Influence diagnostics in a multivariate normal regression model with general parameterization

Alexandre G. Patriota[a,*], Artur J. Lemonte[a], Mário de Castro[b]

[a] Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil
[b] Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, Caixa Postal 668, 13560-970, São Carlos-SP, Brazil

Abstract

Patriota and Lemonte (2009) introduced a quite general multivariate normal regression model. This model considers that the mean vector and the covariance matrix share the same vector of parameters. In this paper we present some influence assessment for this model, such as the local influence, total local influence of an individual and generalized leverage are discussed. Additionally, the normal curvatures for local influence studies are derived under some perturbation schemes.

Key words: Influence diagnostic, Generalized leverage, Likelihood displacement, Local influence, Multivariate models.

1. Introduction

It is nowadays a well spread practice, after modeling, to check the model assumptions and conduct diagnostic studies in order to detect possible influential observations that may distort the results of the analysis. Diagnostic analysis is an efficient way to detect influential observations. The first technique developed to assess the individual impact of cases on the estimation process is, perhaps, the case deletion which became a very popular tool. Cook (1977) presents a great development of case deletion diagnostics for a general statistical model. Case deletion is an example of a global influence analysis, that is, the effect of an observation is assessed by completely removing it.

However, case deletion excludes all information from an observation and we can hardly say whether this observation has some influence on a specific aspect of the model. To overcome this problem, one can resort to local influence approach where one investigates the model sensitivity under small perturbations. In this context, Cook (1986) proposes a general framework to detect influential observations which gives a measure of this sensitivity under small perturbations on the data or in the model. Several authors have extended the local influence method to various regression models; see, for example, Beckman *et al.* (1987), Lawrance (1988), Thomas and Cook (1990), Paula (1993), Lesaffre and Verbeke (1998) and more recently Osorio *et al.* (2007), Carrasco *et al.* (2008), Espinheira *et al.* (2008), Paula *et al.* (2009), Vasconcellos and Fernandez (2009), Russo *et al.* (2009), Patriota (2010), among others. In particular, Vasconcellos and Fernandez (2009) discuss the problem of influence analysis for the situation in which the observations must satisfy a set of homogeneous linear restrictions.

In this paper we focus on influence diagnostics based on case deletion (Cook, 1977) and also develop local influence diagnostics (Cook, 1986) based on minor perturbations in the data and in the postulated model. The underlying model is a multivariate normal regression model with the mean vector and the covariance matrix indexed by the same vector of parameters (Patriota and Lemonte, 2009). It includes many of the

*Corresponding author
Email addresses: patriota.alexandre@gmail.com (Alexandre G. Patriota)

existing regression models; for instance, mixed models, nonlinear regression models and errors-in-variables models. Additionally, an expression for the generalized leverage in this general model is derived.

The paper is organized as follows. Section 2 presents the model. In Section 3 we use several diagnostic measures considering case deletion and the normal curvatures of local influence are derived under various perturbation schemes. In Section 4, we present some useful examples of the proposed formulation. Finally, some concluding remarks are made in Section 5.

2. Model and inference

Throughout this paper we consider the situation in which independent $q_i \times 1$ observable random vectors $\boldsymbol{Y}_i$ follow a multivariate normal distribution denoted by

$$\boldsymbol{Y}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}\big(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta})\big), \quad i = 1, \ldots, n, \tag{1}$$

where "$\overset{\text{ind}}{\sim}$" means "independently distributed as" and $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{z}_i)$ are the mean vector and the covariance matrix, respectively. Also, $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are $m_i \times 1$ and $t_i \times 1$ vectors, respectively, of nonstochastic variables associated with the i^{th} observed response $\boldsymbol{Y}_i$. Notice that $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ may have common components. Additionally, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters. The functional forms of $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are known and twice continuously differentiable with respect to each element of $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ must be identifiable in model (1), we suppose that the model fulfills this requirement. Note that, if the location vector and the scale matrix have just a few parameters in common, say $\boldsymbol{\mu}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3)$, we can always write $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \boldsymbol{\theta}_3^\top)$ and then $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ may represent $\boldsymbol{\mu}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3)$, respectively, without lost of generality. Therefore, we can just consider $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to have the more general setting. In other words, to avoid excess of notation, all model parameters can be entirely described only by $\boldsymbol{\theta}$. Lange et al. (1989) introduce a very general Student-t model where the location vector and the scale matrix do not share parameters. This model is roboust against outliers. Our model is a particular case of this model when $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_i(\boldsymbol{\gamma})$ for all $i = 1, \ldots, n$.

The class of models presented in (1) is quite broad and includes many 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 (see, for instance, Kulathinal *et al.*, 2002; Cheng and Riu, 2006; Kelly, 2007; de Castro *et al.*, 2008; Patriota *et al.*, 2009). These models can also be formulated as in (1). 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 (1). Simultaneous equations models (e.g., Zhao and Lee, 1998; Magnus and Neudecker, 2007, Ch. 16) comprise endogenous and exogenous variables and, in the reduced form, they are a particular case of the general model in (1). As can be seen, model (1) encompasses a wide range of models and our list of examples is by no means exhaustive. Section 4 presents some important special cases.

To simplify notation, let $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{x}_i)$, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{z}_i)$ and $\boldsymbol{u}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$. The log-likelihood function associated with (1), apart from an unimportant constant, is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{2}$$

where $\ell_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\boldsymbol{u}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{u}_i$. We make some assumptions (Cox and Hinkley, 1974, Ch. 9) on the behavior of $\ell(\boldsymbol{\theta})$ as $n \to \infty$, such as the regularity of the first two derivatives of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and the uniqueness of the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$.

We define the quantities

$$\boldsymbol{a}_{i(r)} = \frac{\partial \boldsymbol{\mu}_i}{\partial \theta_r}, \quad \boldsymbol{a}_{i(sr)} = \frac{\partial^2 \boldsymbol{\mu}_i}{\partial \theta_s \partial \theta_r}, \quad \boldsymbol{C}_{i(r)} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \theta_r}, \quad \boldsymbol{C}_{i(sr)} = \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \theta_s \partial \theta_r} \quad \text{and} \quad \boldsymbol{A}_{i(r)} = -\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}_{i(r)}\boldsymbol{\Sigma}_i^{-1},$$

for $r, s = 1, \ldots, p$. To compute the derivatives of $\ell(\boldsymbol{\theta})$ we make use of matrix differentiation methods (Magnus and Neudecker, 2007).

The score function for $\boldsymbol{\theta}$ is

$$\boldsymbol{U}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{u}_i - \frac{1}{2}\sum_{i=1}^{n} \boldsymbol{V}_i^\top(\boldsymbol{\Sigma}_i^{-1}\otimes\boldsymbol{\Sigma}_i^{-1})\mathrm{vec}(\boldsymbol{\Sigma}_i - \boldsymbol{u}_i\boldsymbol{u}_i^\top),$$

where $\boldsymbol{D}_i = (\boldsymbol{a}_{i(1)},\ldots,\boldsymbol{a}_{i(p)})$ and $\boldsymbol{V}_i = (\mathrm{vec}(\boldsymbol{C}_{i(1)}),\ldots,\mathrm{vec}(\boldsymbol{C}_{i(p)}))$. 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. Let

$$\boldsymbol{F}_i = \begin{pmatrix}\boldsymbol{D}_i\\ \boldsymbol{V}_i\end{pmatrix}, \quad \boldsymbol{H}_i = \begin{bmatrix}\boldsymbol{\Sigma}_i^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{1}{2}\boldsymbol{\Sigma}_i^{-1}\otimes\boldsymbol{\Sigma}_i^{-1}\end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_i = \begin{pmatrix}\boldsymbol{u}_i\\ -\mathrm{vec}(\boldsymbol{\Sigma}_i - \boldsymbol{u}_i\boldsymbol{u}_i^\top)\end{pmatrix}, \tag{3}$$

where $(\boldsymbol{F}_1^\top,\ldots,\boldsymbol{F}_n^\top)^\top$ has rank p. Then, $\boldsymbol{U}_{\boldsymbol{\theta}}$ can be written as

$$\boldsymbol{U}_{\boldsymbol{\theta}} = \sum_{i=1}^{n}\boldsymbol{F}_i^\top\boldsymbol{H}_i\boldsymbol{v}_i. \tag{4}$$

The observed information matrix of $\boldsymbol{\theta}$ is given by $-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$, which is obtained from

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \sum_{i=1}^{n}\big(\boldsymbol{G}_i - \boldsymbol{F}_i^\top\boldsymbol{H}_i\boldsymbol{M}_i\boldsymbol{H}_i\boldsymbol{F}_i\big),$$

where $\boldsymbol{G}_i$ is a $p\times p$ matrix with (s,r) element given by $(\boldsymbol{a}_{i(sr)}^\top \mathrm{vec}(\boldsymbol{C}_{i(sr)})^\top)\boldsymbol{H}_i\boldsymbol{v}_i$ and

$$\boldsymbol{M}_i = \begin{bmatrix}\boldsymbol{\Sigma}_i & 2\boldsymbol{\Sigma}_i\otimes\boldsymbol{u}_i^\top\\ 2\boldsymbol{\Sigma}_i\otimes\boldsymbol{u}_i & \boldsymbol{M}_{i;22}\end{bmatrix},$$

with $\boldsymbol{M}_{i;22} = 2\{(\boldsymbol{u}_i\boldsymbol{u}_i^\top)\otimes\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i\otimes(\boldsymbol{u}_i\boldsymbol{u}_i^\top) - \boldsymbol{\Sigma}_i\otimes\boldsymbol{\Sigma}_i\}$. Additionally, noting that $\mathrm{E}(\boldsymbol{M}_i) = \boldsymbol{H}_i^{-1}$ and $\mathrm{E}(\boldsymbol{G}_i) = \boldsymbol{0}$, the expected information matrix of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{K}_{\boldsymbol{\theta}} = -\mathrm{E}(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}) = \sum_{i=1}^{n}\boldsymbol{F}_i^\top\boldsymbol{H}_i\boldsymbol{F}_i.$$

The MLE $\widehat{\boldsymbol{\theta}}$ satisfies the equation $\boldsymbol{U}_{\boldsymbol{\theta}} = \boldsymbol{0}$. The Fisher scoring method can be used to estimate $\boldsymbol{\theta}$ by iteratively solving the equation

$$(\boldsymbol{F}^{(m)\top}\boldsymbol{H}^{(m)}\boldsymbol{F}^{(m)})\boldsymbol{\theta}^{(m+1)} = \boldsymbol{F}^{(m)\top}\boldsymbol{H}^{(m)}\boldsymbol{v}^{*(m)}, \quad m = 0,1,\ldots, \tag{5}$$

where

$$\boldsymbol{F}^{(m)} = (\boldsymbol{F}_1^{(m)\top},\ldots,\boldsymbol{F}_n^{(m)\top})^\top, \quad \boldsymbol{H}^{(m)} = \mathrm{bdiag}(\boldsymbol{H}_1^{(m)},\ldots,\boldsymbol{H}_n^{(m)}),$$

$\boldsymbol{v}^{*(m)} = \boldsymbol{F}^{(m)}\boldsymbol{\theta}^{(m)} + \boldsymbol{v}^{(m)}$, $\boldsymbol{v}^{(m)} = (\boldsymbol{v}_1^{(m)\top},\ldots,\boldsymbol{v}_n^{(m)\top})^\top$, $\boldsymbol{H} = \mathrm{bdiag}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_n)$ is a block diagonal matrix with the matrices $\boldsymbol{H}_1,\ldots,\boldsymbol{H}_n$ in the diagonal and m is the iteration counter. The cycles through the scheme (5) consist of iterative re-weighted least squares steps and the iterations go on until convergence is achieved. Equation (5) shows that the calculation of the MLE $\widehat{\boldsymbol{\theta}}$ can be carried out using any software with a matrix algebra library.

Since the estimating function (4) is an unbiased estimating function just by assuming that $\mathrm{E}(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i$ and $\mathrm{Var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i$, we can also resort to the estimating equation theory. We suppose valid the regularity conditions stated in Gimenez and Bolfarine (1997) (see also Van der Vaart, 1998, Ch. 5) regarding the estimating function (4). The sequences $\{\boldsymbol{\mu}_i\}$, $\{\boldsymbol{\Sigma}_i\}$, $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{z}_i\}$ must be defined to satisfy such conditions. Then, a consistent estimate for the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be formulated as the following sandwich form

$$\widehat{\mathrm{Var}}_a(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\widehat{\boldsymbol{\Psi}}_n^{-1}\widehat{\boldsymbol{\Lambda}}_n\widehat{\boldsymbol{\Psi}}_n^{-1}, \tag{6}$$

where

$$\boldsymbol{\Psi}_n = -\frac{1}{n}\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} \quad \text{and} \quad \boldsymbol{\Lambda}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{F}_i^\top \boldsymbol{H}_i \boldsymbol{v}_i \boldsymbol{v}_i^\top \boldsymbol{H}_i \boldsymbol{F}_i,$$

with $\widehat{\boldsymbol{\Psi}}_n = \boldsymbol{\Psi}_n(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\Lambda}}_n = \boldsymbol{\Lambda}_n(\widehat{\boldsymbol{\theta}})$. We assume that the limits of $\widehat{\boldsymbol{\Lambda}}_n$ and $\widehat{\boldsymbol{\Psi}}_n$ exist and are positive definite.

Thus, if the normal distribution is not tenable (which can be verified by plotting simulated envelopes or also by using the transformed distance plots proposed by Lange et al., 1989), we can use the estimator in (6) instead of the expected information matrix to build confidence regions and test statistics, since we have also that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{a}{\sim} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1})$, for n large, where $\overset{a}{\sim}$ denotes approximately distributed, $\boldsymbol{\Psi} = \lim_{n\to\infty}\boldsymbol{\Psi}_n$ and $\boldsymbol{\Lambda} = \lim_{n\to\infty}\boldsymbol{\Lambda}_n$. Under normality we have that $\boldsymbol{\Psi} = \boldsymbol{\Lambda} = \bar{\boldsymbol{K}}_{\boldsymbol{\theta}}$, where $\bar{\boldsymbol{K}}_{\boldsymbol{\theta}} = \lim_{n\to\infty}\frac{1}{n}\boldsymbol{K}_{\boldsymbol{\theta}}$. It is clear that, if the normal distribution is suitable, we should use the inverse of the information matrix, since it reaches the Cramér–Rao lower bound and, consequently, the confidence intervals will be more precise.

3. Diagnostic analysis

The first tool to perform diagnostic analysis is by means of global influence starting from case deletion with Cook (1977). Case deletion is a very common methodology to assess the effect of removing the i^{th} observation from the data set. Let $\ell_{[i]}(\boldsymbol{\theta})$ be the log-likelihood function without the i^{th} and $\widehat{\boldsymbol{\theta}}_{[i]}$ be the corresponding MLE of $\boldsymbol{\theta}$. Thus, the generalized Cook distance is given by $GD_i = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})$ and the likelihood displacement $LD_i = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{[i]})\}$. For the purpose of avoiding the estimation varying for all observations, we can use the one step approximation $\widehat{\boldsymbol{\theta}}_{[i]} = \widehat{\boldsymbol{\theta}} + \ddot{\boldsymbol{L}}^{-1}_{\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}}}\boldsymbol{U}_{\widehat{\boldsymbol{\theta}}[i]}$ (Cook and Weisberg, 1982), where $\boldsymbol{U}_{\boldsymbol{\theta}[i]} = \partial\ell_{[i]}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Also, $\boldsymbol{U}_{\widehat{\boldsymbol{\theta}}[i]}$ and $\ddot{\boldsymbol{L}}^{-1}_{\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}}}$ denote $\boldsymbol{U}_{\boldsymbol{\theta}[i]}$ and $\ddot{\boldsymbol{L}}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, respectively.

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 $C_{\boldsymbol{d}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top\boldsymbol{\Delta}^\top\ddot{\boldsymbol{L}}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{\Delta}\boldsymbol{d}|$, where $\boldsymbol{\Delta} = \partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$, both $\boldsymbol{\Delta}$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ 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 $\boldsymbol{B} = -\boldsymbol{\Delta}^\top\ddot{\boldsymbol{L}}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{\Delta}$ 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

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

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 $C_{\boldsymbol{d};\boldsymbol{\theta}_1}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top\boldsymbol{\Delta}^\top(\ddot{\boldsymbol{L}}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}\boldsymbol{d}|$, where

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

and $\boldsymbol{d}_{\max;\boldsymbol{\theta}_1}$ here is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{B}_1 = -\boldsymbol{\Delta}^\top(\ddot{\boldsymbol{L}}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}$ (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 $\boldsymbol{d}$ at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, given by $B_{\boldsymbol{d}}(\boldsymbol{\theta}) = C_{\boldsymbol{d}}(\boldsymbol{\theta})/||2\boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{\Delta}||_F$, where "$||\cdot||_F$" denotes the Frobenius norm defined as $||\boldsymbol{Z}||_F = \{\mathrm{tr}(\boldsymbol{Z}^\top\boldsymbol{Z})\}^{1/2}$ with $\boldsymbol{Z}$ being a $l \times c$ matrix. An interesting property of the conformal normal curvature is that $0 \leq B_{\boldsymbol{d}}(\boldsymbol{\theta}) \leq 1$. Therefore, this quantity can be seen as a normalized version of $C_{\boldsymbol{d}}(\boldsymbol{\theta})$.

Another procedure is the total local curvature corresponding to the i^{th} element, which follows by taking the direction $\boldsymbol{d}_i$ as a vector of zeros with one at the i^{th} position. Thus, the curvature at the direction $\boldsymbol{d}_i$ assumes the form $C_i(\boldsymbol{\theta}) = 2|\boldsymbol{\Delta}_i^\top \ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{\Delta}_i|$, where $\boldsymbol{\Delta}_i^\top$ denotes the i^{th} row of $\boldsymbol{\Delta}$. This is named total local influence by Lesaffre and Verbeke (1998). It is also possible to compute the total local influence of the i^{th} individual when estimating a subset of the elements of $\boldsymbol{\theta}$. For instance, if the interest lies on $\boldsymbol{\theta}_1$, we have that $C_{i;\boldsymbol{\theta}_1}(\boldsymbol{\theta}) = 2|\boldsymbol{\Delta}_i^\top(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}_i|$. Verbeke and Molenberghs (2000, § 11.3) proposed consider $\bar{C} = \sum_{i=1}^n C_i/n$ as a cut-off value, so that an element is potentially influential if $C_i \geq 2\bar{C}$.

Curvature calculations

Next, we derive for three perturbation schemes the matrix

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

considering the model defined in (1) and its log-likelihood function given by (2). Also, k is the dimension of the perturbation vector $\boldsymbol{\omega}$ for the scheme under consideration. Equations (7), (8) and (9) represent the main results of the paper.

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 $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$, with $0 \leq \omega_i \leq 1$, for $i = 1,\ldots,n$, and $\boldsymbol{\omega}_0 = \boldsymbol{1}_n = (1,\ldots,1)^\top$ is the vector of no perturbations. Using matrix differentiation rules we find

$$\Delta_{ri} = \frac{1}{2}\mathrm{tr}\{\widehat{\boldsymbol{A}}_{i(r)}(\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^\top)\} + \widehat{\boldsymbol{a}}_{i(r)}^\top \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \tag{7}$$

for $r = 1,\ldots,p$ and $i = 1,\ldots,n$. The matrix version of (7) is

$$\boldsymbol{\Delta} = (\boldsymbol{F}_1^\top \boldsymbol{H}_1 \boldsymbol{v}_1, \ldots, \boldsymbol{F}_n^\top \boldsymbol{H}_n \boldsymbol{v}_n).$$

In the factor analysis model studied by Kwan and Fung (1998) this perturbation scheme is equivalent to perturb the covariance matrix of the observed variables.

Perturbations affecting location and scale

Taking into account the generality of our formulation in (1), it is worth studying perturbation schemes that lead to changes only in the mean vector, only in the covariance matrix or in both, as well. For instance, in Vasconcellos and Cordeiro (1997) is analyzed a dataset for the growth of winter tillers (Faivre and Masle, 1988). The univariate response, Y_i, is the dry weight of the tillers for plants harvested from the same area. The covariate, x_i, is the time measured on a cumulative degree-days scale with a 0°C base temperature. In their working model it is considered that $\mathrm{E}(Y_i) = \mu_i = \beta_1 \exp(\beta_2 x_i)$ and $\mathrm{Var}(Y_i) = \Sigma_i = \sigma_1 \exp(\sigma_2 x_i)$. Hence, we can use the perturbation scheme derived in this section to verify the sensitivity of this model. Since the covariate is linked to the mean and the variance, the proposed perturbation scheme is justified.

In the structural equation model investigated by Lee *et al.* (2006), perturbations on the covariance matrix of the unique factors, on the manifest and latent variables, as well as on all unknown parameters are examples in which, once again, the perturbation scheme in this section is meaningful.

Here we consider that

$$\boldsymbol{Y}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*), \quad i = 1, \ldots, n,$$

where $\boldsymbol{\mu}_i^* = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i)$ and $\boldsymbol{\Sigma}_i^* = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i)$ are the perturbed mean vector and covariance matrix with $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{ik_i})^\top$ and $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$. The vector of no perturbations is defined as $\boldsymbol{\omega}_0 = (\boldsymbol{\omega}_{01}^\top, \ldots, \boldsymbol{\omega}_{0n}^\top)^\top$ such that $\boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_{0i}) = \boldsymbol{\mu}_i(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_{0i}) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$. For instance, consider the model studied by Vasconcellos and Cordeiro (1997), if we perturb the covariate as $x_i^* = x_i + \omega_i$, we will have that $\mu_i^* = \beta_1 \exp(\beta_2(x_i + \omega_i))$ and $\Sigma_i^* = \sigma_1 \exp(\sigma_2(x_i + \omega_i))$, so the mean and variance are perturb at the same time.

Define

$$\boldsymbol{a}_{i(r)}^* = \frac{\partial \boldsymbol{\mu}_i^*}{\partial \theta_r}, \quad \boldsymbol{a}_{\omega_{is}}^* = \frac{\partial \boldsymbol{\mu}_i^*}{\partial \omega_{is}}, \quad \boldsymbol{a}_{i(rs)}^* = \frac{\partial^2 \boldsymbol{\mu}_i^*}{\partial \theta_r \partial \omega_{is}}, \quad \boldsymbol{C}_{i(r)}^* = \frac{\partial \boldsymbol{\Sigma}_i^*}{\partial \theta_r}, \quad \boldsymbol{C}_{i(rs)}^* = \frac{\partial^2 \boldsymbol{\Sigma}_i^*}{\partial \theta_r \partial \omega_{is}},$$

$$\boldsymbol{A}_{i(r)}^* = -\boldsymbol{\Sigma}_i^{*-1} \boldsymbol{C}_{i(r)}^* \boldsymbol{\Sigma}_i^{*-1}, \quad \boldsymbol{A}_{\omega_{is}}^* = -\boldsymbol{\Sigma}_i^{*-1} \boldsymbol{C}_{\omega_{is}}^* \boldsymbol{\Sigma}_i^{*-1} \quad \text{and} \quad \boldsymbol{C}_{\omega_{is}}^* = \frac{\partial \boldsymbol{\Sigma}_i^*}{\partial \omega_{is}}.$$

Clearly, the expressions above depends on the chosen perturbation scheme. Then,

$$\begin{aligned}\Delta_{ris} = {} & \frac{1}{2}\operatorname{tr}\{(\widehat{\boldsymbol{A}}_{i(r)}^* \widehat{\boldsymbol{\Sigma}}_i \widehat{\boldsymbol{A}}_{\omega_{is}}^* + \widehat{\boldsymbol{A}}_{\omega_{is}}^* \widehat{\boldsymbol{\Sigma}}_i \widehat{\boldsymbol{A}}_{i(r)}^* - \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{C}}_{i(rs)}^* \widehat{\boldsymbol{\Sigma}}_i^{-1})(\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{u}}_i \widehat{\boldsymbol{u}}_i^\top) + \widehat{\boldsymbol{A}}_{\omega_{is}}^* \widehat{\boldsymbol{C}}_{i(r)}^*\} \\ & + \widehat{\boldsymbol{a}}_{i(r)}^{*\top} \widehat{\boldsymbol{A}}_{\omega_{is}}^* \widehat{\boldsymbol{u}}_i + \widehat{\boldsymbol{a}}_{\omega_{is}}^{*\top} \widehat{\boldsymbol{A}}_{i(r)}^* \widehat{\boldsymbol{u}}_i + \widehat{\boldsymbol{a}}_{i(rs)}^{*\top} \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{u}}_i - \widehat{\boldsymbol{a}}_{i(r)}^{*\top} \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{a}}_{\omega_{is}}^*,\end{aligned} \tag{8}$$

for $r = 1, \ldots, p$, $s = 1, \ldots, k_i$ and $i = 1, \ldots, n$. Here, $\boldsymbol{\Delta}$ is a $p \times \sum_{i=1}^n k_i$ matrix formed by the $1 \times k_i$ vectors $\boldsymbol{\Delta}_{ri}$ with $\boldsymbol{\Delta}_{ri} = (\Delta_{ri1}, \ldots, \Delta_{rik_i})$.

However, if one is interested only on the scale perturbation $\boldsymbol{\Sigma}_i^* = \omega_i \boldsymbol{\Sigma}_i$, then (8) becomes

$$\Delta_{ri} = \frac{1}{2} \widehat{\boldsymbol{u}}_i^\top \widehat{\boldsymbol{A}}_{i(r)} \widehat{\boldsymbol{u}}_i - \widehat{\boldsymbol{a}}_{i(r)}^\top \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{u}}_i,$$

for $r = 1, \ldots, p$ and $i = 1, \ldots, n$, since $\boldsymbol{a}_{\omega_{is}}^* = \boldsymbol{a}_{i(rs)}^* = \boldsymbol{0}$ and $\boldsymbol{a}_{i(r)}^* = \boldsymbol{a}_{i(r)}$, $\boldsymbol{A}_{\omega_{is}}^* = -\boldsymbol{\Sigma}_i^{-1}$, $\boldsymbol{C}_{i(rs)}^* = \boldsymbol{C}_{i(r)}$ and $\boldsymbol{A}_{i(r)}^* = \boldsymbol{A}_{i(r)}$ for $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Here, $\boldsymbol{\Delta} = \{\Delta_{ri}\}$ is a $p \times n$ matrix. In this context, the vector of no perturbation is the unit one.

On the other hand, if one is interested only on the location perturbation $\boldsymbol{\mu}_i^* = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i)$, we have that $\boldsymbol{A}_{w_{is}}^* = \boldsymbol{C}_{i(rs)}^* = \boldsymbol{C}_{\omega_{is}}^* = \boldsymbol{0}$ and $\boldsymbol{A}_{i(r)}^* = \boldsymbol{A}_{i(r)}$ for $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Thus, (8) becomes

$$\Delta_{ris} = \widehat{\boldsymbol{u}}_i^\top \widehat{\boldsymbol{A}}_{i(r)} \widehat{\boldsymbol{a}}_{\omega_{is}}^* + \widehat{\boldsymbol{u}}_i^\top \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{a}}_{i(rs)}^* - \widehat{\boldsymbol{a}}_{i(r)}^{*\top} \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{a}}_{\omega_{is}}^*,$$

for $r = 1, \ldots, p$, $s = 1, \ldots, k_i$ and $i = 1, \ldots, n$. Again, $\boldsymbol{\Delta}$ is a $p \times \sum_{i=1}^n k_i$ matrix with element given by the $1 \times k_i$ vector $\boldsymbol{\Delta}_{ri}$ with $\boldsymbol{\Delta}_{ri} = (\Delta_{ri1}, \ldots, \Delta_{rik_i})$ and the vector of no perturbation is the zero one.

Response perturbation

We consider here that each $\boldsymbol{Y}_i$ 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 $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$, so that the no perturbation vector is $\boldsymbol{\omega}_0 = \boldsymbol{0}$, where $\boldsymbol{\omega} \in \mathbb{R}^N$, where $N = \sum_{i=1}^n q_i$. In this case, the perturbed log-likelihood function is also given by $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\boldsymbol{\omega})$. We obtain

$$\boldsymbol{\Delta}_{ri} = -\widehat{\boldsymbol{A}}_{i(r)} \widehat{\boldsymbol{u}}_i + \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\boldsymbol{a}}_{i(r)}, \tag{9}$$

for $r = 1, \ldots, p$ and $i = 1, \ldots, n$. In matrix notation defined previously, we have that

$$\boldsymbol{\Delta} = \big(\widehat{\boldsymbol{F}}_1^\top \widehat{\boldsymbol{H}}_1 \widehat{\boldsymbol{G}}_1, \ldots, \widehat{\boldsymbol{F}}_n^\top \widehat{\boldsymbol{H}}_n \widehat{\boldsymbol{G}}_n\big), \quad \text{where} \quad \boldsymbol{G}_i = \begin{bmatrix} \boldsymbol{I}_{q_i} \\ 2\boldsymbol{u}_i \otimes \boldsymbol{I}_{q_i} \end{bmatrix}.$$

3.1. *Generalized leverage*

Let $\boldsymbol{Y} = \mathrm{vec}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ and $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathrm{vec}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)$. 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 $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{Y}^\top$. The main idea behind the concept of leverage is that of evaluating the influence of $\boldsymbol{Y}_i$ on its own predicted value. As noted by the authors, the generalized leverage is invariant under reparameterizations and observations with large GL_{ii} are leverage points.

Under the model defined in (1), we have that

$$\boldsymbol{D}_{\boldsymbol{\theta}} = (\boldsymbol{D}_1^\top, \boldsymbol{D}_2^\top, \ldots, \boldsymbol{D}_n^\top)^\top \quad \text{and} \quad \ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}} = \{\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}_i}\},$$

where $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}_i} = \{-\boldsymbol{A}_{i(r)}\boldsymbol{u}_i + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{a}_{i(r)}\}$, for $r = 1, \ldots, p$ and $i = 1, \ldots, n$. Again, by using the matrix notation defined previously, we have $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}} = \big(\boldsymbol{F}_1^\top\boldsymbol{H}_1\boldsymbol{G}_1, \ldots, \boldsymbol{F}_n^\top\boldsymbol{H}_n\boldsymbol{G}_n\big)$. Index plots of GL_{ii} may reveal those observations with high influence on their own predicted values.

3.2. *Connection between local influence and generalized leverage*

There is a connection between local influence and generalized leverage. In order to deduce such relationship we must define some quantities. Define $\boldsymbol{F}_i = (\boldsymbol{F}_{1i}^\top, \boldsymbol{F}_{2i}^\top)^\top$, $\boldsymbol{G}_i = (\boldsymbol{G}_{1i}^\top, \boldsymbol{G}_{2i}^\top)^\top$ where $\boldsymbol{F}_{1i} = \boldsymbol{D}_i$, $\boldsymbol{F}_{2i} = \boldsymbol{V}_i$, $\boldsymbol{G}_{1i} = \boldsymbol{I}_{q_i}$ and $\boldsymbol{G}_{2i} = 2\boldsymbol{u}_i \otimes \boldsymbol{I}_{q_i}$. Define also,

$$\boldsymbol{F}^* = (\boldsymbol{F}_1^\top, \ldots, \boldsymbol{F}_n^\top)^\top \quad \text{and} \quad \widetilde{\boldsymbol{F}} = (\widetilde{\boldsymbol{F}}_1^\top, \widetilde{\boldsymbol{F}}_2^\top)^\top,$$

where $\widetilde{\boldsymbol{F}}_j = (\boldsymbol{F}_{j1}^\top, \ldots, \boldsymbol{F}_{jn}^\top)^\top$ for $j = 1, 2$. Notice that there exists a permutation matrix $\boldsymbol{I}^*$ such that $\boldsymbol{F}^* = \boldsymbol{I}^*\widetilde{\boldsymbol{F}}$ (see, for instance, Magnus and Neudecker, 2007, Ch. 1). Therefore, we have that

$$\boldsymbol{\Delta}^\top(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\boldsymbol{\Delta} = \boldsymbol{P}^\top\boldsymbol{I}^*\widetilde{\boldsymbol{F}}(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\boldsymbol{\Delta},$$

where $\boldsymbol{P} = \mathrm{bdiag}(\boldsymbol{H}_1\boldsymbol{G}_1, \ldots, \boldsymbol{H}_n\boldsymbol{G}_n)$. Note also that $\boldsymbol{P}^\top\boldsymbol{I}^* = (\boldsymbol{P}_1^{*\top}, \boldsymbol{P}_2^{*\top})$, where

$$\boldsymbol{P}_1^* = \mathrm{bdiag}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{G}_{11}, \ldots, \boldsymbol{\Sigma}_n^{-1}\boldsymbol{G}_{1n}) \quad \text{and} \quad \boldsymbol{P}_2^* = (1/2)\,\mathrm{bdiag}\big((\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{\Sigma}_1^{-1})\boldsymbol{G}_{21}, \ldots, (\boldsymbol{\Sigma}_n^{-1} \otimes \boldsymbol{\Sigma}_n^{-1})\boldsymbol{G}_{2n}\big).$$

Hence, the relationship between the local influence (under additive response perturbations) and the generalized leverage is

$$\boldsymbol{\Delta}^\top(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\boldsymbol{\Delta} = \boldsymbol{P}_1^{*\top}\boldsymbol{GL}(\boldsymbol{\theta}) + \boldsymbol{P}_2^{*\top}\widetilde{\boldsymbol{F}}_2(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\boldsymbol{\Delta}.$$

Then, the normal curvature under additive perturbations in the response values can be rewritten as $C_{\boldsymbol{d}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top\boldsymbol{P}_1^{*\top}\boldsymbol{GL}(\boldsymbol{\theta})\boldsymbol{d} - \boldsymbol{d}^\top\boldsymbol{P}_2^{*\top}\widetilde{\boldsymbol{F}}_2(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\boldsymbol{\Delta}\boldsymbol{d}|$ in the general case.

4. Special models

In this section we present some particular cases of our general model defined in (1). To the best of our knowledge, the results obtained here had not been reported in the statistical literature.

4.1. Heteroskedastic mixed model with nonlinear fixed effects

The mixed model considered here generalizes many of the most important mixed models presented in the literature, for example, Beckman *et al.* (1987) and Lesaffre and Verbeke (1998). We define the heteroskedastic mixed model with nonlinear fixed effects by the stochastic equation

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n, \tag{10}$$

where $\boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathcal{N}_{m_i}(\boldsymbol{0}, \boldsymbol{\Gamma}_i(\boldsymbol{\sigma}_1))$ independent of $\boldsymbol{e}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{0}, \boldsymbol{R}_i(\boldsymbol{\sigma}_2))$. The dimensions of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are p_1, p_2 and $p - p_1 - p_2$, respectively. Defining $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1^\top, \boldsymbol{\sigma}_2^\top)^\top$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\sigma}) = \boldsymbol{Z}_i\boldsymbol{\Gamma}_i(\boldsymbol{\sigma}_1)\boldsymbol{Z}_i^\top + \boldsymbol{R}_i(\boldsymbol{\sigma}_2)$, the marginal distribution of $\boldsymbol{Y}_i$ can be written as $\boldsymbol{Y}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}), \boldsymbol{\Sigma}_i(\boldsymbol{\sigma}))$. It is noteworthy that the mean vector and the covariance-variance matrix do not share parameters and, of course, it is a particular case of general model (1).

The mean vector may depend on extra covariates, i.e., $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{x}_i)$. In addition, it allows nonlinear relationships with them. The covariance-variance matrices $\boldsymbol{\Gamma}_i(\boldsymbol{\sigma}_1)$ and $\boldsymbol{R}_i(\boldsymbol{\sigma}_2)$ may also vary with the observation; for instance, each element of these matrices may depend on $\boldsymbol{x}_i$, characterizing a heteroskedastic model. It is noteworthy that model (10) reduces to the model studied by Beckman *et al.* (1987) just by taking $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = \boldsymbol{X}_i\boldsymbol{\beta}$, $\boldsymbol{R}_i(\boldsymbol{\sigma}_2) = \sigma^2\boldsymbol{I}_{q_i}$ and $\boldsymbol{\Gamma}_i(\boldsymbol{\sigma}_1)$ is a diagonal matrix with an appropriated structure, where $\boldsymbol{I}_{q_i}$ is the $q_i \times q_i$ identity matrix.

Next, we obtain the matrix $\boldsymbol{\Delta}$ for four perturbation schemes (case weighting, scale matrix, location and response perturbation). Define $\boldsymbol{\Gamma}_{i(r)} = \partial\boldsymbol{\Gamma}_i(\boldsymbol{\sigma}_1)/\partial\theta_r$ and $\boldsymbol{R}_{i(r)} = \partial\boldsymbol{R}_i(\boldsymbol{\sigma}_2)/\partial\theta_r$. Notice that, from the general expressions given in (7), (8) and (9), it is easy to obtain

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \\ \boldsymbol{\Delta}_{\boldsymbol{\sigma}_1} \\ \boldsymbol{\Delta}_{\boldsymbol{\sigma}_2} \end{pmatrix}$$

for the scheme under consideration. We have immediately that:

Case weight perturbation

$$\Delta_{ri} = \begin{cases} \widehat{\boldsymbol{a}}_{i(r)}^\top \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, & \text{for } r = 1, \ldots, p_1, \\ -\frac{1}{2}\text{tr}\{\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{i(r)}\boldsymbol{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}(\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^\top)\}, & \text{for } r = p_1 + 1, \ldots, p_1 + p_2, \\ -\frac{1}{2}\text{tr}\{\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{R}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}(\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^\top)\}, & \text{for } r = p_1 + p_2 + 1, \ldots, p, \end{cases}$$

for $i = 1, \ldots, n$.

Scale matrix perturbation – $\boldsymbol{\Sigma}_i^* = \omega_i\boldsymbol{\Sigma}_i$

$$\Delta_{ri} = \begin{cases} -\widehat{\boldsymbol{a}}_{i(r)}^\top \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, & \text{for } r = 1, \ldots, p_1, \\ -\frac{1}{2}\widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{i(r)}\boldsymbol{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, & \text{for } r = p_1 + 1, \ldots, p_1 + p_2, \\ -\frac{1}{2}\widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{R}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, & \text{for } r = p_1 + p_2 + 1, \ldots, p, \end{cases}$$

for $i = 1, \ldots, n$.

Location perturbation – $\boldsymbol{\mu}_i^* = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i)$

$$\Delta_{ris} = \begin{cases} \widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{i(rs)}^* - \widehat{\boldsymbol{a}}_{i(r)}^{*\top}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{\omega_{is}}^*, & \text{for } r = 1, \ldots, p_1, \\ -\widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{i(r)}\boldsymbol{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{\omega_{is}}^*, & \text{for } r = p_1 + 1, \ldots, p_1 + p_2, \\ -\widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{R}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{\omega_{is}}^*, & \text{for } r = p_1 + p_2 + 1, \ldots, p, \end{cases}$$

for $i = 1, \ldots, n$.

Response perturbation

$$\boldsymbol{\Delta}_{ri} = \begin{cases} \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{i(r)}, \quad \text{for } r = 1, \dots, p_1, \\ \widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Z}_i\widehat{\boldsymbol{\Gamma}}_{i(r)}\boldsymbol{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \quad \text{for } r = p_1+1, \dots, p_1+p_2, \\ \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{R}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \quad \text{for } r = p_1+p_2+1, \dots, p, \end{cases}$$

for $i = 1, \dots, n$.

4.2. Heteroskedastic multivariate nonlinear regression

This section considers a heteroskedastic multivariate nonlinear regression. The model is

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) + \boldsymbol{e}_i, \quad i = 1, \dots, n,$$

where $\boldsymbol{e}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\sigma}))$. The response variable is such that $\boldsymbol{Y}_i \overset{\text{ind}}{\sim} \mathcal{N}_{q_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}), \boldsymbol{\Sigma}_i(\boldsymbol{\sigma}))$. The vector of parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\sigma}^\top)^\top$. Let p_1 and $p - p_1$ be the dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$, respectively.

Again, from the general expressions given in (7), (8) and (9), it is easy to obtain

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \\ \boldsymbol{\Delta}_{\boldsymbol{\sigma}} \end{pmatrix}$$

for the scheme under consideration. We have immediately the Delta matrices for different perturbation schemes:

Case weight perturbation

$$\Delta_{ri} = \begin{cases} \widehat{\boldsymbol{a}}_{i(r)}^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \quad \text{for } r = 1, \dots, p_1, \\ -\frac{1}{2}\text{tr}\{\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{C}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}(\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^\top)\}, \quad \text{for } r = p_1+1, \dots, p, \end{cases}$$

for $i = 1, \dots, n$.

Scale matrix perturbation — $\boldsymbol{\Sigma}_i^* = \omega_i\boldsymbol{\Sigma}_i$

$$\Delta_{ri} = \begin{cases} -\widehat{\boldsymbol{a}}_{i(r)}^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \quad \text{for } r = 1, \dots, p_1, \\ -\frac{1}{2}\widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{C}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \quad \text{for } r = p_1+1, \dots, p, \end{cases}$$

for $i = 1, \dots, n$.

Location perturbation – $\boldsymbol{\mu}_i^* = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i)$

$$\boldsymbol{\Delta}_{ri} = \begin{cases} \widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{i(rs)}^* - \widehat{\boldsymbol{a}}_{i(r)}^{*\top}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{\omega_{is}}^*, \quad \text{for } r = 1, \dots, p_1, \\ -\widehat{\boldsymbol{u}}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{C}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{\omega_{is}}^*, \quad \text{for } r = p_1+1, \dots, p, \end{cases}$$

for $i = 1, \dots, n$.

Response perturbation

$$\boldsymbol{\Delta}_{ri} = \begin{cases} \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{a}}_{i(r)}, \quad \text{for } r = 1, \dots, p_1, \\ \widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{C}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{u}}_i, \quad \text{for } r = p_1+1, \dots, p, \end{cases}$$

for $i = 1, \dots, n$.

4.3. Heteroskedastic measurement error model

Kulathinal *et al.* (2002) analyzed an epidemiological dataset where the response variable, y_i, is an index of the cardiovascular mortality and the covariate, x_i, is an index composed by its risk factors. A linear relationship $y_i = \alpha + \beta x_i + \eta_i$ is postulated, where η_i represents the equation error. Both response and covariate are subject to measurement errors and the working model is an errors-in-variables one with observations $Y_i = y_i + e_i$ and $X_i = x_i + u_i$, where $e_i \overset{\text{ind}}{\sim} \mathcal{N}(0, \tau_{e_i})$ and $u_i \overset{\text{ind}}{\sim} \mathcal{N}(0, \tau_{u_i})$, respectively, are the measurement errors with variances τ_{e_i} and τ_{u_i} supposedly known. It is also assumed that $x_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$, $\eta_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\eta^2)$ and the variables (e, u, η, x) are all independent. In this setup, the mean vector and the covariance matrix of the observed vector $\boldsymbol{Y}_i = (Y_i, X_i)^\top$ are given, respectively, by

$$\boldsymbol{\mu}_i = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_i = \begin{bmatrix} \beta^2\sigma_x^2 + \sigma_\eta^2 + \tau_{e_i} & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \tau_{u_i} \end{bmatrix}, \quad i = 1, \ldots, n.$$

Therefore, one might be interested if small perturbations on τ_{e_i} and τ_{u_i} have high influence on the ML estimates of the model parameters $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma_\eta^2)^\top$. Thus, the perturbation scheme

$$\boldsymbol{\Sigma}_i^* = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \boldsymbol{\omega}) = \begin{bmatrix} \beta^2\sigma_x^2 + \sigma_\eta^2 + \omega_i\tau_{e_i} & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \omega_i\tau_{u_i} \end{bmatrix}$$

might be considered, with $\omega_i > 0$ and $\boldsymbol{\omega}_0 = \mathbf{1}_n$. For this case, we have immediately from (8) that

$$\Delta_{ri} = \frac{1}{2}\text{tr}\{(\widehat{\boldsymbol{A}}_{i(r)}\widehat{\boldsymbol{\Sigma}}_i\widehat{\boldsymbol{A}}^*_{\omega_i} + \widehat{\boldsymbol{A}}^*_{\omega_i}\widehat{\boldsymbol{\Sigma}}_i\widehat{\boldsymbol{A}}_{i(r)})(\widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^\top) + \widehat{\boldsymbol{A}}^*_{\omega_i}\widehat{\boldsymbol{C}}_{i(r)}\} + \widehat{\boldsymbol{a}}_{i(r)}^\top\widehat{\boldsymbol{A}}^*_{\omega_i}\widehat{\boldsymbol{u}}_i,$$

for $r = 1, \ldots, p$ and $i = 1, \ldots, n$, since $\boldsymbol{a}^*_{\omega_{is}} = \boldsymbol{a}^*_{i(rs)} = \mathbf{0}$ and $\boldsymbol{C}^*_{i(rs)} = \mathbf{0}$, and $\boldsymbol{a}^*_{i(r)} = \boldsymbol{a}_{i(r)}$, $\boldsymbol{A}^*_{i(r)} = \boldsymbol{A}_{i(r)}$ and $\boldsymbol{C}^*_{i(r)} = \boldsymbol{C}_{i(r)}$ for $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Here, $\boldsymbol{C}^*_{\omega_i} = \text{diag}(\tau_{e_i}, \tau_{u_i})$.

5. Concluding remarks

We have discussed in this paper applications of local influence and generalized leverage methods in a multivariate normal regression model with general parameterization. This model considers that the mean vector and the covariance matrix share the same vector of parameters. It includes many regression models as particular cases. Appropriate matrices for assessing local influence on the parameter estimates under different perturbation schemes are obtained. Our results are very general and can be applied to any model as defined by (1), that is, this paper can be used as a guide for computing diagnostics measures in practical applications. In particular, we derive the normal curvature of local influence under some perturbation schemes for a large class of heteroskedastic nonlinear models with longitudinal structure (mixed model).

Acknowledgments

The first two authors gratefully acknowledge grants from FAPESP (Brazil). The third author acknowledges partial support from CNPq, Brazil (contract #305705/2009-1). The authors are also grateful to a referee for helpful comments and suggestions.

References

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

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

Carrasco, J.M.F., Ortega, E.M.M., Paula, G.A. (2008). Log-modified Weibull regression models with censored data: sensitivity and residual analysis. *Computational Statistics and Data Analysis* **52**, 4021–4039.

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. (1977). Detection of influential observation in linear regression. *Technometrics* **19**, 15–18.

Cook, R.D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* **48**, 133–169.
Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
Cox, D.R., Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall, London.
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.
Espinheira, P.L., Ferrari, S.L.P., Cribari-Neto, F. (2008). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis* **52**, 4417–4431.
Faivre, R., Masle, J. (1988). Modeling potential growth of tillers in winter-wheat. *Acta Oecologica–Oecologia Generalis* **9**, 179–196.
Gimenez, P., Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameters models. *The Australian Journal of Statistics*, **39**, 325–344.
Kelly, B.C. (2007). Some aspects of measurement error in linear regression of astronomical data. *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 errors vary between the observations. *Statistics in Medicine* **21**, 1089–1101.
Kwan, C.W., Fung, W.K. (1998). Assessing local influence for specific restricted likelihood: application to factor analysis. *Psychometrika* **63**, 35–46.
Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989). Robust statistical modelling using the tdistribution. *Journal of the American Statistical Association* **84**, 881–896.
Lawrance, A.J. (1988). Regression transformation diagnostics using local influence. *Journal of the American Statistical Association* **83**, 1067–1072.
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.
Lesaffre, E., Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **54**, 570–582.
Magnus, J.R., Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, third edition.
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. (2010). A note on influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis* doi: 10.1016/j.csda.2010.06.007.
Patriota, A.G., Bolfarine, H., de Castro, M. (2009). A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology* **6**, 408–423
Patriota, A.G., Lemonte, A.J. (2009). Bias correction in a multivariate normal regression model with general parameterization. *Statistics and Probability Letters* **79**, 1655–1662.
Paula, G.A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis* **16**, 63–79.
Paula, G.A., Medeiros, M., Vilca-Labra, F.E. (2009). Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics and Probability Letters* **79**, 339–346.
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.
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.
Thomas, W., Cook, R.D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics* **32**, 59–65.
Van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge.
Vasconcellos, K.L.P., Cordeiro, G.M. (1997). Approximate bias for multivariate nonlinear heteroscedastic regressions. *Brazilian Journal of Probability and Statistics* **11**, 141–159.
Vasconcellos, K.L.P., Fernandez, L.M.Z. (2009). Influence analysis with homogeneous linear restrictions. *Computational Statistics and Data Analysis* **53**, 3787–3794.
Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.
Wei, B.-C., Hu, Y.-Q., Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25–37.
Zhao, Y.J., Lee, A.H. (1998). Influence diagnostics for simultaneous equations models. *Australian and New Zealand Journal of Statistics* **40**, 345–357.