

Diagnostics analysis for skew-normal linear regression models: Applications to a quality of life dataset

Clécio da Silva Ferreira^a, Filidor Vilca^b and Heleno Bolfarine^c

^a*Universidade Federal de Juiz de Fora*

^b*Universidade Estadual de Campinas*

^c*Universidade de São Paulo*

Abstract. The skew-normal distribution has been used successfully in various statistical applications. The main purpose of this paper is to consider local influence analysis, which is recognized as an important step of data analysis. Motivated to simplify expressions of the conditional expectation of the complete-data log-likelihood function, used in the EM algorithm, diagnostic measures are derived from the case-deletion approach and the local influence approach inspired by Zhu et al. [*Biometrika* **88** (2001) 727–737] and Zhu and Lee [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** (2001) 111–126]. Finally, the results obtained are applied to a dataset from a study to evaluate quality of life (QOL) and to identify its associated factors in climacteric women with a history of breast cancer.

1 Introduction

Linear symmetrical models have been investigated by various authors. For example, Lange, Little and Taylor (1989) presented an approach to model Student-t distributions, while Galea-Rojas, Paula and Bolfarine (1997) and Liu (2000) discussed diagnostic methods for symmetrical linear regression models. Although the class of symmetric distributions is a better alternative than that of the normal distribution, it is not appropriate in situations where the sample distribution is asymmetrical. For example, Hill and Dixon (1982) discussed and presented examples with asymmetrical structures. From a practical viewpoint, many authors have transformed variables to achieve normality, and in many cases their methods are satisfactory. However, Azzalini and Capitanio (1999) pointed out some problems, mentioning, for example, that the transformed variables are more difficult to deal with and to interpret.

A family of skew-normal distributions that accommodates practical values of skewness and kurtosis, and includes the normal distribution as a special case, was introduced by Azzalini (1985). From then on, many authors have considered these distributions in different areas, such as economics, oceanography, engineering and biomedical sciences, among others. Azzalini (2005) presented a discussion

Key words and phrases. Case-deletion, local influence, skew-normal distribution, approach, EM algorithm.

Received August 2015; accepted January 2017.

on skew-normal distributions with applications in regression models. Also, Lachos et al. (2007) considered an application of diagnostics analysis in linear mixed models, and in the context of errors-in-variables models some results can be found in Lachos, Montenegro and Bolfarine (2008). Sahu, Dey and Branco (2003) proposed a new skew-normal distribution, equivalent to that of Azzalini (1985), in which the EM algorithm used to obtain the maximum likelihood estimates is simpler in the M-step. Consequently, some of its properties and applications can be easily derived, such as diagnostics analysis based the conditional expectation of the complete-data log-likelihood; see Zhu and Lee (2001) and Zhu et al. (2001). Some applications of the skew-normal of Sahu, Dey and Branco (2003) can be found in Azevedo, Bolfarine and Andrade (2011), Lee and McLachlan (2013) and Dagne (2016), among others.

After the model is fitted, diagnostics analysis is a key step in data analysis subsequent to parameter estimation. This procedure can be carried out by conducting an influence analysis to detect influential observations. There are two usual approaches to detect influential observations. The first is the case-deletion approach, which has received a great deal of attention due to the paper by Cook (1977), which that presented an intuitively appealing method. Since then, the Cook distance or the likelihood distance have been applied to many statistical models. This approach allows assessment of the individual impact of cases in the estimation process (see, Cook and Weisberg (1982)). The second approach, which is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs, is the local influence approach, due to Cook (1986). This method has received considerable attention recently in the statistical literature on regression models. Both approaches can be studied on the basis of the Q -function considered by Zhu et al. (2001) to introduce the case-deletion measures and by Zhu and Lee (2001) to study local influence diagnostics as can be seen in Ferreira, Lachos and Bolfarine (2015), Massuia et al. (2015), and Zeller, Lachos and Vilca (2014).

Although several diagnostic studies on regression models have appeared in the literature, no application of local influence has been considered for regression models under the skew-normal distribution of Sahu, Dey and Branco (2003) with emphasis on the case-deletion approach and local influence. Based on the Q -function, which is closely related to the conditional expectation of the complete-data log-likelihood in the E-step of the EM algorithm, we develop a study of local influence following the approach of Zhu and Lee (2001), which produces results very similar to those obtained from Cook's method, but with simpler application. We also develop methods to obtain case-deletion measures by using the method of Zhu et al. (2001).

Following Sahu, Dey and Branco (2003), we say that an random variable Y has an univariate skew-normal distribution with location parameter μ , scale parameter

σ^2 and skewness parameter δ , if its probability density function (pdf) is given by

$$f(y) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} \phi\left(\frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}}\right) \Phi\left(\frac{\delta}{\sigma} \frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}}\right), \tag{1.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cumulative distribution function (cdf) of the $N(0, 1)$ distribution, respectively. This distribution is denoted by $Y \sim SN(\mu, \sigma^2, \delta)$. It is easy to see that for $\delta = 0$, the pdf in (1.1) reduces to the normal distribution. The mean and variance of Y are

$$E(Y) = \mu + \sqrt{\frac{2}{\pi}} \delta \quad \text{and} \quad \text{Var}(Y_j) = \sigma^2 + \left(1 - \frac{2}{\pi}\right) \delta^2, \tag{1.2}$$

respectively. The stochastic representation is given by $Y \stackrel{d}{=} \mu + \delta|X_0| + \sigma X_1$, where X_0 and X_1 are independent with distribution $N(0, 1)$. The notation “ $\stackrel{d}{=}$ ” means that both variables have the same distribution.

The paper is organized as follows. Section 2 introduces the skew-normal linear regression model and describes an EM algorithm for obtaining the maximum likelihood (ML) estimates. Section 3 provides a brief sketch of the case-deletion and local influence approaches for models with incomplete data, and also develops a method pertinent to linear regression models under the skew-normal distribution. Moreover, the generalized leverage is discussed. Section 4 illustrates the method with analysis of two examples involving life quality of women with breast cancer and simulation studies. Finally, some concluding remarks are made in Section 5.

2 The skew-normal linear regression model

In this section, we present the linear regression models under the skew-normal distribution (SN-LRM). Specifically, we consider the linear regression model under the skew-normal distribution proposed by Sahu, Dey and Branco (2003), as follows:

$$Y_j = \mathbf{x}_j^\top \boldsymbol{\beta} + \delta U_j + \varepsilon_j, \tag{2.1}$$

where $\mathbf{x}_j^\top = (1, x_{j2}, \dots, x_{jp})$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$, $\varepsilon_j \sim N(0, \sigma^2)$ and $U_j \sim HN(0, 1)$, $j = 1, \dots, n$, all independent, where $HN(0, 1)$ denotes the standard half-normal distribution. So, $Y_j \sim SN(\mathbf{x}_j^\top \boldsymbol{\beta}, \sigma^2, \delta)$ with pdf give by

$$f(y_j; \boldsymbol{\theta}) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} \phi\left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta}}{\sqrt{\sigma^2 + \delta^2}}\right) \Phi\left(\frac{\delta}{\sigma} \frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta}}{\sqrt{\sigma^2 + \delta^2}}\right), \tag{2.2}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are as in (1.1). The mean and variance of Y_j are given by

$$\mathbb{E}(Y_j) = \mathbf{x}_j^\top \boldsymbol{\beta} + \sqrt{\frac{2}{\pi}} \delta \quad \text{and} \quad \text{Var}(Y_j) = \sigma^2 + \left(1 - \frac{2}{\pi}\right) \delta^2, \tag{2.3}$$

respectively. The log-likelihood function for $\theta = (\boldsymbol{\beta}^\top, \sigma^2, \delta)$, given y_1, \dots, y_n , is given by

$$\begin{aligned} \ell(\theta) = & n \log 2 - \frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2 + \delta^2) \\ & - \frac{1}{2(\sigma^2 + \delta^2)} \sum_{j=1}^n (y_j - \mathbf{x}_j^\top \boldsymbol{\beta})^2 + \sum_{j=1}^n \log \Phi(B_j), \end{aligned} \quad (2.4)$$

where $B_j = \frac{\delta}{\sigma \sqrt{\sigma^2 + \delta^2}} (y_j - \mathbf{x}_j^\top \boldsymbol{\beta})$. After some algebraic manipulations, the information matrix denoted by $I_F(\theta)$ is given by $I_F(\theta) = [I_{\boldsymbol{\gamma}\boldsymbol{\psi}}]$, $\boldsymbol{\gamma}, \boldsymbol{\psi} = \boldsymbol{\beta}, \sigma^2, \delta$, whose components are given by

$$\begin{aligned} I_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{1}{\sigma^2 + \delta^2} \left[1 + \frac{\delta^2}{\sigma^2} a_{02}(\delta/\sigma) \right] \mathbf{X}^\top \mathbf{X}, \\ I_{\boldsymbol{\beta}\sigma^2} &= \frac{\delta}{(\sigma^2 + \delta^2)^{3/2}} \left[\frac{c}{(\sigma^2 + \delta^2)^{1/2}} - \frac{2\sigma^2 + \delta^2}{2\sigma^3} a_1(\delta, \sigma) \right] \mathbf{X}, \\ I_{\boldsymbol{\beta}\delta} &= \frac{\sigma}{(\sigma^2 + \delta^2)^{3/2}} \left[2c \frac{\delta^2}{(\sigma^2 + \delta^2)^{1/2}} - \delta a_2(\delta, \sigma) + \sigma a_{01}(\delta/\sigma) \right] \mathbf{X}, \\ I_{\sigma^2\sigma^2} &= \frac{n}{(\sigma^2 + \delta^2)^2} \left[\frac{1}{2} + \frac{(2\sigma^2 + \delta^2)^2 \delta^2}{4\sigma^6} a_{22}(\delta/\sigma) \right], \\ I_{\sigma^2\delta} &= \frac{n\delta}{(\sigma^2 + \delta^2)^2} \left[1 - \frac{(2\sigma^2 + \delta^2)}{2\sigma^2} a_{22}(\delta/\sigma) \right], \\ I_{\delta\delta} &= \frac{n}{(\sigma^2 + \delta^2)^2} [2\delta^2 + \sigma^2 a_{22}(\delta/\sigma)], \end{aligned}$$

where $a_1(\delta, \sigma) = a_{01}(\delta/\sigma) - \frac{\delta}{\sigma} a_{12}(\delta/\sigma) - \frac{\delta^2}{\sigma^2} a_{21}(\delta/\sigma)$, $a_2(\delta, \sigma) = \frac{\delta}{\sigma} a_{21}(\delta/\sigma) + a_{12}(\delta/\sigma)$, $c = \sqrt{2/\pi}$ and $a_{hk}(x) = \mathbb{E}[Z^h W_{\Phi}^k(xZ)]$, with $a_{01}(x) = c(x^2 + 1)^{-1/2}$ and $a_{21}(x) = c(x^2 + 1)^{-3/2}$, and \mathbf{X} is the model matrix. The other values $a_{12}(\cdot)$ and $a_{22}(\cdot)$ are obtained through the approximation given in [Rodríguez and Branco \(2007\)](#).

Next, we present the Q -function as an alternative to the observed-data log-likelihood function, which is associated with the conditional expectation of the complete-data log-likelihood function in the EM algorithm. The use of the Q -function will be useful to detect influential observations in regression models when the errors follow a skew-normal distribution because it is significantly simpler. This idea is inspired by the works of [Zhu and Lee \(2001\)](#) and [Zhu et al. \(2001\)](#).

2.1 The Q -function and the EM algorithm

We implement the EM algorithm introduced by [Dempster, Laird and Rubin \(1977\)](#), in which the M-step involves only the Q -function complete data, which in this case

is computationally simple. The model (2.1) can be expressed by the hierarchical representation

$$Y_j|U_j = u_j \sim N(\mathbf{x}_j^\top \boldsymbol{\beta} + \delta u_j, \sigma^2) \quad \text{and} \quad U_j \sim HN(0, 1). \tag{2.5}$$

Hence, $U_j|Y_j = y_j \sim HN(\frac{\delta}{\sigma^2 + \delta^2}(y_j - \mathbf{x}_j^\top \boldsymbol{\beta}), \frac{\sigma^2}{\sigma^2 + \delta^2})$. Now, letting $\mathbf{y} = (y_1, \dots, y_n)^\top$ and $\mathbf{u} = (u_1, \dots, u_n)^\top$, treating \mathbf{u} as missing data, we have the complete-data log-likelihood function associated with $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ is given by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{j=1}^n \ell_c(\boldsymbol{\theta}|y_j, u_j)$, where (without the additive constant)

$$\ell_c(\boldsymbol{\theta}|y_j, u_j) = -\frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y_j - \mathbf{x}_j^\top \boldsymbol{\beta})^2 + \frac{\delta}{\sigma^2} u_j (y_j - \mathbf{x}_j^\top \boldsymbol{\beta}) - \frac{\delta^2}{2\sigma^2} u_j^2.$$

Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\beta}}^{(k)\top}, \widehat{\sigma}^{2(k)}, \widehat{\delta}^{(k)})^\top$ of $\boldsymbol{\theta}$ at the k th iteration, the E-step calculates the conditional expectation of the complete-data log-likelihood function or simply the Q -function

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{j=1}^n \mathbb{E}[\ell_c(\boldsymbol{\theta}|y_j, u_j)|y_j, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}] = \sum_{j=1}^n Q_j(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}), \tag{2.6}$$

where

$$Q_j(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = C^* - \frac{1}{2} \log \widehat{\sigma}^{2(k)} - \frac{1}{2\widehat{\sigma}^{2(k)}} \widehat{\epsilon}_j^{2(k)} + \frac{\widehat{\delta}^{(k)}}{\widehat{\sigma}^{2(k)}} \widehat{u}_j^{(k)} \widehat{\epsilon}_j^{(k)} - \frac{\widehat{\delta}^{2(k)}}{2\widehat{\sigma}^{2(k)}} \widehat{u}_j^{2(k)},$$

with $\widehat{\epsilon}_j^{(k)} = y_j - \mathbf{x}_j^\top \widehat{\boldsymbol{\beta}}^{(k)}$, and the quantities $\widehat{u}_j^{(k)} = \mathbb{E}(U_j|y_j, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)})$ and $\widehat{u}_j^{2(k)} = \mathbb{E}(U_j^2|y_j, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)})$ are obtained from properties of the half-normal distribution

$$\widehat{u}_j^{(k)} = \frac{\widehat{\delta}^{(k)}}{\widehat{\sigma}^{2(k)} + \widehat{\delta}^{2(k)}} \widehat{\epsilon}_j^{(k)} + \frac{\widehat{\sigma}^{(k)}}{(\widehat{\sigma}^{2(k)} + \widehat{\delta}^{2(k)})^{1/2}} W_\Phi(B_j^{(k)}), \tag{2.7}$$

$$\begin{aligned} \widehat{u}_j^{2(k)} &= \frac{\widehat{\sigma}^{2(k)}}{\widehat{\sigma}^{2(k)} + \widehat{\delta}^{2(k)}} + \frac{\widehat{\delta}^{2(k)}}{(\widehat{\sigma}^{2(k)} + \widehat{\delta}^{2(k)})^2} \widehat{\epsilon}_j^{2(k)} \\ &\quad + \frac{\widehat{\delta}^{(k)} \widehat{\sigma}^{(k)}}{(\widehat{\sigma}^{2(k)} + \widehat{\delta}^{2(k)})^{3/2}} \widehat{\epsilon}_j^{(k)} W_\Phi(B_j^{(k)}), \end{aligned} \tag{2.8}$$

where B_j is as in (2.4), $j = 1, \dots, n$ and $W_\Phi(u) = \phi(u)/\Phi(u)$. Thus, we have the following EM-algorithm:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, compute $\widehat{u}_j^{(k)}$ and $\widehat{u}_j^{2(k)}$ for $j = 1, \dots, n$, using (2.7) and (2.8).

M-step: Update $\widehat{\boldsymbol{\theta}}^{(k+1)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following closed form expressions

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(k+1)} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} - \widehat{\delta}^{(k)} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{u}}^{(k)}, \\ \widehat{\sigma}^2{}^{(k+1)} &= \frac{1}{n} \left[\widehat{\boldsymbol{\epsilon}}^{(k)\top} \widehat{\boldsymbol{\epsilon}}^{(k)} - 2\widehat{\delta}^{(k)} \widehat{\mathbf{u}}^{(k)\top} \widehat{\boldsymbol{\epsilon}}^{(k)} + \widehat{\delta}^2{}^{(k)} \sum_{j=1}^n \widehat{u}_j^2{}^{(k)} \right], \\ \widehat{\delta}^{(k+1)} &= \frac{\widehat{\mathbf{u}}^{(k)\top} \widehat{\boldsymbol{\epsilon}}^{(k)}}{\sum_{j=1}^n \widehat{u}_j^2{}^{(k)}},\end{aligned}$$

where $\widehat{\mathbf{u}}^{(k)} = (\widehat{u}_1^{(k)}, \dots, \widehat{u}_n^{(k)})^\top$, $\widehat{\boldsymbol{\epsilon}}^{(k)} = (\widehat{\epsilon}_1^{(k)}, \dots, \widehat{\epsilon}_n^{(k)})^\top$ and \mathbf{X} is the model matrix.

3 Diagnostics analysis

Influence diagnostic techniques are used to identify anomalous observations that impact model fit or statistical inference for the assumed statistical model. In the literature, there are primarily two approaches to detect influential observations. The case-deletion approach of Cook (1977) is the most popular one for identifying influential observations. For assessing the impact of influential observations on parameter estimates, some metrics have been used to measure the distance between $\widehat{\boldsymbol{\theta}}_{[j]}$ and $\widehat{\boldsymbol{\theta}}$, such as the likelihood distance and Cook's distance. The second approach is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs proposed by Cook (1986). We study here the case-deletion measures and the local influence diagnostics for linear regression models on the basis of the Q -function inspired by the results of Zhu et al. (2001), Zhu and Lee (2001) and Lee and Xu (2004). We first consider the case-deletion measures, then the local influence measures based on some perturbation schemes and generalized leverage.

3.1 Case-deletion measures

Case-deletion is a classic approach to study the effects of dropping the j th case from the dataset. Let $\mathbf{y}_c = (\mathbf{y}, \mathbf{u})$ be the augmented dataset, where a quantity with a subscript “[j]” denotes the original one with the j th observation deleted. The complete-data log-likelihood function based on the data with the j th case deleted is denoted by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_{c[j]})$. Let $\widehat{\boldsymbol{\theta}}_{[j]} = (\widehat{\boldsymbol{\beta}}_{[j]}^\top, \widehat{\sigma}^2_{[j]}, \widehat{\delta}_{[j]})^\top$ be the maximizer of the function $Q_{[j]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{c[j]}|\mathbf{y}_{[j]}, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})]$, where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma}^2, \widehat{\delta})^\top$ is the ML estimate of $\boldsymbol{\theta}$, and the estimates $\widehat{\boldsymbol{\theta}}_{[j]}$ are obtained by using the EM algorithm based on the remaining $n - 1$ observations, with $\widehat{\boldsymbol{\theta}}$ as the initial value. To assess the influence of the i th case on $\widehat{\boldsymbol{\theta}}$, we compare $\widehat{\boldsymbol{\theta}}_{[j]}$ with $\widehat{\boldsymbol{\theta}}$. If $\widehat{\boldsymbol{\theta}}_{[j]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some

sense, then the i th case is regarded as influential. As $\widehat{\boldsymbol{\theta}}_{[j]}$ is needed for every case, the required computational effort can be quite heavy, especially when the sample size is large. Hence, to calculate the case-deletion estimate $\widehat{\boldsymbol{\theta}}_{[j]}^1$ of $\boldsymbol{\theta}$, Zhu et al. (2001) proposed the following one-step approximation based on the Q -function:

$$\widehat{\boldsymbol{\theta}}_{[j]}^1 = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[j]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{3.1}$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\dot{Q}_{[j]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[j]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ are the Hessian matrix and the gradient vector evaluated at $\widehat{\boldsymbol{\theta}}$, respectively. The Hessian matrix is an essential element in the method developed by Zhu and Lee (2001) to obtain the measures for case-deletion diagnosis. To develop the case-deletion measures, we have to obtain the elements in (3.1), $\dot{Q}_{[j]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ and $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$, which can be obtained quite easily from (2.6). The vector $\dot{Q}_{[j]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = (\dot{Q}_{[j]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top, \dot{Q}_{[j]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}_{[j]\delta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top)^\top$ has its elements given by

$$\begin{aligned} \dot{Q}_{[j]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{1}{\widehat{\sigma}^2} \mathbf{X}_{[j]}^\top (\widehat{\boldsymbol{\epsilon}}_{[j]} - \widehat{\boldsymbol{\delta}} \widehat{\mathbf{u}}_{[j]}), \\ \dot{Q}_{[j]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{1}{2\widehat{\sigma}^2} \left[-n + 1 + \frac{1}{\widehat{\sigma}^2} \left(\widehat{\boldsymbol{\epsilon}}_{[j]}^\top \widehat{\boldsymbol{\epsilon}}_{[j]} - 2\widehat{\boldsymbol{\delta}} \widehat{\mathbf{u}}_{[j]}^\top \widehat{\boldsymbol{\epsilon}}_{[j]} + \widehat{\delta}^2 \sum_{i=1, i \neq j}^n \widehat{u}_i^2 \right) \right], \\ \dot{Q}_{[j]\delta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{1}{\widehat{\sigma}^2} \left(\widehat{\mathbf{u}}_{[j]}^\top \widehat{\boldsymbol{\epsilon}}_{[j]} - \widehat{\delta} \sum_{i=1, i \neq j}^n \widehat{u}_i^2 \right). \end{aligned}$$

Moreover, the Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ and its inverse matrix are respectively given by

$$\begin{aligned} \ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \begin{pmatrix} \ddot{Q}_{11} & \ddot{Q}_{12} \\ \ddot{Q}_{21} & \ddot{Q}_{22} \end{pmatrix} \quad \text{and} \\ \ddot{Q}^{-1}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \begin{pmatrix} \ddot{Q}_{11}^{-1} + \frac{1}{K} \ddot{Q}_{11}^{-1} \ddot{Q}_{12} \ddot{Q}_{21} \ddot{Q}_{11}^{-1} & -\frac{1}{K} \ddot{Q}_{11}^{-1} \ddot{Q}_{12} \\ -\frac{1}{K} \ddot{Q}_{12} \ddot{Q}_{11}^{-1} & 1/K \end{pmatrix}, \end{aligned} \tag{3.2}$$

where $K = \ddot{Q}_{22} - \ddot{Q}_{21} \ddot{Q}_{11}^{-1} \ddot{Q}_{12}$,

$$\begin{aligned} \ddot{Q}_{11}(\widehat{\boldsymbol{\theta}}) &= -\frac{1}{\widehat{\sigma}^2} \begin{pmatrix} \mathbf{X}^\top \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{n}{\widehat{\sigma}^2} \end{pmatrix}, \quad \ddot{Q}_{12}(\widehat{\boldsymbol{\theta}}) = -\frac{1}{\widehat{\sigma}^2} \begin{pmatrix} \mathbf{X}^\top \widehat{\mathbf{u}} \\ \mathbf{0} \end{pmatrix} \quad \text{and} \\ \ddot{Q}_{22}(\widehat{\boldsymbol{\theta}}) &= -\frac{n}{\widehat{\sigma}^2} \overline{\widehat{\mathbf{u}}_2}, \end{aligned}$$

with $\widehat{\mathbf{u}} = (\widehat{u}_1, \dots, \widehat{u}_n)^\top$ and $\overline{\widehat{\mathbf{u}}_2}$ denoting the mean of the components of vector $\widehat{\mathbf{u}}_2 = (\widehat{u}_1^2, \dots, \widehat{u}_n^2)^\top$. Inspired by the classic case-deletion measures, Cook's distance and the likelihood displacement, Zhu et al. (2001) and Lee and Xu (2004) presented analogous measures based on the Q -function. These measures are:

(i) *Generalized Cook distance*: This measure is defined similar to the usual Cook distance, which is based on the genuine likelihood

$$GD_j = (\hat{\theta}_{[j]} - \hat{\theta})^\top \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}(\hat{\theta}_{[j]} - \hat{\theta}). \tag{3.3}$$

Upon substituting (3.1) into (3.3), we obtain the approximate distance given by

$$GD_j^1 = \dot{Q}_{[j]}(\hat{\theta}|\hat{\theta})^\top \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} \dot{Q}_{[j]}(\hat{\theta}|\hat{\theta}).$$

(ii) *Q-distance*: This measure of the influence of the i th case is based on the Q -distance function, similar to the likelihood distance LD_j discussed by Cook (1977), defined as

$$QD_j = 2\{Q(\hat{\theta}|\hat{\theta}) - Q(\hat{\theta}_{[j]}|\hat{\theta})\}. \tag{3.4}$$

We can calculate an approximation of the likelihood displacement QD_j by substituting (3.1) into (3.4), resulting in the following approximation QD_j^1 of QD_j :

$$QD_j^1 = 2\{Q(\hat{\theta}|\hat{\theta}) - Q(\hat{\theta}_{[j]}^1|\hat{\theta})\}.$$

If the interest is to consider the influence of the i th case on some subset of parameters, then it can be obtained quite easy as follows: Let $\theta = (\theta_1^\top, \theta_2^\top)^\top$, where here $\theta_1 = (\beta^\top, \sigma^2)^\top$ is the parameter of the usual regression model and $\theta_2 = \delta$ is the skewness parameter. From the expressions of $\dot{Q}_{[j]}(\hat{\theta}|\hat{\theta})$ and $\ddot{Q}(\hat{\theta}|\hat{\theta})$, the generalized Cook distance for the parameters $(\beta^\top, \sigma^2)^\top$ and δ can be defined as follows:

(1) Both β and σ^2 are the parameters of interest and δ is the nuisance parameter

$$GD_j^1(\beta, \sigma^2) = \dot{Q}_{[j]\theta_1}(\hat{\theta}|\hat{\theta})^\top [\mathbf{I}_{p+1}, \mathbf{0}] \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} \times [\mathbf{I}_{p+1}, \mathbf{0}]^\top \dot{Q}_{[j]\theta_1}(\hat{\theta}|\hat{\theta}), \tag{3.5}$$

where $\dot{Q}_{[j]\theta_1}(\hat{\theta}|\hat{\theta}) = (\dot{Q}_{[j]\beta}(\hat{\theta}|\hat{\theta})^\top, \dot{Q}_{[j]\sigma^2}(\hat{\theta}|\hat{\theta})^\top)^\top$, with $\theta_1 = (\beta^\top, \sigma^2)^\top$;

(2) δ is the parameter of interest and both β and σ^2 are nuisance parameters

$$GD_j^1(\delta) = \dot{Q}_{[j]\delta}(\hat{\theta}|\hat{\theta})^\top \mathbf{b}_2^\top \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} \mathbf{b}_2 \dot{Q}_{[j]\delta}(\hat{\theta}|\hat{\theta}) = (\dot{Q}_{[j]\delta}(\hat{\theta}|\hat{\theta}))^2 \mathbf{b}_2^\top \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} \mathbf{b}_2, \tag{3.6}$$

where $\dot{Q}_{[j]\delta}(\hat{\theta}|\hat{\theta})$ is the third element of $\dot{Q}_{[j]}(\hat{\theta}|\hat{\theta})$ and \mathbf{b}_2 is the $(p+2) \times 1$ vector with one at $(p+2)$ th position.

3.2 The local influence approach

Consider a perturbation vector ω varying in an open region $\Omega \in \mathbb{R}^q$. Let $\ell_c(\theta, \omega|y_c)$, $\theta \in \mathbb{R}^p$ be the complete-data log-likelihood of the perturbed model. We assume there is a ω_0 such that $\ell_c(\theta, \omega_0|y_c) = \ell_c(\theta|y_c)$ for all θ . Let $\hat{\theta}_\omega$ be

the maximizer of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{y}, \hat{\boldsymbol{\theta}}]$. Then the influence graph is defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q -displacement function defined as

$$f_Q(\boldsymbol{\omega}) = 2[Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_\omega|\hat{\boldsymbol{\theta}})].$$

Following the approach developed in [Zhu and Lee \(2001\)](#), the normal curvature $C_{f_Q, \mathbf{d}}$, of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of a unit vector \mathbf{d} can be used to summarize the local behavior of the Q -displacement function. It can be shown that

$$C_{f_Q, \mathbf{d}}(\boldsymbol{\theta}) = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_0} \mathbf{d} = 2\mathbf{d}^\top \Delta_{\boldsymbol{\omega}_0}^\top \{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1} \Delta_{\boldsymbol{\omega}_0} \mathbf{d}^\top$$

where $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\Delta_{\boldsymbol{\omega}} = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}}$. As in [Cook \(1986\)](#), the expression $-\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\omega}_0}^\top \{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1} \Delta_{\boldsymbol{\omega}_0}$ is fundamental to detect influential observations. It may be of interest to assess the influence on a subset $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^\top$. For example, one may be interested in $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$ or $\boldsymbol{\theta}_1 = \delta$. In such situations, the curvature in the direction \mathbf{d} is given by

$$C_{f_Q, \mathbf{d}}(\boldsymbol{\theta}_1) = 2\mathbf{d}^\top \Delta_{\boldsymbol{\omega}_0}^\top (-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^{-1} - B_{22}) \Delta_{\boldsymbol{\omega}_0} \mathbf{d}^\top, \tag{3.7}$$

where

$$B_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\ddot{Q}_{22}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^{-1} \end{pmatrix},$$

and $\ddot{Q}_{22}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ is obtained from the partition of $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ according to the partition of $\boldsymbol{\theta}$. A clear picture of $-\ddot{Q}_{\boldsymbol{\omega}_0}$ (a symmetric matrix) is given by its spectral decomposition

$$-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^n \lambda_k \mathbf{e}_k \mathbf{e}_k^\top,$$

where $(\lambda_1, \mathbf{e}_1), \dots, (\lambda_n, \mathbf{e}_n)$ are the eigenvalue-eigenvector pairs of the matrix $-2\ddot{Q}_{\boldsymbol{\omega}_0}$, with $\lambda_1 \geq \dots \geq \lambda_q > \lambda_{q+1} = \dots = \lambda_n = 0$ and $\mathbf{e}_1, \dots, \mathbf{e}_n$ are elements of the associated orthonormal basis. [Zhu and Lee \(2001\)](#) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for more revealing information, but this can be computationally intensive for large n . Following [Zhu and Lee \(2001\)](#) and [Lu and Song \(2006\)](#), we consider an aggregated contribution vector of all eigenvectors corresponding to nonzero eigenvalues. Starting with some notation, let $\bar{\lambda}_k = \lambda_k/(\lambda_1 + \dots + \lambda_q)$, $\mathbf{e}_k^2 = (e_{k1}^2, \dots, e_{kn}^2)$ and

$$M(0) = \sum_{k=1}^q \bar{\lambda}_k \mathbf{e}_k^2.$$

Hence, the assessment of influential cases is based on $\{M(0)_l, l = 1, \dots, n\}$ and one can obtain $M(0)_l$ via $B_{f_Q, \mathbf{u}_l} = -2\mathbf{u}_l^\top \ddot{Q}_{\boldsymbol{\omega}_0} \mathbf{u}_l / \text{tr}[-2\ddot{Q}_{\boldsymbol{\omega}_0}]$, where \mathbf{u}_l is a column vector in \mathbb{R}^n with the l th entry equal to one and all other entries zero. Refer to

Zhu and Lee (2001) for other theoretical properties of $B_{f_{Q,\mathbf{w}}}$, such as invariance under reparameterization of $\boldsymbol{\theta}$. Additionally, Lee and Xu (2004) propose to use $1/n + c^*SM(0)$ as a benchmark to regard the l th case as influential, where c^* is an arbitrary constant (depending on the real application) and $SM(0)$ is the standard deviation of $\{M(0)_l, l = 1, \dots, n\}$.

Finally, following the idea of Verbeke and Molenberghs (2000), we consider another important direction given by $\mathbf{d} = \mathbf{d}_{kq}$ which corresponds to a $q \times 1$ vector with one in the k th position and all other entries zero. In that case, the normal curvature is called the total local influence of k th observation, which is given by

$$C_k = \mathbf{d}_{kq}^\top \Delta_{\omega_0}^\top \{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1} \Delta_{\omega_0} \mathbf{d}_{kq}, \quad k = 1, \dots, q. \tag{3.8}$$

Moreover, the k th term is an influential observation if C_k is larger than the cutoff value $c^* = \sum_{k=1}^q C_k/q$.

Now, we are ready to evaluate the matrix Δ_{ω_0} under three different perturbation schemes: case weights perturbation, response perturbation and explanatory perturbation. Each perturbation scheme has the partitioned form of $\Delta_{\omega_0} = (\Delta_{\beta}^\top, \Delta_{\sigma^2}^\top, \Delta_{\delta}^\top)^\top$.

(i) *Case weight perturbation*

Let $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ be an $n \times 1$ dimensional vector with $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top \in \mathbb{R}^n$. Then, consider the following arbitrary allocation of weights for the Q -function, which may capture departures in general directions $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{k=1}^n \omega_k Q_k(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$. In this case, the matrix Δ_{ω_0} is given by

$$\Delta_{\omega_0} = \frac{1}{\sigma^2} \begin{pmatrix} \mathbf{X}^\top \mathbf{D}(\hat{\boldsymbol{\epsilon}}) \\ \frac{1}{2\sigma^2} \hat{\boldsymbol{\epsilon}}^\top \mathbf{D}(\hat{\boldsymbol{\epsilon}}) - \frac{1}{2} \mathbf{1}_n^\top \\ \mathbf{0} \end{pmatrix} + \frac{1}{\sigma^2} \begin{pmatrix} \hat{\delta} \mathbf{X}^\top \mathbf{D}(\hat{\mathbf{u}}) \\ \frac{\hat{\delta}}{\sigma^2} \hat{\boldsymbol{\epsilon}}^\top \mathbf{D}(\hat{\mathbf{u}}) - \frac{\hat{\delta}^2}{2\sigma^2} \hat{\mathbf{u}}_2^\top \\ -\hat{\boldsymbol{\epsilon}}^\top \mathbf{D}(\hat{\mathbf{u}}) + \hat{\delta} \hat{\mathbf{u}}_2^\top \end{pmatrix}, \tag{3.9}$$

where $\mathbf{D}(\mathbf{a})$ denotes the diagonal matrix of vector \mathbf{a} .

(ii) *Response perturbation*

A perturbation of the response variables $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ is introduced by replacing Y_k by $Y_{k\omega} = Y_k + \omega_k S_y$, where S_y is the standard deviation of \mathbf{Y} . The perturbed Q -function is given by $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})$, switching $Y_{k\omega}$ with Y_k and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$. Under this perturbation scheme, the vector $\boldsymbol{\omega}_0 = \mathbf{0}$ and the matrix Δ_{ω_0} is expressed as

$$\Delta_{\omega_0} = \frac{S_y}{\sigma^2} \begin{pmatrix} \mathbf{X}^\top \\ \frac{1}{\sigma^2} \hat{\boldsymbol{\epsilon}}^\top \\ \mathbf{0} \end{pmatrix} + \frac{S_y}{\sigma^2} \begin{pmatrix} \mathbf{0} \\ -\frac{\hat{\delta}}{\sigma^2} \hat{\mathbf{u}}^\top \\ \hat{\mathbf{u}}^\top \end{pmatrix}. \tag{3.10}$$

(iii) *Explanatory perturbation*

We are interested in perturbing a specific explanatory variable. Under this condition, we have the perturbed explanatory variable $\mathbf{x}_{t\omega} = \mathbf{x}_t + S_t \boldsymbol{\omega}$, $t \in \{1, \dots, p\}$,

where S_t is the standard deviation of the explanatory variable \mathbf{x}_t and $\boldsymbol{\omega}_0 = \mathbf{0}$. The perturbed Q -function is given by $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})$, switching $\mathbf{x}_t\boldsymbol{\omega}$ with \mathbf{x}_t . The $(p+2) \times n$ matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ is given by

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \frac{S_t}{\sigma^2} \begin{pmatrix} \mathbf{e}_p(t)\hat{\boldsymbol{\epsilon}}^\top - \hat{\beta}_t\mathbf{X}^\top \\ -\frac{\hat{\beta}_t}{\sigma^2}\hat{\boldsymbol{\epsilon}}^\top \\ \mathbf{0} \end{pmatrix} + \frac{S_t}{\sigma^2} \begin{pmatrix} -\hat{\delta}\mathbf{e}_p(t)\hat{\mathbf{u}}^\top \\ \frac{\hat{\beta}_t}{\sigma^2}\hat{\delta}\hat{\mathbf{u}}^\top \\ -\hat{\beta}_t\hat{\mathbf{u}}^\top \end{pmatrix}, \tag{3.11}$$

where β_t is the t th element of $\boldsymbol{\beta}$, $\mathbf{e}_p(t)$ is a $p \times 1$ vector with one in the t th position and zeros elsewhere.

3.3 Generalized leverage

Another concept that has been useful in the development of diagnostics in linear regression is the leverage. The main idea behind the concept of leverage is that of evaluating the influence of y_j on its own predicted value (see [Wei, Qu and Fung \(1998\)](#)). This influence can be represented by the derivative $\partial\hat{y}_j/\partial y_j$. Under usual linear regression model, $\partial\hat{y}_j/\partial y_j = h_{jj}$ that is the j th principal diagonal element of the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Inspired by the generalized leverage proposed by [Wei, Qu and Fung \(1998\)](#), we consider a generalized leverage matrix for models with incomplete data defined by

$$GL(\hat{\boldsymbol{\theta}}) = D_{\boldsymbol{\theta}}[-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^{-1}\ddot{Q}_{\boldsymbol{\theta},y}(\hat{\boldsymbol{\theta}}), \tag{3.12}$$

where $D_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$, $\ddot{Q}_{\boldsymbol{\theta},y}(\hat{\boldsymbol{\theta}}) = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial y^\top|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ is the Hessian matrix.

In the linear regression models under skew-normal distribution, the expectation of \mathbf{Y} is given by $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta} + \sqrt{\frac{2}{\pi}}\delta\mathbf{1}_n$, then $\hat{\mathbf{Y}} = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ is the predicted response vector. In this case, we have that $D_{\boldsymbol{\theta}} = [\mathbf{X}, \mathbf{0}_n, c\mathbf{1}_n]$, $\ddot{Q}_{\boldsymbol{\theta},y}(\hat{\boldsymbol{\theta}})$ is as in (3.10) without the term S_y and $\ddot{Q}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ is given in (3.2). We use c_0p_0/n , $p_0 = \text{tr}(GL(\hat{\boldsymbol{\theta}}))$, as a benchmark to regard the j th case as a leverage point, where c_0 is a selected constant (depending on the real application). So, observations with values $GL_{jj}(\hat{\boldsymbol{\theta}}) > c_0p_0/n$ are considered leverage points.

4 Numerical application

In this section, a simulation study and a real example are presented to illustrate the performance of the developed method. First, we carry out a numerical illustration with simulated data. Finally, we analyze real data to illustrate the usefulness of the proposed method.

Table 1 Simulated and adjusted values for model (4.1), with $n = 100$, with standard errors in parentheses

Parameters	Simulated	Adjusted		
		Model A	Model B	Model C
β_0	5.0	5.06 (0.15)	5.02 (0.15)	6.26 (0.25)
β_1	1.0	1.01 (0.21)	1.12 (0.21)	1.09 (0.27)
σ^2	0.1	0.09 (0.04)	0.08 (0.04)	0.41 (0.11)
δ	1.0	0.98 (0.12)	1.06 (0.12)	-0.57 (0.26)

Table 2 Simulated and adjusted values for model (4.1) with $n = 500$, with standard errors in parentheses

Parameters	Simulated	Adjusted		
		Model A	Model B	Model C
β_0	5.0	5.04 (0.07)	4.98 (0.07)	6.14 (0.13)
β_1	1.0	1.00 (0.09)	1.01 (0.09)	0.98 (0.12)
σ^2	0.1	0.12 (0.03)	0.09 (0.02)	0.46 (0.05)
δ	1.0	1.01 (0.06)	1.08 (0.06)	-0.41 (0.14)

4.1 Simulation study

In this section, we examine the performance of the developed method based on simulated data. We consider a dataset of size $n = 100, 500$ from linear regression models defined in (4.1). That is,

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j, \quad (4.1)$$

where the true values of the parameters are given by $\beta_0 = 5.0$, $\beta_1 = 1.0$, $\sigma^2 = 0.1$ and $\delta = 1.0$, where x_j is generated from the uniform distribution $U(0, 1)$, $j = 1, \dots, n$. For each sample, we choose an observation and disturb it in the response variable. So, we consider three linear regression models:

Model A: model without perturbation;

Model B: $y_B(ind) = a * y(ind)$;

Model C: $y_C(ind) = b * y(ind)$,

where $a > 1$ and $0 < b < 1$. For $n = 100$, $a = 1.4$, $b = 0.5$ and $ind = 58$; for $n = 500$, $a = 1.7$, $b = 0.5$ and $ind = 308$. Tables 1 and 2 shows the maximum likelihood (ML) estimates for the parameters of the three models for $n = 100$ and $n = 500$, respectively. In Model B, the estimate of β_0 is slightly lower than that of Model A, while the estimate of δ is slightly higher. However, as $\hat{y}_j = \mathbf{x}_j^\top \hat{\boldsymbol{\beta}} + \sqrt{\frac{2}{\pi}} \hat{\delta}$, the fitted line is very similar to that in Model A. In Model C, the differences of the

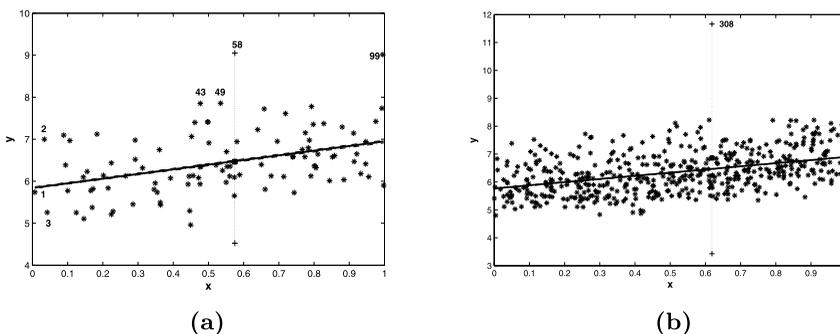


Figure 1 Simulated data with $n = 100$ (a) and $n = 500$ (b). Scatterplot with adjusted line for Models A (solid line), B (dashed line) and C (dashdot line).

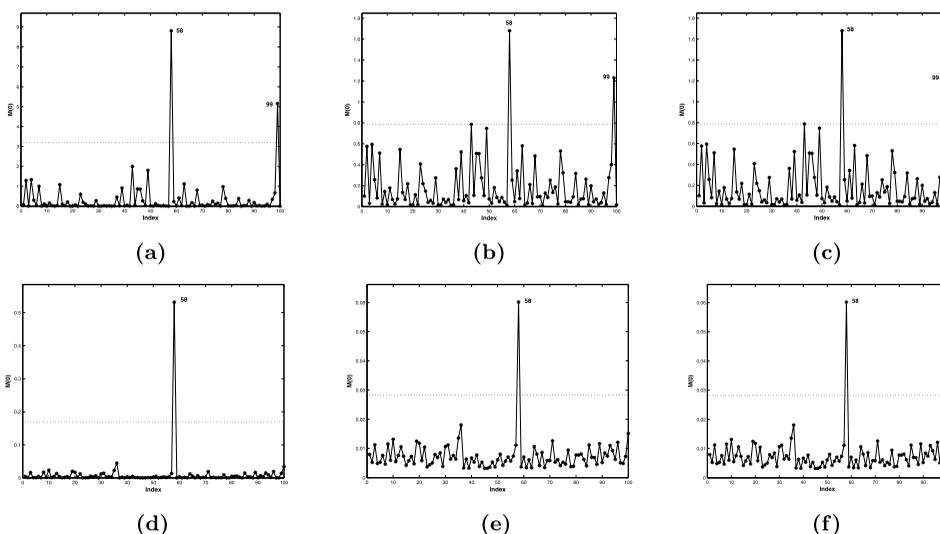


Figure 2 Simulated data with $n = 100$. Model B: Case weight perturbation (a), Response perturbation (b) and Explanatory perturbation (c). Model C: Case weight perturbation (d), Response perturbation (e) and Explanatory perturbation (f).

estimates for the parameters β_0 , σ^2 and δ are more pronounced, but the fitted line is also similar to that in Model A, as can be seen in Figure 1.

Local influence analysis under three perturbation schemes and generalized leverage measures for simulated data are described next. For $n = 100$, Model A yielded observations #49 and #99 as influential in case weight perturbation, and #43, #49 and #99 under response and explanatory variables. Observations #1, #2 and #3 were classified as leverage points; see Figure 1. Model B indicated observations #58 and #99 are influential for the three perturbation schemes; see Figures 2(a)–(c). For the three perturbation schemes under Model C, only obser-

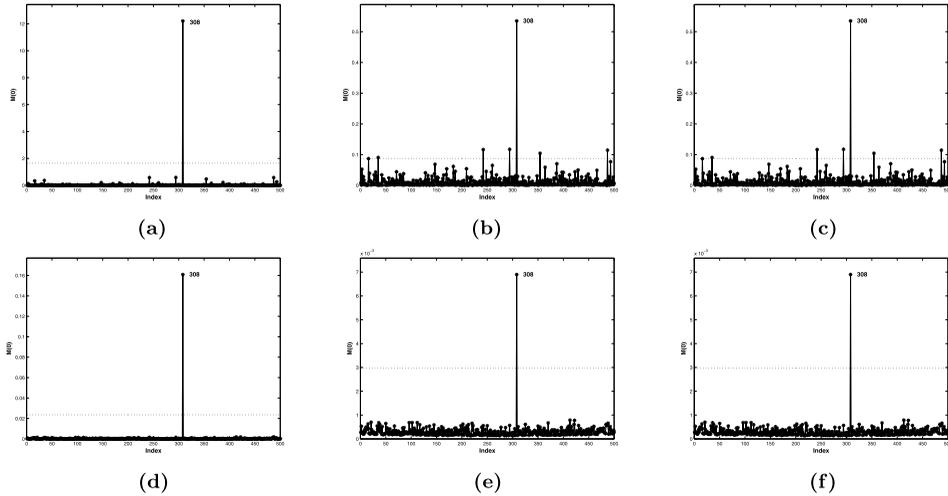


Figure 3 Simulated data with $n = 500$. Model B: Case weight perturbation (a), Response perturbation (b) and Explanatory perturbation (c). Model C: Case weight perturbation (d), Response perturbation (e) and Explanatory perturbation (f).

vation #58 is detected; see Figures 2(d)–(f). Following the same way, we have similar results for $n = 500$, with observation #308 as influential, as presented in Figures 3(a)–(c) and 3(e)–(f). The results do not change for large samples size, and are omitted.

4.2 Real dataset

A dataset on life quality of women with breast cancer obtained by the Center for Integral Attention to Women’s Health (State University of Campinas, Brazil) was adjusted to the skew-normal regression model and the diagnostic analysis approach was applied. The index of life quality was evaluated by the Medical Outcome Study 36-item Short-Form Health Survey (SF-36) questionnaire. These indexes condense eight components into two: a physical component summary and a mental component summary. Conde et al. (2005) analyzed this dataset, evaluating the associated factors of the life quality of women with breast cancer.

The response variable is the physical component summary of the life quality index (pcs) and the explanatory variables are the indicator variable *dizziness* and the body mass index (*bmi*) of the individual. In this study, we discuss these data based on the skew-normal linear regression model

$$pcs_j = \beta_0 + \beta_1 * dizziness_j + \beta_2 * bmi_j + e_j, \quad j = 1, \dots, 97, \quad (4.2)$$

where e_j are independent such that $e_j \sim SN(0, \sigma^2, \delta)$.

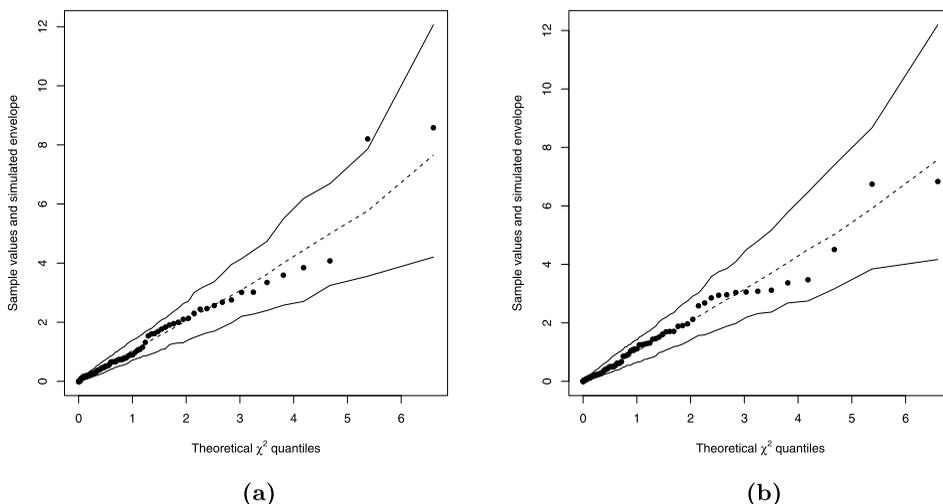


Figure 4 Simulated envelope for quality of life dataset. (a) Adjusted under normal linear model and (b) Adjusted under skew-normal linear model.

Table 3 ML estimates of the life quality dataset for normal and skew-normal linear regression models

Parameter	SN		Normal	
	Estimate	SE	Estimate	SE
β_0	68.57	4.64	61.29	4.75
β_1	-7.62	1.91	-7.92	1.91
β_2	-0.35	0.17	-0.43	0.16
σ^2	25.47	14.76	79.13	11.73
δ	-11.99	2.33	-	-
Log-likelihood	-346.36		-348.14	
AIC	702.73		704.27	

4.2.1 *Goodness-of-fit and estimation:*. The ML estimates of parameters and their corresponding standard deviations (calculated using the observed information matrix) are reported in Table 3, which shows that the estimates of β in both models are similar. In contrast, the estimates of the parameter σ^2 are different. We have constructed the QQ-plots and envelopes in Figure 4 (lines represent the 5th percentile, the mean and the 95th percentile of 100 simulated points for each observation) based on $r_j^2 = \frac{(y_j - \mathbf{x}_j^T \hat{\beta})^2}{\sigma^2 + \hat{\delta}^2}$, which has an approximate χ_1^2 distribution. The random variable r_j also enables us to check the model and detect the presence of outlying observations. The simulated envelope graph plotted to validate the skew-normal regression model indicates no points lie outside the confidence bounds (Figure 4(b)).

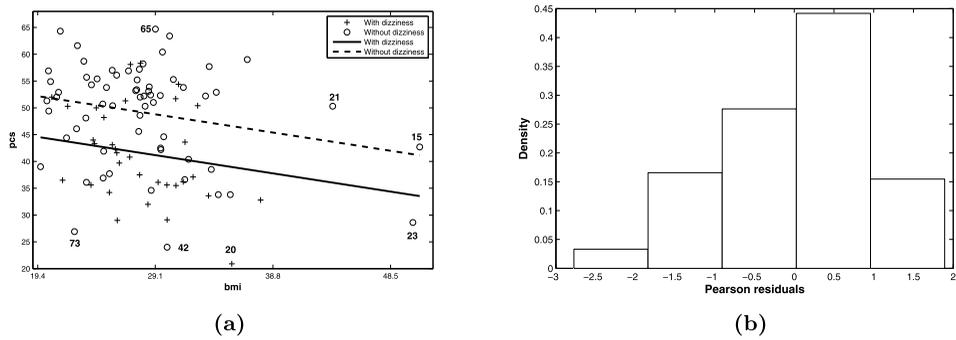


Figure 5 Quality of life dataset. (a) Scatter plot and adjusted lines, (b) Pearson residuals

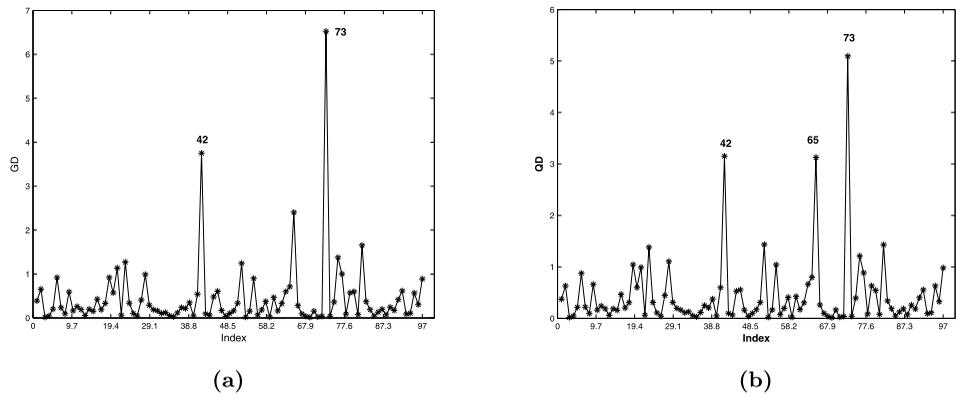


Figure 6 Quality of life dataset. (a) Generalized Cook distance, (b) Q-Distance

On the other hand, the simulated envelope for the normal linear regression model (Figure 4(a)) indicates some exterior points.

Figure 5 displays the scatter plot of data and the adjusted lines for each status of dizziness, showing a negative relationship between *pcs* and *bmi*. Figure 5(b) shows the histogram of the model waste set (4.2), indicating the presence of asymmetric data. Based on the normal approximation of the ML estimates, the 95% confidence interval for δ (CI: $\hat{\delta} \pm 1.96SE$) is $[-16.56, -7.42]$, which does not include the value zero. That is, a skew-normal model is more suitable than a normal linear model. A similar result is obtained when testing the null hypothesis $H_0 : \delta = 0$ against the alternative $H_1 : \delta \neq 0$ when we use the likelihood-ratio (LR) statistic, which has a χ^2_1 distribution under H_0 . In this case, $LR = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_0)) = 3.88$, with a *p*-value of 0.05.

4.2.2 Influence diagnostics analysis. First, we identify outlying observations under the fitted model based on case-deletion measures, the Q-distance and generalized Cook distance. The results are reported in Figure 6. We note from these

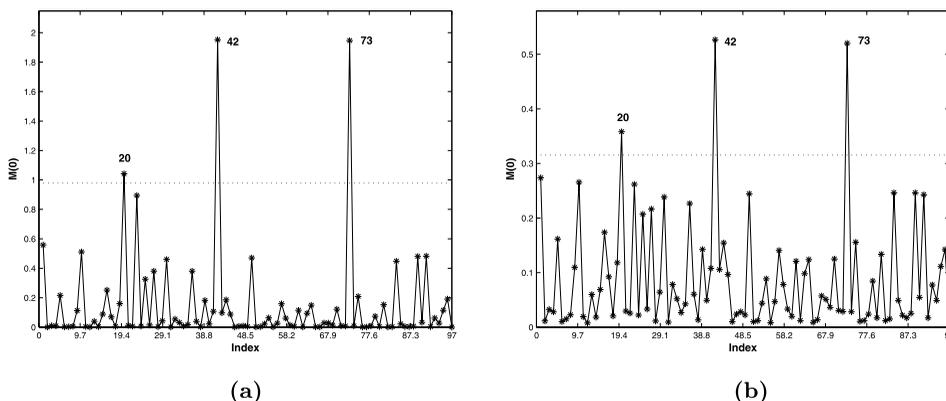


Figure 7 Quality of life dataset. Diagnostics of case weight perturbation (a) and response variable perturbation (b) (benchmark with $c^* = 3$).

figures that observations #42 and #73 are potentially influential on the parameter estimates. These refer to women without dizziness with low life quality index. Observation #65 shows a slight influence and represents a woman with high life quality index. We also used the generalized Cook distance when the interest is a subset of the parameters. All results were similar, so they are not shown here to save space.

Second, we conducted the local influence diagnostics analysis to detect outlying observations, using the perturbation schemes discussed previously and the generalized leverage. For the perturbation schemes, we obtained the values of $M(0)$ and the figures present the index graphs of $M(0)$. The horizontal lines delimit the Lee and Xu (2004) benchmark for $M(0)$, with $c^* = 3$. So, under the perturbation schemes, observations #20, #42 and #73 are potentially influential on the parameter estimates. These results are reported in Figure 7 and Figure 8(a). Note that observations #42 and #73 are potentially influential under all perturbation schemes, which correspond to the smallest values of pcs for women without dizziness. On the other hand, observation #20 is influential under case weight perturbation and response perturbation. This observation is a woman with smaller values of pcs and with dizziness. Moreover, we performed diagnostics analysis based on C_i measures defined in (3.8), and the influential observations were the same observations #42 and #73. We then used the generalized leverage and observations #15, #21 and #23 were indicated as influential, as can be seen in Figure 8(b). These observations are more distant from the mass data (right), which are women without dizziness and high bmi . It is important to note that these observations were not detected using the perturbation schemes, so the usefulness of the generalized leverage can be seen.

Finally, we also discuss the influence of some observations on the estimates of β and δ based on $C_{f_Q, \mathbf{d}}(\beta)$ and $C_{f_Q, \mathbf{d}}(\delta)$, as defined in (3.7). For the three

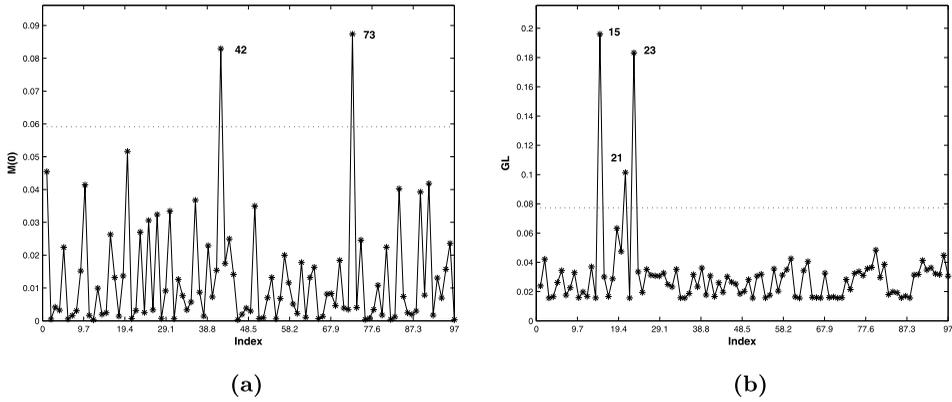


Figure 8 *Quality of life dataset. (a) Diagnostic of bmi variable perturbation (benchmark with $c^* = 3$) and (b) Generalized leverage.*

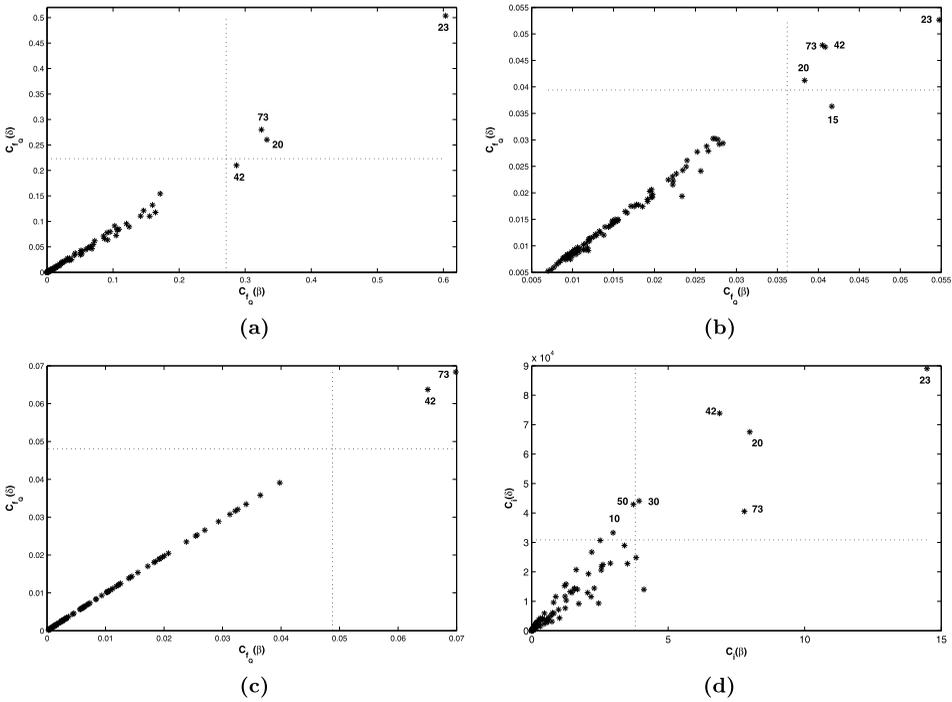


Figure 9 *Quality of life dataset. Diagnostic of influence in β and δ (benchmark with $c^* = 3$): (a) Case weight perturbation, (b) Response variable, (c) bmi variable and (d) C_i for case weight.*

perturbation schemes, Figures 9(a)–(c) show the scatter plots of $C_{f_{Q,a}}(\beta)$ versus $C_{f_{Q,a}}(\delta)$. From these figures, we observe that #73 is potentially influential on both parameter estimates β and δ , regardless of the perturbation scheme. Moreover,

Figure 9(d) shows that observation #23 is found to have relatively large $C_i(\beta)$ and $C_i(\delta)$, which corresponds to the woman without dizziness and high *bmi*.

5 Conclusion

In this paper, we have provided a diagnostics analysis for skew-normal regression models based on the approaches proposed by Zhu et al. (2001) and Zhu and Lee (2001). We have proposed a procedure for computing case-deletion measures and local influence diagnostics on the basis of the conditional expectation of the complete-data log-likelihood function in relation to the EM algorithm and also we have considered a generalized leverage. Case weight, response and explanatory perturbations are considered. Explicit expressions are obtained for the Hessian matrix $\hat{Q}(\hat{\theta}|\hat{\theta})$ and for the matrix Δ_{ω_0} under different perturbation schemes. The generalized leverage developed here is a necessary supplement to the results obtained based on the perturbation schemes. This diagnostic measure helps to detect observations that were not detected by perturbation schemes.

References

- Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Comput. Statist. Data Anal.* **55**, 353–365. [MR2736560](#)
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171–178. [MR0808153](#)
- Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Stat.* **32**, 159–188. [MR2188669](#)
- Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61**, 579–602.
- Conde, D. M., Pinto-Neto, A. M., Cabello, C., Santos-Sá, D., Costa-Paiva, C. and Martinez, E. S. (2005). Quality of life in Brazilian breast cancer survivors age 45–65 years: Associated factors. *Breast J.* **11**, 425–432.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics* **19**, 5–18.
- Cook, R. D. (1986). Assessment of local influence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **48**, 133–169. [MR0867994](#)
- Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Boca Raton, FL: Chapman & Hall/CRC. [MR0675263](#)
- Dagne, G. A. (2016). Bayesian segmental growth mixture Tobit models with skew distributions. *Comput. Statist.* **31**, 121–137.
- Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **39**, 1–38.
- Ferreira, C. S., Lachos, V. H. and Bolfarine, H. (2015). Inference and diagnostics in skew scale mixtures of normal regression models. *J. Stat. Comput. Simul.* **85**, 517–537. [MR3275462](#)
- Galea-Rojas, M., Paula, G. A. and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *Statistician* **46**, 71–79.

- Hill, M. A. and Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrics* **38**, 377–396.
- Lachos, V. H., Montenegro, L. C. and Bolfarine, H. (2008). Inference and influence diagnostics for skew-normal null intercept measurement errors models. *J. Stat. Comput. Simul.* **78**, 395–419. [MR2432253](#)
- Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. and Montenegro, L. C. (2007). Likelihood based inference for multivariate skew-normal regression models. *Comm. Statist. Theory Methods* **36**, 1769–1786.
- Lange, K. L., Little, R. and Taylor, J. (1989). Robust statistical modeling using t distribution. *J. Amer. Statist. Assoc.* **84**, 881–896. [MR1134486](#)
- Lee, S. X. and McLachlan, G. J. (2013). On mixtures of skew normal and skew t -distributions. *Adv. Data Anal. Classif.* **7**, 241–266. [MR3103965](#)
- Lee, S. Y. and Xu, L. (2004). Influence analysis of nonlinear mixed-effects models. *Comput. Statist. Data Anal.* **45**, 321–341. [MR2045635](#)
- Liu, S. Z. (2000). On local influence for elliptical linear models. *Statist. Papers* **41**, 211–224.
- Lu, B. and Song, X.-Y. (2006). Local influence of multivariate probit latent variable models. *J. Multivariate Anal.* **97**, 1783–1798. [MR2298889](#)
- Massuia, M. B., Cabral, C. R. B., Matos, L. A. and Lachos, V. H. (2015). Influence diagnostics for Student- t censored linear regression models. *Statistics* **49**, 1074–1094.
- Rodríguez, C. L. B. and Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Braz. J. Probab. Stat.* **21**, 141–163. [MR2397043](#)
- Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate distributions with applications to Bayesian regression models. *Canad. J. Statist.* **31**, 129–150. [MR2016224](#)
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Wei, B. C., Qu, Y. Q. and Fung, W. K. (1998). Generalized leverage and its applications. *Scand. J. Stat.* **25**, 25–37.
- Zeller, C. B., Lachos, V. H. and Vilca, F. V. (2014). Influence diagnostics for Grubbs's model with asymmetric heavy-tailed distributions. *Statist. Papers* **55**, 671–690. [MR3227546](#)
- Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63**, 111–126.
- Zhu, H., Lee, S., Wei, B. and Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika* **88**, 727–737.

C. S. Ferreira
 Departamento de Estatística
 Universidade Federal de Juiz de Fora
 CEP: 36036-900, Juiz de Fora-MG
 Brazil
 E-mail: clecio.ferreira@ufjf.edu.br

F. Vilca
 Departamento de Estatística
 Universidade Estadual de Campinas
 CEP: 13083-859, Campinas-SP
 Brazil
 E-mail: fily@ime.unicamp.br

H. Bolfarine
 Departamento de Estatística
 Universidade de São Paulo
 CEP: 05508-090, São Paulo-SP
 Brazil
 E-mail: hbolfar@ime.usp.br