	Notation	Generating function
Normal	$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$g(u) = c\exp(-u/2), \quad u \geq 0$
Student-t	$t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$g(u) = c(1 + u/\nu)^{-(\nu+d)/2}, \quad u \geq 0$
Contaminated normal	$\mathcal{CN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta, \tau)$	$g(u) = c\{(1-\delta)\exp(-u/2)$ $+\delta\tau^{-d/2}\exp(-u/(2\tau))\}, \quad u \geq 0$
Cauchy	$\mathcal{C}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$g(u) = c(1+u)^{-(d+1)/2}, \quad u \geq 0$
Logistic	$\mathcal{L}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$g(u) = c\mathrm{e}^{-u}/(1+\mathrm{e}^{-u})^2, \quad u \geq 0$
Power Exponential	$\mathcal{PE}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha)$	$g(u) = c\exp(-u^{\alpha}/2), \quad u \geq 0$

[†] c denotes the normalizing constant.

Assuming $g(\cdot)$ continuous and differentiable, it is useful to define the following quantities in the elliptic context:

$$W_g(u) = \frac{\mathrm{d}}{\mathrm{d}u}\log g(u) = \frac{g'(u)}{g(u)} \qquad \text{and} \qquad W_g'(u) = \frac{\mathrm{d}}{\mathrm{d}u}W_g(u) = \frac{g''(u)}{g(u)} - \left(\frac{g'(u)}{g(u)}\right)^2,$$

where $g'(u) = \mathrm{d}g(u)/\mathrm{d}u$ and $g''(u) = \mathrm{d}g'(u)/\mathrm{d}u$. For example, we have

$$W_g(u) = -\frac{1}{2}\left(\frac{\nu+d}{\nu+u}\right) \qquad \text{and} \qquad W_g'(u) = \frac{1}{2}\left(\frac{\sqrt{\nu+d}}{\nu+u}\right)^2$$

for Student-t distribution with ν degrees of freedom (Lange et al., 1989) and

$$W_g(u) = -\frac{\alpha u^{\alpha-1}}{2} \qquad \text{and} \qquad W_g'(u) = -\frac{\alpha(\alpha-1)u^{\alpha-2}}{2}$$

for power exponential (Gómes et al., 1998).

Multivariate elliptical regression models have been extensively studied in the statistical literature. In fact, some important references are Lange et al. (1989), Welsh and Richardson (1997), Kowalski et al. (1999), Fernández and Steel (1999), Galea et al. (2000), Liu (2000, 2002), Díaz–García et al. (2003), Cysneiros and Paula (2004), Savalli et al. (2006), Díaz–García et al. (2007), Osorio et al. (2007) and Russo et al. (2009), among others. The class of models proposed in this article includes all the models considered in the papers above and several others as special cases, e.g., multivariate errors-in-variables models, either with homoskedastic or heteroskedastic structures, mixed models with some regressors subject to measurement errors, and so forth. Our approach unifies several important models which can be thought from a multivariate elliptical model. Here, the modelling is made directly in the (observable) response variable (in mixed models context, it is known as marginal model).

In this paper we introduce a class of multivariate regression models with general parameterization based on the elliptical distribution given in (1). Here, general parameterization has the same meaning as defined by Patriota and Lemonte (2009). We consider that the mean vector and the positive definite scale matrix share parameters. For example, in structural errors-in-variables models some variables cannot be measured exactly, but instead it is observed surrogate variables contaminated with errors. This characteristic makes the mean vector shares parameters with the scale matrix of the observed variables. Thus, the model proposed in this paper is justified. Additionally, we develop local influence diagnostics based on minor perturbations in the data and in the assumed model and derive an expression for the generalized leverage.

The rest of the paper is organized as follows. Section 2 presents the model and discusses the estimation of the model parameters by maximum likelihood. We present the score function, Fisher information matrix and an iterative process to obtain the maximum likelihood estimates. Section 3 deals with some basic calculations related with local influence. The normal curvatures of local influence are derived under some perturbation schemes in Section 4. An expression for the generalized leverage is derived in Section 5. A special model is considered in Section 6. Section 7 contains an empirical application. Finally, some concluding remarks are made in Section 8.

2 The model

Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ be observable independent vectors which the number of responses measured in the ith observation is q_i. Following the same idea introduced in Patriota and Lemonte (2009), the multivariate elliptical regression model with general parameterization can be written as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}) + \boldsymbol{u}_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

with $\boldsymbol{u}_i \sim \mathcal{E}_{q_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$ and hence $\boldsymbol{Y}_i \sim \mathcal{E}_{q_i}(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$. Also, $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$ is the mean and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{w}_i)$ is the positive definite scale matrix, where $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ are $m_i \times 1$ and $k_i \times 1$ nonstochastic vectors of auxiliary variables, respectively, associated with the ith observed response $\boldsymbol{Y}_i$ which may have common components. Both $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ have known functional forms and are twice differentiable with respect to each element of $\boldsymbol{\theta}$. Additionally, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$ is a p-vector of unknown parameters of interest (where $p < n$ and it is fixed). Since $\boldsymbol{\theta}$ must be identifiable in model (2), the functions $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are defined to accomplish such restriction.

It is important to observe that $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ is proportional to the variance-covariance matrix of $\boldsymbol{Y}_i$ by a quantity $\xi_i > 0$ which depends on the assumed elliptical distribution. For example, under normal and Student-t models, $\xi_i = 1$ and $\xi_i = \nu/(\nu - 2)$, respectively, for $\nu > 2$. For further details the reader is referred to Fang et al. (1990).

The class of models presented in (2) is quite broad and includes several important statistical models. As a first example, we can mention linear and nonlinear regression models, either homoskedastic or heteroskedastic. Recently, heteroskedastic structural measurement error models have been studied by many authors, for instance, Kulathinal et al. (2002), Cheng and Riu (2006), Kelly (2007), de Castro et al. (2008) and Patriota et al. (2009). These models can also be formulated as in (2). Structural equation models (e.g., Bollen, 1989; Lee et al., 2006) is a rich class of models with latent variables that can be put as in (2). As can be seen, model (2) encompasses a wide range of models and our list of examples is by no means exhaustive. Section 6 presents an important special case that shows the applicability of the general formulation.

Let $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{w}_i)$, $\boldsymbol{z}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$ and $u_i = \boldsymbol{z}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{z}_i$. The log-likelihood function associated with (2), except for a constant term, is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{3}$$

where $\ell_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log g(u_i)$. The functions $g(\cdot)$, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ must be defined in such way that $\ell(\boldsymbol{\theta})$ be a regular function with respect to $\boldsymbol{\theta}$ (Cox and Hinkley, 1974, Ch. 9). To obtain the score function and the Fisher information matrix, we need to derive $\ell(\boldsymbol{\theta})$ with respect to the unknown parameters and then computing some moments of such derivatives. We suppose that such derivatives exist. To compute the derivatives of $\ell(\boldsymbol{\theta})$ we make use of matrix differentiation methods (Magnus and Neudecker, 1988).

Some additional notation is in order. Let

$$\boldsymbol{a}_{i(r)} = \frac{\partial \boldsymbol{\mu}_i}{\partial \theta_r}, \qquad \boldsymbol{C}_{i(r)} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \theta_r}, \qquad \boldsymbol{A}_{i(r)} = -\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}_{i(r)}\boldsymbol{\Sigma}_i^{-1},$$

for $r = 1, 2, \ldots, p$. Additionally, let

$$\boldsymbol{F}_i = \begin{pmatrix} \boldsymbol{D}_i \\ \boldsymbol{V}_i \end{pmatrix}, \quad \boldsymbol{H}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{0} \\ \boldsymbol{0} & 2\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i \end{pmatrix}^{-1}, \quad \boldsymbol{s}_i = \begin{bmatrix} v_i \boldsymbol{z}_i \\ -\mathrm{vec}(\boldsymbol{\Sigma}_i - v_i \boldsymbol{z}_i \boldsymbol{z}_i^\top) \end{bmatrix}, \tag{4}$$

where $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\theta}^\top$, $\boldsymbol{V}_i = \partial \mathrm{vec}(\boldsymbol{\Sigma}_i)/\partial \boldsymbol{\theta}^\top$ and $v_i = -2W_g(u_i)$. Here, we assume that $\boldsymbol{F} = (\boldsymbol{F}_1^\top, \boldsymbol{F}_2^\top, \ldots, \boldsymbol{F}_n^\top)$ has rank p, i.e. the functions $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ must be defined to hold such condition. Also, the "vec" operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other and "$\otimes$" indicates the Kronecker product. These quantities are used to find the derivatives of the log-likelihood function which will be required to compute the score function, Fisher information and all the normal curvatures of local influence.

By using the defined quantities in (4) and after some straightforward matrix algebra, the score function for $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{U}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \boldsymbol{F}_i^\top \boldsymbol{H}_i \boldsymbol{s}_i. \tag{5}$$

The expected Fisher information matrix for $\boldsymbol{\theta}$ is (see Appendix A)

$$\boldsymbol{K}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \boldsymbol{F}_i^{\top} \boldsymbol{H}_i \boldsymbol{M}_i \boldsymbol{H}_i \boldsymbol{F}_i, \tag{6}$$

with

$$\boldsymbol{M}_i = \begin{bmatrix} \frac{4d_{gi}}{q_i}\boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & \frac{8f_{gi}}{q_i(q_i+2)}\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\frac{4f_{gi}}{q_i(q_i+2)} - 1\right)\mathrm{vec}(\boldsymbol{\Sigma}_i)\mathrm{vec}(\boldsymbol{\Sigma}_i)^{\top} \end{bmatrix},$$

where $d_{gi} = \mathbb{E}(W_g^2(U_i)U_i)$ and $f_{gi} = \mathbb{E}(W_g^2(U_i)U_i^2)$, with $U_i = ||\boldsymbol{L}_i||^2$, $\boldsymbol{L}_i \sim \mathcal{E}_{q_i}(\mathbf{0}, \boldsymbol{I}_{q_i})$. Naturally, function $g(u)$ is such that d_{gi} and f_{gi} exist for all $i = 1, 2, \ldots, n$. It is possible to obtain closed-form expressions for d_{gi} and f_{gi} for some multivariate elliptical distributions. For example, we have (Mitchell, 1989)

$$d_{gi} = \frac{q_i}{4}\left(\frac{q_i + \nu}{q_i + \nu + 2}\right) \qquad \text{and} \qquad f_{gi} = \frac{q_i(q_i+2)}{4}\left(\frac{q_i + \nu}{q_i + \nu + 2}\right)$$

for Student-t distribution and

$$d_{gi} = \frac{\alpha^2}{2^{1/\alpha}}\Gamma\left(\frac{q_i - 2}{2\alpha} + 2\right)\Gamma\left(\frac{q_i}{2\alpha}\right)^{-1} \qquad \text{and} \qquad f_{gi} = \frac{q_i(q_i + 2\alpha)}{4}$$

for power exponential, where $\Gamma(\cdot)$ denotes the gamma function. It should be noticed that matrix $\boldsymbol{M}_i$ has all information about the adopted distribution. Also, note that the expected Fisher information is a quadratic form which can be attained through simple matrix operations. If $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ have not parameters in common, i.e. $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\theta}_1)$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_2)$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\top}, \boldsymbol{\theta}_2^{\top})^{\top}$, then $\boldsymbol{F}_i = \text{block–diag}\{\boldsymbol{F}_i(\boldsymbol{\theta}_1), \boldsymbol{F}_i(\boldsymbol{\theta}_2)\}$ and the parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ will be orthogonal.

It is noteworthy that d_{gi} and f_{gi} may have unknown quantities (for instance, the degrees of freedom ν in the Student-t distribution and the shape parameter α in the power exponential distribution) and one may want to estimate these quantities via maximum-likelihood estimation. However, Lucas (1997) studies some robustness aspects of the Student-t M-estimators using influence functions and shows that the protection against "large" observations is only valid when the degrees of freedom parameter is kept fixed. Therefore, for the purpose of avoiding possible lack of protection against outliers, we do not estimate d_{gi} and f_{gi} by maximum likelihood and instead of it, we kept fixed all quantities involved with them. Otherwise, unboundedness problems may arise for the influence functions and the elliptical distribution will lose its main goal. It is worth emphasizing that, the problem with the influence functions verified by Lucas (1997) is proven only for the Student-t distribution, but it may also happen for other distributions when one estimates d_{gi} and f_{gi} via maximum likelihood (it happens at least with Student-t distribution). This issue is an open problem and needs more attention but it is beyond of the main scope of this paper. In practice, one can use model selection procedures, such as the Akaike information criterion (AIC), to choose the more appropriate values of such unknown parameters.

The Fisher scoring method can be used to estimate $\boldsymbol{\theta}$ by iteratively solving the equation

$$(\boldsymbol{F}^{(m)\top}\boldsymbol{W}^{(m)}\boldsymbol{F}^{(m)})\boldsymbol{\theta}^{(m+1)} = \boldsymbol{F}^{(m)\top}\boldsymbol{W}^{(m)}\boldsymbol{s}^{*(m)}, \quad m = 0, 1, \ldots, \tag{7}$$

where

$$\boldsymbol{W}^{(m)} = \boldsymbol{H}^{(m)}\boldsymbol{M}^{(m)}\boldsymbol{H}^{(m)}, \qquad \boldsymbol{F}^{(m)} = (\boldsymbol{F}_1^{(m)\top}, \boldsymbol{F}_2^{(m)\top}, \ldots, \boldsymbol{F}_n^{(m)\top})^\top,$$

$$\boldsymbol{H}^{(m)} = \text{block–diag}\{\boldsymbol{H}_1^{(m)}, \boldsymbol{H}_2^{(m)}, \ldots, \boldsymbol{H}_n^{(m)}\},$$

$$\boldsymbol{M}^{(m)} = \text{block–diag}\{\boldsymbol{M}_1^{(m)\top}, \boldsymbol{M}_2^{(m)\top}, \ldots, \boldsymbol{M}_n^{(m)\top}\},$$

$$\boldsymbol{s}^{*(m)} = \boldsymbol{F}^{(m)}\boldsymbol{\theta}^{(m)} + \boldsymbol{H}^{-1(m)}\boldsymbol{M}^{-1(m)}\boldsymbol{s}^{(m)}, \qquad \boldsymbol{s}^{(m)} = (\boldsymbol{s}_1^{(m)\top}, \boldsymbol{s}_2^{(m)\top}, \ldots, \boldsymbol{s}_n^{(m)\top})^\top,$$

and m is the iteration counter. Each loop, through the iterative scheme (7), consists of an iterative re-weighted least squares algorithm to optimize the log-likelihood (3). Using equation (7) and any software (for instance, `MAPLE`, `MATLAB`, `Ox`, `R`, `SAS`) with a weighted linear regression routine one can compute the MLE $\widehat{\boldsymbol{\theta}}$ iteratively. The iterations continue until convergence is achieved (a stopping criterion must be defined). Sometimes this iterative algorithm does not converge, neither find the actual maximum of the likelihood function nor a relative maximum point which is an interior point of a restricted parametric space. In these cases, other numerical methods can be used such as the Gauss-Newton and Quasi-Newton methods.

Note that the score function and the Fisher information matrix for $\boldsymbol{\theta}$ can be written as, respectively, $\boldsymbol{U}_{\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{s}$ and $\boldsymbol{K}_{\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{W}\boldsymbol{F}$. We have $v_i = 1$ and $\boldsymbol{M} = \boldsymbol{H}^{-1}$ for the normal model, which implies that $\boldsymbol{W} = \boldsymbol{H}$. Thus, equations (5)-(7) agree with the result due to Patriota and Lemonte (2009).

3 Local influence

The local influence method is recommended when the concern is related to investigate the model sensitivity under some minor perturbations in the model (or data). Let $\boldsymbol{\omega}$ be a k-dimensional vector of perturbations restricted to some open subset $\boldsymbol{\Omega}$ of $\mathbb{R}^k$. The perturbed log-likelihood function is denoted by $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. We consider that exists a no perturbation vector $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ such that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$, for all $\boldsymbol{\theta}$. The influence of minor perturbations on the MLE $\widehat{\boldsymbol{\theta}}$ can be assessed by using the likelihood displacement $LD_{\boldsymbol{\omega}} = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the maximizer of $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$.

The idea for assessing local influence as advocated by Cook (1986) is essentially the analysis of the local behavior of $LD_{\boldsymbol{\omega}}$ around $\boldsymbol{\omega}_0$ by evaluating the curvature of the plot of $LD_{\boldsymbol{\omega}_0 + a\boldsymbol{d}}$ against a, where $a \in \mathbb{R}$ and $\boldsymbol{d}$ is a unit direction. One of the measures of particular interest is the direction $\boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{\boldsymbol{d}_{\max}}$. The index plot of $\boldsymbol{d}_{\max}$ may evidence those observations that have considerable influence on $LD_{\boldsymbol{\omega}}$ under minor perturbations. Also, plots of $\boldsymbol{d}_{\max}$ against covariate values may be helpful for identifying atypical patterns. Cook (1986) showed that the normal curvature at the direction $\boldsymbol{d}$ is given by $C_{\boldsymbol{d}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top\boldsymbol{\Delta}^\top\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\boldsymbol{\Delta}\boldsymbol{d}|$, where $\boldsymbol{\Delta} = \partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$

and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$, both $\boldsymbol{\Delta}$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}$ are evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Moreover, $C_{\boldsymbol{d}_{\max}}$ is twice the largest eigenvalue of $\boldsymbol{B} = -\boldsymbol{\Delta}^\top\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{\Delta}$ and $\boldsymbol{d}_{\max}$ is the corresponding eigenvector. The index plot of $\boldsymbol{d}_{\max}$ may reveal how to perturb the model (or data) to obtain large changes in the estimate of $\boldsymbol{\theta}$.

Assume that the parameter vector $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. The dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are p_1 and $p - p_1$, respectively. Let

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}} = \begin{pmatrix} \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} & \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2} \\ \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}^\top & \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} \end{pmatrix},$$

where $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\top$, $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top$. If the interest lies on $\boldsymbol{\theta}_1$, the normal curvature in the direction of the vector $\boldsymbol{d}$ is $C_{\boldsymbol{d};\boldsymbol{\theta}_1}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top\boldsymbol{\Delta}^\top(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}\boldsymbol{d}|$, where

$$\ddot{\boldsymbol{L}}_{22} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}^{-1} \end{pmatrix}$$

and $\boldsymbol{d}_{\max;\boldsymbol{\theta}_1}$ here is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{B}_1 = -\boldsymbol{\Delta}^\top(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}$ (see Cook, 1986). The index plot of the $\boldsymbol{d}_{\max;\boldsymbol{\theta}_1}$ may reveal those influential elements on $\widehat{\boldsymbol{\theta}}_1$.

In order to have a curvature invariant under a uniform change of scale, Poon and Poon (1999) introduce the conformal normal curvature $B_{\boldsymbol{d}}(\boldsymbol{\theta})$ in the direction of the unit vector $\boldsymbol{d}$, given by

$$B_{\boldsymbol{d}}(\boldsymbol{\theta}) = -\frac{\boldsymbol{d}^\top\boldsymbol{\Delta}^\top\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{\Delta}\boldsymbol{d}}{\sqrt{\mathrm{tr}\{(\boldsymbol{\Delta}^\top\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{\Delta})^2\}}},$$

evaluated at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. An interesting property of the conformal normal curvature is that $0 \leq B_{\boldsymbol{d}}(\boldsymbol{\theta}) \leq 1$. Thus, it can be easily computed once $C_{\boldsymbol{d}}(\boldsymbol{\theta})$ was obtained. This quantity can be seen as a normalized version of $C_{\boldsymbol{d}}(\boldsymbol{\theta})$.

4 Curvature calculations

In the section, we derive the matrix $\boldsymbol{\Delta}$ for different perturbation schemes. These matrices are obtained using results of matrix differentiation (Magnus and Neudecker, 1988). We shall consider the case-weight perturbation, scale matrix and response variable perturbation schemes, that is, we derive (for three perturbation schemes) the matrix

$$\boldsymbol{\Delta} = \{\Delta_{ri}\} = \left\{\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_r\partial\omega_i}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\,\boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad i = 1, 2, \ldots, k \quad \text{and} \quad r = 1, 2, \ldots, p,$$

considering the defined model in (2) and its log-likelihood function given by (3). The quantities distinguished by the addition of "$\widehat{\ }$" are evaluated at $\widehat{\boldsymbol{\theta}}$. The observed information matrix used in the calculation of the normal curvature is given in Appendix B.

4.1 Case weight perturbation

The perturbation of cases is done by attaching some weight to each observation in the log-likelihood resulting in $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\theta})$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^\top$, with $\omega_i \geq 0$, for $i = 1, 2, \dots, n$, and $\boldsymbol{\omega}_0 = \mathbf{1}_n = (1, 1, \dots, 1)^\top$ is the vector of no perturbations. Using matrix differentiation rules along with the notations defined in Section 2, we find

$$\Delta_{ri} = \frac{1}{2}\mathrm{tr}\{\widehat{\boldsymbol{A}}_{i(r)}(\widehat{\boldsymbol{\Sigma}}_i - \widehat{v}_i\widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top)\} + \widehat{v}_i\widehat{\boldsymbol{a}}_{i(r)}^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{z}}_i, \qquad (8)$$

for $r = 1, 2, \dots, p$ and $i = 1, 2, \dots, n$. Here, $\widehat{v}_i = v_i(\widehat{\boldsymbol{\theta}})$ for $i = 1, 2, \dots, n$. In matrix notation, the $p \times n$ matrix $\boldsymbol{\Delta}$ can be written as

$$\boldsymbol{\Delta} = \Big(\widehat{\boldsymbol{F}}_1^\top\widehat{\boldsymbol{H}}_1\widehat{\boldsymbol{s}}_1, \widehat{\boldsymbol{F}}_2^\top\widehat{\boldsymbol{H}}_2\widehat{\boldsymbol{s}}_2, \dots, \widehat{\boldsymbol{F}}_n^\top\widehat{\boldsymbol{H}}_n\widehat{\boldsymbol{s}}_n\Big),$$

where $\boldsymbol{F}_i$, $\boldsymbol{H}_i$ and $\boldsymbol{s}_i$ (for $i = 1, 2, \dots, n$) were defined in Section 2. For normal models, expression (8) reduces to the one derived by Patriota et al. (2010).

4.2 Scale matrix perturbation

The scale matrix perturbation is introduced by considering

$$\boldsymbol{Y}_i \sim \mathcal{E}_{q_i}(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \omega_i^{-1}\boldsymbol{\Sigma}_i(\boldsymbol{\theta})), \quad i = 1, 2, \dots, n,$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^\top \in \mathbb{R}^n - \{\mathbf{0}\}$ and $\boldsymbol{\omega}_0 = \mathbf{1}_n$ such that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$ given in (3). Thus, we have

$$\Delta_{ri} = \{W_g(\widehat{u}_i) + \widehat{u}_i W_g'(\widehat{u}_i)\}\big[\widehat{\boldsymbol{z}}_{i(r)}^\top\widehat{\boldsymbol{A}}_{i(r)}\widehat{\boldsymbol{z}}_i - 2\widehat{\boldsymbol{a}}_{i(r)}^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{z}}_i\big], \qquad (9)$$

for $r = 1, 2, \dots, p$ and $i = 1, 2, \dots, n$. In matrix notation, the $p \times n$ matrix $\boldsymbol{\Delta}$ takes the form

$$\boldsymbol{\Delta} = \Big(\widehat{\boldsymbol{F}}_1^\top\widehat{\boldsymbol{H}}_1\widehat{\boldsymbol{r}}_1, \widehat{\boldsymbol{F}}_2^\top\widehat{\boldsymbol{H}}_2\widehat{\boldsymbol{r}}_2, \dots, \widehat{\boldsymbol{F}}_n^\top\widehat{\boldsymbol{H}}_n\widehat{\boldsymbol{r}}_n\Big),$$

where

$$\widehat{\boldsymbol{r}}_i = -2\{W_g(\widehat{u}_i) + \widehat{u}_i W_g'(\widehat{u}_i)\}\begin{bmatrix}\widehat{\boldsymbol{z}}_i \\ \mathrm{vec}(\widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top)\end{bmatrix}.$$

Expression (9) reduces to the one given by Patriota et al. (2010) for normal models.

4.3 Response perturbation

Here, the response variable $\boldsymbol{Y}_i$ is perturbed according to $\boldsymbol{Y}_i^* = \boldsymbol{Y}_i + \boldsymbol{\omega}_i$, where $\boldsymbol{\omega}_i$ denotes the $q_i \times 1$ perturbation vector and $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \boldsymbol{\omega}_2^\top, \dots, \boldsymbol{\omega}_n^\top)^\top$, so that the no perturbation vector is $\boldsymbol{\omega}_0 = \mathbf{0}$, where $\boldsymbol{\omega} \in \mathbb{R}^N$ ($N = \sum_{i=1}^{n} q_i$). In this case, the perturbed log-likelihood function is also given by $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\boldsymbol{\omega})$, where

$$\ell_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log g(u_{iw}),$$

where $u_{iw} = \boldsymbol{z}_{iw}^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{z}_{iw}$, with $\boldsymbol{z}_{iw} = \boldsymbol{z}_i + \boldsymbol{\omega}_i$. We obtain

$$\begin{aligned}\Delta_{ri} = &-2\widehat{\boldsymbol{a}}_{i(r)}^\top \widehat{\boldsymbol{\Sigma}}_i^{-1}\big\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + 2W_g'(\widehat{u}_i)\widehat{\boldsymbol{z}}_{i(r)}\widehat{\boldsymbol{z}}_{i(r)}^\top\big\}\widehat{\boldsymbol{\Sigma}}_i^{-1} \\ &+ 2\widehat{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{A}}_{i(r)}\big\{W_g(\widehat{u}_i)\widehat{\boldsymbol{\Sigma}}_i + W_g'(\widehat{u}_i)\widehat{\boldsymbol{z}}_{i(r)}\widehat{\boldsymbol{z}}_{i(r)}^\top\big\}\widehat{\boldsymbol{\Sigma}}_i^{-1},\end{aligned} \tag{10}$$

for $r = 1, 2, \ldots, p$ and $i = 1, 2, \ldots, n$. In matrix notation, the $p \times N$ matrix $\boldsymbol{\Delta}$ is given by

$$\boldsymbol{\Delta} = \Big(\widehat{\boldsymbol{F}}_1^\top \widehat{\boldsymbol{H}}_1 \widehat{\boldsymbol{G}}_1, \widehat{\boldsymbol{F}}_2^\top \widehat{\boldsymbol{H}}_2 \widehat{\boldsymbol{G}}_2, \ldots, \widehat{\boldsymbol{F}}_n^\top \widehat{\boldsymbol{H}}_n \widehat{\boldsymbol{G}}_n\Big),$$

where

$$\widehat{\boldsymbol{G}}_i = -2\begin{pmatrix} W_g(\widehat{u}_i)\boldsymbol{I}_{q_i} + 2W_g'(\widehat{u}_i)\widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1} \\ 2\widehat{\boldsymbol{z}}_i \otimes \big[W_g(\widehat{u}_i)\boldsymbol{I}_{q_i} + W_g'(\widehat{u}_i)\widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\big]\end{pmatrix}.$$

For normal models the expression (10) reduces to the one given by Patriota et al. (2010).

5 Generalized leverage

Let $\boldsymbol{Y} = \mathrm{vec}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$ and $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathrm{vec}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n)$. In what follows we shall use the generalized leverage proposed by Wei et al. (1998). The authors have shown that the generalized leverage is obtained by evaluating the $N \times N$ matrix

$$\boldsymbol{GL}(\boldsymbol{\theta}) = \boldsymbol{D}_{\boldsymbol{\theta}}(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}},$$

at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, where $\boldsymbol{D}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{Y}^\top$. As noted by the authors, the generalized leverage is invariant under reparameterization and observations with large GL_{ij} are leverage points. The main idea behind the concept of leverage is that of evaluating the influence of $\boldsymbol{Y}_i$ on its own predicted value.

Under the model defined in (2), we have that

$$\boldsymbol{D}_{\boldsymbol{\theta}} = (\boldsymbol{D}_1^\top, \boldsymbol{D}_2^\top, \ldots, \boldsymbol{D}_n^\top)^\top$$

and

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}} = \Big(\boldsymbol{F}_1^\top \boldsymbol{H}_1 \boldsymbol{G}_1, \boldsymbol{F}_2^\top \boldsymbol{H}_2 \boldsymbol{G}_2, \ldots, \boldsymbol{F}_n^\top \boldsymbol{H}_n \boldsymbol{G}_n\Big).$$

Index plots of GL_{ii} may reveal those observations with high influence on their own predicted values.

6 Special model

In order to illustrate the usefulness and applicability of the proposed formulation, we consider a general elliptical mixed-effects model with nonlinear mixed effects and some covariates subject to measurement error. Here, the equation of interest is

$$\boldsymbol{z}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{x}_i + \boldsymbol{f}(\boldsymbol{l}_i, \boldsymbol{\alpha}) + \boldsymbol{W}_i\boldsymbol{b}_i + \boldsymbol{q}_i, \quad i = 1, 2, \ldots, n, \tag{11}$$

where $\boldsymbol{z}_i$ is a $v \times 1$ latent response vector, $\boldsymbol{x}_i$ is a $m \times 1$ latent vector of covariates, $\boldsymbol{\beta}_0$ is a $v \times 1$ vector of intercepts, $\boldsymbol{\beta}_1$ is a $v \times m$ matrix which elements are inclinations, $\boldsymbol{f}(\boldsymbol{l}_i, \boldsymbol{\alpha})$ is a v-dimensional nonlinear function of $\boldsymbol{\alpha}$, $\boldsymbol{l}_i$ is a vector of explanatory known covariates, $\boldsymbol{W}_i$ is a $v \times r$ matrix of known constants, $\boldsymbol{b}_i$ is a $r \times 1$ vector of unobserved random coefficients (random effects of the model) and $\boldsymbol{q}_i$ is the equation error. Model (11) is an errors-in-variables nonlinear mixed model which generalizes the one considered in Russo et al. (2009). In model (11) we cannot observe directly the variables $\boldsymbol{z}_i$ and $\boldsymbol{x}_i$, instead we observe $\boldsymbol{Z}_i$ and $\boldsymbol{X}_i$, respectively, with the following additive relationship

$$\boldsymbol{Z}_i = \boldsymbol{z}_i + \boldsymbol{e}_i \qquad \text{and} \qquad \boldsymbol{X}_i = \boldsymbol{x}_i + \boldsymbol{u}_i,$$

where $\boldsymbol{e}_i$ and $\boldsymbol{u}_i$ are measurement errors. We consider that the vector of full random vector $\boldsymbol{r}_i = \big((\boldsymbol{x}_i - \boldsymbol{\mu}_x)^\top, \boldsymbol{b}_i^\top, \boldsymbol{q}_i^\top, \boldsymbol{e}_i^\top, \boldsymbol{u}_i^\top\big)^\top$ follows the following elliptical distribution

$$\boldsymbol{r}_i \sim \mathcal{E}_d(\boldsymbol{0}, \boldsymbol{\Omega}_i),$$

where $d = 2v + 2m + r$ and $\boldsymbol{\Omega}_i = \text{block-diag}\{\boldsymbol{R}_x(\boldsymbol{\sigma}_1), \boldsymbol{R}_b(\boldsymbol{\sigma}_2), \boldsymbol{R}_q(\boldsymbol{\sigma}_3), \boldsymbol{\tau}_{ei}, \boldsymbol{\tau}_{ui}\}$, with $\boldsymbol{\tau}_{ei}$ and $\boldsymbol{\tau}_{ui}$ the variances of the measurement errors assumed to be known for all $i = 1, \ldots, n$. These "known" matrices may be attained, for example, through an analytical treatment of the data collection mechanism, replications, machine precision, etc. Here, we consider that the matrices $\boldsymbol{R}_x = \boldsymbol{R}_x(\boldsymbol{\sigma}_1)$, $\boldsymbol{R}_b = \boldsymbol{R}_b(\boldsymbol{\sigma}_2)$ and $\boldsymbol{R}_q = \boldsymbol{R}_q(\boldsymbol{\sigma}_3)$ are completely specified by the vectors of parameters $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ and $\boldsymbol{\sigma}_3$, respectively. Therefore, the marginal distribution for the observable vector $\boldsymbol{Y}_i = (\boldsymbol{Z}_i^\top, \boldsymbol{X}_i^\top)^\top$ is

$$\boldsymbol{Y}_i \sim \mathcal{E}_d(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta})),$$

where

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mu}_{zi} \\ \boldsymbol{\mu}_x \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{zi} + \boldsymbol{\tau}_{zi} & \boldsymbol{\beta}_1 \boldsymbol{R}_x \\ \boldsymbol{R}_x \boldsymbol{\beta}_1^\top & \boldsymbol{R}_x + \boldsymbol{\tau}_{xi} \end{pmatrix},$$

with $\boldsymbol{\mu}_{zi} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{\mu}_x + \boldsymbol{f}(\boldsymbol{l}_i, \boldsymbol{\alpha})$, $\boldsymbol{\Sigma}_{zi} = \boldsymbol{\beta}_1 \boldsymbol{R}_x \boldsymbol{\beta}_1^\top + \boldsymbol{W}_i \boldsymbol{R}_b \boldsymbol{W}_i^\top + \boldsymbol{R}_q$. In this model, $\boldsymbol{\theta} = (\boldsymbol{\beta}_0^\top, \text{vec}(\boldsymbol{\beta}_1)^\top, \boldsymbol{\mu}_x^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\sigma}_1^\top, \boldsymbol{\sigma}_2^\top, \boldsymbol{\sigma}_3^\top)^\top$. Note that the mean vector and the covariance-variance matrix of observed variables have the matrix $\boldsymbol{\beta}_1$ in common, i.e. they share mv parameters. Kulathinal et al. (2002) study the linear univariate case under normality (i.e. $v = 1$, $m = 1$, $\boldsymbol{f} = \boldsymbol{0}$ and $\boldsymbol{D} = \boldsymbol{0}$).

Notice that the matrix $\boldsymbol{F}_i$ is the only thing that we have to find for computing all quantities presented in this paper. For this special model, it is given by

$$\boldsymbol{F}_i = \begin{pmatrix} \boldsymbol{D}_i^{(1)} & \boldsymbol{D}_i^{(2)} & \boldsymbol{D}_i^{(3)} & \boldsymbol{D}_i^{(4)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_i^{(2)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{V}_i^{(5)} & \boldsymbol{V}_i^{(6)} & \boldsymbol{V}_i^{(7)} \end{pmatrix},$$

where $\boldsymbol{D}_i^{(1)} = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}_0^\top$, $\boldsymbol{D}_i^{(2)} = \partial \boldsymbol{\mu}_i / \partial \text{vec}(\boldsymbol{\beta}_1)^\top$, $\boldsymbol{D}_i^{(3)} = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\mu}_x^\top$, $\boldsymbol{D}_i^{(4)} = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\alpha}^\top$, $\boldsymbol{V}_i^{(2)} = \partial \text{vec}(\boldsymbol{\Sigma}_i) / \partial \text{vec}(\boldsymbol{\beta}_1)^\top$, $\boldsymbol{V}_i^{(5)} = \partial \text{vec}(\boldsymbol{\Sigma}_i) / \partial \boldsymbol{\sigma}_1)^\top$, $\boldsymbol{V}_i^{(6)} = \partial \text{vec}(\boldsymbol{\Sigma}_i) / \partial \boldsymbol{\sigma}_2^\top$ and $\boldsymbol{V}_i^{(7)} = \partial \text{vec}(\boldsymbol{\Sigma}_i) / \partial \boldsymbol{\sigma}_3^\top$. As a special case of the model above we have the nonlinear mixed model considered by Russo et al. (2009), which emerges by taking $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\sigma}_2^\top, \sigma_3^2)^\top$, $\boldsymbol{Y}_i = \boldsymbol{Z}_i$, $\boldsymbol{\mu}_i = \boldsymbol{f}(\boldsymbol{l}_i, \boldsymbol{\alpha})$, $\boldsymbol{\Sigma}_i = \boldsymbol{W}_i \boldsymbol{D} \boldsymbol{W}_i^\top + \sigma_3^2 \boldsymbol{I}_{q_i}$ and the matrix $\boldsymbol{F}_i$ becomes

$$\boldsymbol{F}_i = \begin{pmatrix} \boldsymbol{D}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_i \end{pmatrix},$$

where $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\alpha}^\top$ and $\boldsymbol{V}_i = \partial \mathrm{vec}(\boldsymbol{\Sigma}_i)/\partial \boldsymbol{\gamma}^\top$ with $\boldsymbol{\gamma} = (\boldsymbol{\sigma}_2, \sigma_3^2)^\top$. Other special models are nonlinear heteroscedastic models, nonlinear model with a first-order autoregressive covariance matrix to the error terms, heteroscedastic multivariate errors-in-variables models, among several others.

As can be seen, several important models can be adjusted just by appropriately defining the vector of parameters $\boldsymbol{\theta}$, the location function $\boldsymbol{\mu}_i(\boldsymbol{\theta})$, the dispersion function $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ and the model-specification matrix $\boldsymbol{F}_i$. With this, all the quantities derived in this paper become available.

7 Application

In this section, for illustrative purposes, we analyze the radioimmunoassay data, reported in Tiede and Pagano (1979), which were obtained from the Nuclear Medicine Department of the Veteran's Administration Hospital, Buffalo, New York. All the computations were done using the `Ox` matrix programming language (Doornik, 2006). `Ox` is freely distributed for academic purposes and available at http://www.doornik.com.

Following Tiede and Pagano (1979) we shall consider the nonlinear regression model

$$y_i = \theta_1 + \frac{\theta_2}{1 + \theta_3 x_i^{\theta_4}} + u_i, \qquad i = 1, 2, \ldots, 14,$$

where the response variable is the observed count, the covariate corresponds to the dose (measured in micro-international units per milliliter) and the errors follow an appropriate elliptical distribution. According to Tiede and Pagano (1979), the model above yields parameters which have physical interpretations, i.e. the estimate of θ_1 is an estimate of the background counts or noise. The zero dose count is estimated by the estimate of θ_2 and the midrange of the assay, also referred to as the effective dose for 50% response is estimated by the estimate of θ_4. The estimated value of this parameter, which is in the neighborhood of 1.0, provides an indication of the sharpness of the bend in the curve. For further details the reader is referred to Tiede and Pagano (1979).

Maximum likelihood estimates of the model parameters for the normal and Student-t (with $\nu = 4$) models are presented in Table 2 as well as the corresponding approximate standard errors. We have considered $\nu = 4$ for the Student-t model for modeling the current data based on the arguments given in Lange et al. (1989, Example 2). Additionally, the scale parameter is assumed to be known for both models. From Table 2 all the parameters seem to be highly significant for the adopted models. Figure 1 gives the scatter-plot of the data, together with the fitted curves of the normal and Student-t models. As can be seen from this figure, the t model fits satisfactorily to the radioimmunoassay data.

Table 2: Maximum likelihood estimates and standard errors in parentheses.

	Model	
Parameter	Normal	Student-t
θ_1	444.8649 (1.4326)	929.2840 (0.9075)
θ_2	7549.7624 (1.7093)	6881.7149 (1.3252)
θ_3	0.1329 (0.0001)	0.0781 (0.0001)
θ_4	0.9583 (0.0004)	1.3562 (0.0006)

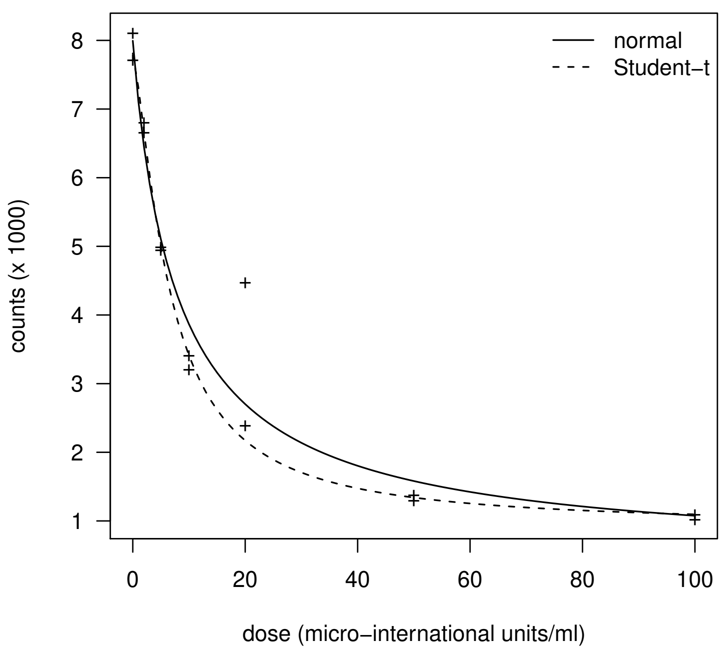

Figure 1: Scatter-plot and the fitted models.

In what follows we shall apply the local influence method developed in the previous sections for the purpose of identifying influential observations in the normal and Student-t regression models fitted to the data. In order to consider the general results derived before, we define $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^\top$ and $\mu_i(\boldsymbol{\theta}) = \theta_1 + \theta_2/(1 + \theta_3 x_i^{\theta_4})$, for $i = 1, 2, \ldots, 14$. We have that

$$\boldsymbol{F}_i = \begin{pmatrix} 1 & \dfrac{1}{1+\theta_3 x_i^{\theta_4}} & -\dfrac{\theta_2 x_i^{\theta_4}}{(1+\theta_3 x_i^{\theta_4})^2} & -\dfrac{\theta_2\theta_3 x_i^{\theta_4}\log(x_i)}{(1+\theta_3 x_i^{\theta_4})^2} \end{pmatrix}$$

and hence the quantities derived in this paper become available. Figure 2 presents the index plots of $|\boldsymbol{d}_{\max}|$ for the maximum likelihood estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$. From this figure we can notice that case #9 appears as the most influential for the normal model. However, for the Student-t model the influence of case #9 reduces substantially, even though some points appear with moderate influence. Based

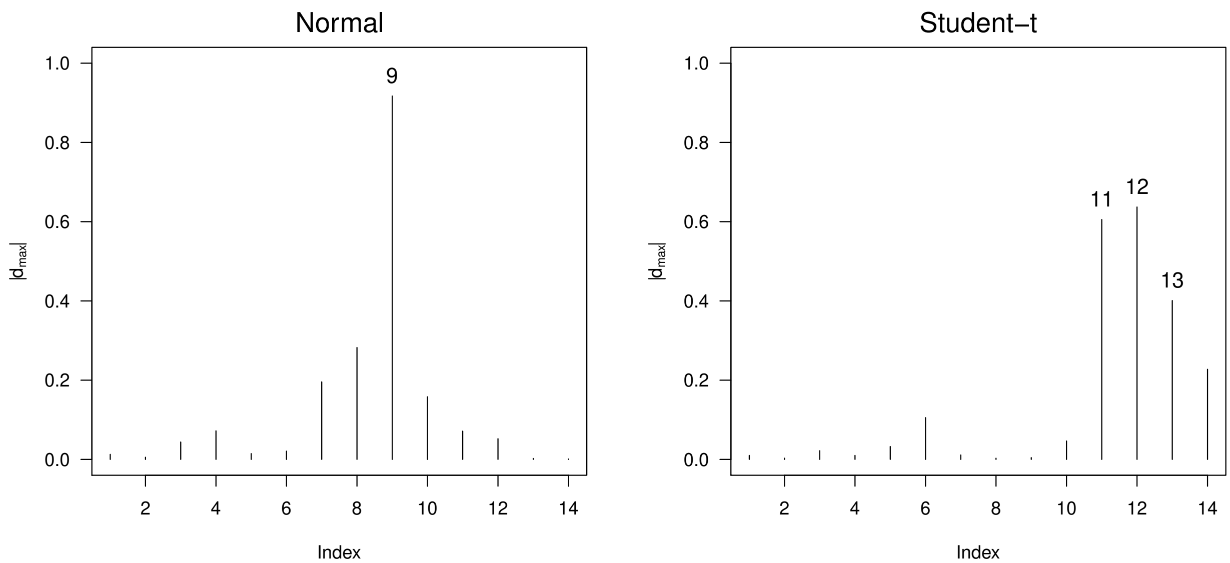

Figure 2: Index plots of $|\boldsymbol{d}_{\max}|$ for $\widehat{\boldsymbol{\theta}}$.

on Figure 2, we eliminated those most influential observations and refitted the normal and Student-t models. In Table 3 we have the relative changes of each parameter estimate, defined by RC $= |(\widehat{\theta}_j - \widehat{\theta}_{j(i)})/\widehat{\theta}_j| \times 100\%$, where $\widehat{\theta}_{j(i)}$ denotes the maximum likelihood estimate of θ_j, after removing the ith observation. It should be noticed from Table 3 that the relative changes for the maximum likelihood estimates of the parameters of the Student-t model are very little pronounced. On the other hand, the maximum likelihood estimates of the parameters of the normal model are extremely affected by the indicated cases, mainly by the case #9. According to Tiede and Pagano (1979) this case is an outlier. Therefore, this table confirms the robustness of the Student-t model against the extreme values.

Table 3: Relative changes (%) dropping the cases indicated.

	Normal				Student-t			
Dropping	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_4$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_4$
#9	107.51	7.12	35.50	38.86	0.00	0.00	0.04	0.00
#11	15.95	0.88	3.02	0.76	0.00	0.01	0.09	0.02
#12	11.65	0.65	2.20	0.56	0.00	0.01	0.00	0.01
#13	15.19	0.91	0.35	1.23	0.01	0.00	0.04	0.00

8 Concluding remarks

In this paper, we introduce a multivariate elliptical model with general parameterization which unifies several important models (e.g., (non)linear regressions models, (non)linear mixed models, errors-in-variables models, and so forth). We also consider diagnostic techniques that can be employed to identify influential observations. Appropriate matrices for assessing local influence on the parameter estimates under different perturbation schemes are obtained. Although the complexity of the postulated model, the expressions derived in this paper are simple, compact and can be easily implemented into any mathematical or statistical/econometric programming environment with numerical linear algebra facilities, such as `R` (R Development Core Team, 2009) and `Ox` (Doornik, 2006), among others, i.e. our formulas related with this class of models are manageable, and with the use of modern computer resources, may turn into adequate tools comprising the arsenal of applied statisticians. Finally, an empirical application to a real data set is presented.

Acknowledgments

We thank two anonymous referees and an Associate Editor for helpful suggestions. We gratefully acknowledge grants from FAPESP (Brazil).

A Appendix

In what follows, we shall derive the Fisher information matrix presented in Section 2, equation (6). As we are considering a function g with regular properties (differentiation and integration are interchangeable), we have that $\mathbb{E}(\boldsymbol{s}_i) = \boldsymbol{0}$ and the Fisher information for $\boldsymbol{\theta}$ is obtained as

$$\mathbb{E}\big(\boldsymbol{U}_{\boldsymbol{\theta}}\boldsymbol{U}_{\boldsymbol{\theta}}^{\top}\big) = \sum_{i=1}^{n} \boldsymbol{F}_i^{\top}\boldsymbol{H}_i\mathbb{E}\big(\boldsymbol{s}_i\boldsymbol{s}_i^{\top}\big)\boldsymbol{H}_i\boldsymbol{F}_i.$$

Note that

$$\mathbb{E}\big(\boldsymbol{s}_i\boldsymbol{s}_i^\top\big) = \begin{bmatrix} \mathbb{E}\big(v_i^2\boldsymbol{z}_i\boldsymbol{z}_i^\top\big) & -\mathbb{E}\big(v_i\boldsymbol{z}_i\text{vec}(\boldsymbol{\Sigma}_i - v_i\boldsymbol{z}_i\boldsymbol{z}_i^\top)^\top\big) \\ -\mathbb{E}\big(v_i\text{vec}(\boldsymbol{\Sigma}_i - v_i\boldsymbol{z}_i\boldsymbol{z}_i^\top)\boldsymbol{z}_i^\top\big) & \mathbb{E}\big(\text{vec}(\boldsymbol{\Sigma}_i - v_i\boldsymbol{z}_i\boldsymbol{z}_i^\top)\text{vec}(\boldsymbol{\Sigma}_i - v_i\boldsymbol{z}_i\boldsymbol{z}_i^\top)^\top\big) \end{bmatrix}.$$

Thus, by adapting the results of Mitchell (1989) for a matrix version, we have the following expectations:

1. $\mathbb{E}\big(v_i\boldsymbol{z}_i\big) = \boldsymbol{0}$,
2. $\mathbb{E}\big(v_i\boldsymbol{z}_i\boldsymbol{z}_i^\top\big) = \boldsymbol{\Sigma}_i$,
3. $\mathbb{E}\big(v_i^2\boldsymbol{z}_i\boldsymbol{z}_i^\top\big) = \dfrac{4d_{gi}}{q_i}\boldsymbol{\Sigma}_i$,
4. $\mathbb{E}\big(v_i^2\text{vec}(\boldsymbol{z}_i\boldsymbol{z}_i^\top)\boldsymbol{z}_i^\top\big) = \boldsymbol{0}$,
5. $\mathbb{E}\big(v_i^2\text{vec}(\boldsymbol{z}_i\boldsymbol{z}_i^\top)\text{vec}(\boldsymbol{z}_i\boldsymbol{z}_i^\top)^\top\big) = \dfrac{4f_{gi}}{q_i(q_i+2)}\Big(\text{vec}(\boldsymbol{\Sigma}_i)\text{vec}(\boldsymbol{\Sigma}_i)^\top + \boldsymbol{\Sigma}_i\otimes\boldsymbol{\Sigma}_i + \boldsymbol{P}_i(\boldsymbol{\Sigma}_i\otimes\boldsymbol{\Sigma}_i)\Big)$, where $\boldsymbol{P}_i$ is a commutation maltrix such that $\text{vec}(\boldsymbol{A}) = \boldsymbol{P}_i\text{vec}(\boldsymbol{A}^\top)$ for any matrix $\boldsymbol{A}$ with appropriated dimensions.

where $\boldsymbol{z}_i \sim \mathcal{E}_{q_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$. Therefore, the main result follows by the properties of the commutation matrix $\boldsymbol{P}_i$.

B Appendix

The observed information matrix for $\boldsymbol{\theta}$ is given by $-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}$, where, after some algebraic manipulation,

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}} = \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \sum_{i=1}^n\Big\{\boldsymbol{F}_i^\top\boldsymbol{H}_i\ddot{\boldsymbol{M}}_i\boldsymbol{H}_i\boldsymbol{F}_i + \big[\boldsymbol{s}_i^\top\boldsymbol{H}_i\big]\Big[\frac{\partial\boldsymbol{F}_i}{\partial\boldsymbol{\theta}}\Big]\Big\},$$

with

$$\ddot{\boldsymbol{M}}_i = 2W_g(u_i)\begin{bmatrix} \boldsymbol{\Sigma}_i & 2\boldsymbol{z}_i^\top\otimes\boldsymbol{\Sigma}_i \\ 2\boldsymbol{\Sigma}_i\otimes\boldsymbol{z}_i & 2(\boldsymbol{\Sigma}_i\otimes(\boldsymbol{z}_i\boldsymbol{z}_i^\top) + (\boldsymbol{z}_i\boldsymbol{z}_i^\top)\otimes\boldsymbol{\Sigma}_i) \end{bmatrix} + 4W_g'(u_i)\begin{bmatrix} \boldsymbol{z}_i\boldsymbol{z}_i^\top & \boldsymbol{z}_i^\top\otimes(\boldsymbol{z}_i\boldsymbol{z}_i^\top) \\ (\boldsymbol{z}_i\boldsymbol{z}_i^\top)\otimes\boldsymbol{z}_i & \text{vec}(\boldsymbol{z}_i\boldsymbol{z}_i^\top)\text{vec}(\boldsymbol{z}_i\boldsymbol{z}_i^\top)^\top \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & 2\boldsymbol{\Sigma}_i\otimes\boldsymbol{\Sigma}_i \end{bmatrix}.$$

Note that $\partial\boldsymbol{F}_i/\partial\boldsymbol{\theta}$ is an array of dimension $q_i(q_i+1)\times p\times p$. Here, $[\cdot][\cdot]$ represents the bracket product of a matrix by an array as defined by Wei (1998, p. 188).

References

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.

Cheng, C.L., Riu, J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics*. **48**:511–519.

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

Cox, D.R., Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cysneiros, F.J.A., Paula, G.A. (2004). One-sided test in linear models with multivariate t-distribution. *Communication in Statistics – Theory and Methods*. **33**:747–771.

de Castro, M., Galea, M., Bolfarine, H. (2008). Hypothesis testing in an errors-in-variables model with heteroscedastic measurement errors. *Statistics in Medicine*. **27**:5217–5234.

Díaz–García, J.A., Galea, M., Leiva, V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communication in Statistics – Theory and Methods*. **32**:625–641.

Díaz–García, J.A., Gutiérrez–Jáimez, R. (2007). The distribution of residuals from a general elliptical distribution. *Journal of Statistics Planning and Inference*. **137**:2347–2354.

Doornik, J.A. (2006). *An Object-Oriented Matrix Language – Ox 4*, 5th ed. Timberlake Consultants Press, London.

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

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

Galea, M., Riquelme, M., Paula, G.A. (2000). Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics*. **14**:167–184.

Gómez, E., Gómez–Villegas, M.A., Martín, J.M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics, Theory and Methods*. **27**:589–600.

Kelly, B.C. (2007). Some aspects of measurement error in linear regression of astronomical data. *The Astrophysical Journal*. **665**:1489–1506.

Kulathinal, S.B., Kuulasmaa, K., Gasbarra, D. (2002). Estimation of an errors-in-variables regression model when the variances of the measurement error vary between the observations. *Statistics in Medicine*. **21**:1089–1101.

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

Lange, K.L., Little, R.J.A., Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the Americam Statistical Association*. **84**:881–896.

Lee, S.Y., Lu, B., Song, X.Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics and Data Analysis*. **50**:1356–1377.

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 Application*. **354**:159–174.

Lucas, A. (1997). Robustness of the student t based M-estimator. *Communications in Statistics, Theory and Methods*. **26**:1165–1182.

Magnus, J.R., Neudecker, H. (1988). *Matrix differential calculus with applications in Statistics and Econometrics*. Wiley: New York.

Mitchell, A.F.S. (1989). The information matrix, skewness tensor and α-connections for the general multivariate elliptical distribution. *Annals of the Institute of Statistical Mathematics*. **41**:289–304.

Osorio, F., Paula, G.A., Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*. **51**:4354–4368.

Patriota, A.G., Bolfarine, H., de Castro, M. (2009). A hetroscedastic errors-in-variables model with equation error. *Statistical Methodology*. **6**:408–423.

Patriota, A.G., Lemonte, A.J. (2009). Bias correction in a multivariate regression model with genereal parameterization. *Statistics and Probability Letters*. **79**:1655–1662.

Patriota, A.G., Lemonte, A.J., de Castro, M. (2010). Influence diagnostics in a multivariate normal regression model with general parameterization. *Statistical Methodology*. **7**:644–654.

Poon, W.Y., Poon, Y.S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*. **61**:51–61.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Russo, C.M., Paula, G.A., Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis* **53**, 4143–4156.

Savalli, C., Paula, G.A., Cysneiros, F.J.A. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modeling*. **6**:59–76.

Tiede, J.J., Pagano, M. (1979). Application of robust calibration to radioimmunoassay. *Biometrics*. **35**:567–574.

Wei, B.C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer.

Wei, B.C., Hu, Y.Q., Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*. **25**:25–37.

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