DIAGNOSTICS ANALYSIS FOR SKEW NORMAL REGRESSION MODELS

CLÉCIO DA SILVA FERREIRA

Centro de Estudos de Opinião Pública, Universidade Estadual de Campinas - UNICAMP, Campinas, SP, Brasil

HELENO BOLFARINE†

Departamento de Estatística, Universidade de São Paulo, São Paulo, SP, Brazil

FILIDOR EDILFONSO VILCA LABRA

Departamento de Estatística, Universidade Estadual de Campinas - UNICAMP, Campinas, SP, Brasil

SUMMARY

The skew normal model (Azzalini, 1985) is being used successfully in various statistical applications. The main purpose of this paper is to consider local influence analysis, which is a well-recognized important step of data analysis. In this paper local influence measures are developed via Zhu and Lee's (2001) approach, that is closely related to the EM-algorithm. The diagnostic measures derived under this approach are invariant under reparameterization. Some useful perturbation schemes are discussed. The Cook's approach is also calculated and compared with Zhu and Lee's approach. Results that are obtained from analysis of a real data example are presented to illustrate the developed methodologies.

Keywords: Skew-normal distribution; Cook's approach; Zhu and Lee's approach, EM algorithm.

1 INTRODUCTION

The normal distribution and normal linear regression models have played an essential role in statistics. However, there is indication that the normality assumption does not work well in certain situations, being specially sensitive to the presence of extreme (outlying) observations. Alternative distributions have been considered, some of which are the Student-t, logistic, exponential-power and contaminated normal, which are particular members of the symmetric distributions. See, for

† *Address for correspondence:* Departamento de Estatística, IME, Universidade de São Paulo, Caixa Postal 66281, CEP 05315 970, São Paulo-SP-Brazil. E-mail: hbolfar@ime.usp.br.

example, Fang et al. (1990) and Fang and Zhang (1990). This class of symmetric distributions contains many distributions with heavier tails than the normal distribution.

Linear symmetrical models have been investigated by various authors. For example, Lange et al. (1989) present an approach for modeling student-t distributions. Galea et al. (1997) and Liu (2000) discuss diagnostics methods for multivariate symmetrical linear regression models, while Galea et al. (2003) present some diagnostic studies for univariate symmetrical models.

Although the class of symmetric distributions represents a better alternative proposal than that of the normal distribution, it is not appropriate in situations where the sample distribution is asymmetrical. For example, Hill and Dixon (1982) discuss and present examples with asymmetrical structures.

From a practical viewpoint, many authors have used transformation of variables to achieve normality, and in many situations their results are satisfactory. However, Azzalini and Capitanio (1999) have pointed out some problems. They point out, among other problems, that the transformed variables are more difficult to deal with because of interpretation problems, specially when each variable is transformed using a different function. In many situations it is not possible to go back to the original parameterization.

A parametric family of asymmetric distributions analytically tractable and able to accommodate practical values of skewness and kurtosis and include the normal distribution as a special case was introduced by Azzalini (1985, 1986) and is known and the shew-normal distribution.

Extensions, further properties and applications of the skew-normal distribution can also be found in Genton (2004).

Based on the work by Azzalini (1985), many authors have considered the skew-normal distributions and applied it in different areas such as economics, finance, oceanography, engineering and biomedical sciences, among other. Additional results on skew-normal distributions and applications can be found in Liseo and Loperfido (2003), Genton et al. (2001), Capitanio et al. (2003).

Recently, Azzalini (2005) presented a discussion on skew-normal distributions with applications in regression models; Bauwens and Laurent (2005) consider a study on autocorrelated regression models; Vilca-Labra and Leiva-Sánchez (2005) consider an asymmetric extension of the Birnbaum-Saunders distribution. In the context of the linear model, Lachos et al. (2008) consider an application of diagnostics in the linear mixed models. Also, the same authors consider a study with the skew-normal distribution proposed by Sahu et al. (2003).

The advantages of Sahu's et al. (2003) skew-normal distribution is that it facilitates the implementation of algorithms that are used to obtain the maximum likelihood estimators (MLE) and the variance of the distribution is finite.

After the model is fitted, the influence diagnostic is an important step in the analysis of a data set since it can indicate bad model fitting or the presence of influential observations. This kind of analysis has received a great deal of attention due to the paper by Cook (1977). Typically, the analysis is based on the case-weight perturbation scheme, where the case (observation) is either deleted or retained. This approach allows the assessment of the individual impact cases in the estimation process (see, for example, Cook and Weisberg, 1982). However, deletion can be viewed as one of the many ways of perturbing a model formulation. In Cook (1986) it is proposed a method of assessing the local influence of minor perturbations of a statistical model. Since then many works has been written with dealing with local influence studies.

Recently, Lachos et al. (2007) apply the local influence method to the Grubbs's model; Lachos et al.(2008) present a study of inference and local influence in skew-normal null intercept measurement error models.

However, no application of local influence has been considered for the skew-normal regression models. Thus, the main objective of this paper is to apply the approach of local influence to the univariate regression models under the Sahu's et al. (2003) skew-normal distribution. Several perturbation schemes are considered. With this we hope to expand some results in Cook (1986) for normal regression models.

Now, we present a small review of the univariate Sahu's et al. (2003) skew-normal distribution. This new class of distributions is obtained by using transformation and conditioning. The developments obtained are applied in a Bayesian regression model.

To describe this class, let ϕ and Φ be the probability density function (pdf) and the cumulative distribution function (cdf), respectively of the $N(0,1)$. We say that a random variable Y has a skew-normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter δ, if the pdf of Y is given by

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

We describe this by using the notation $Y \sim SN(\mu,\sigma^2,\delta)$. For $\delta = 0$ the pdf in (.) corresponds to the normal distributions. The mean and variance of $SN(\mu,\sigma^2,\delta)$ are given by

$$E(Y_j) = \mu + c\delta \quad \text{and} \quad Var(Y_j) = \sigma^2 + (1-c^2)\delta^2, \tag{1.2}$$

where $c = \sqrt{2/\pi}$. Alternatively, it is possible to describe this distribution by using the stochastic representation given by $Y \stackrel{d}{=} \delta|X_0| + X_1$, where $X_0 \sim N(0,1)$ and is independent of $X_1 \sim N(\mu,\sigma^2)$. The notation "$\stackrel{d}{=}$" means that both variables have the same distribution.

If $Y \sim SN(\mu,\sigma^2,\delta)$, the square of the Mahalanobis distance D, is such that

$$D = \frac{(Y-\mu)^2}{\sigma^2+\delta^2} \sim \chi_1^2, \tag{1.3}$$

that is, follows a chi-square distribution with 1 degrees of freedom.

For skew-normal distribution the Fisher information matrix for $\boldsymbol{\theta} = (\mu, \sigma^2, \delta)^\top$, can be written as

$$I_F(\boldsymbol{\theta}) = [I_{\gamma\psi}], \quad \boldsymbol{\gamma}, \boldsymbol{\psi} = \mu, \sigma^2, \delta. \tag{1.4}$$

where the components are given by

$$\begin{aligned}
I_{\mu\mu} &= \frac{1}{\sigma^2+\delta^2}\left[1+\frac{\delta^2}{\sigma^2}a_{02}(\delta/\sigma)\right],\\
I_{\mu\sigma^2} &= \frac{\delta}{(\sigma^2+\delta^2)^{3/2}}\left[\frac{c}{(\sigma^2+\delta^2)^{1/2}}-\frac{2\sigma^2+\delta^2}{2\sigma^3}\left(a_{01}(\delta/\sigma)-\frac{\delta}{\sigma}a_{12}(\delta/\sigma)-\frac{\delta^2}{\sigma^2}a_{21}(\delta/\sigma)\right)\right],\\
I_{\mu\delta} &= \frac{1}{(\sigma^2+\delta^2)^{3/2}}\left[2c\frac{\delta^2}{(\sigma^2+\delta^2)^{1/2}}-\delta\left(\frac{\delta}{\sigma}a_{21}(\delta/\sigma)+a_{12}(\delta/\sigma)\right)+\sigma a_{01}(\delta/\sigma)\right],\\
I_{\sigma^2\sigma^2} &= \frac{1}{(\sigma^2+\delta^2)^2}\left[\frac{1}{2}+\frac{(2\sigma^2+\delta^2)^2\delta^2}{4\sigma^6}a_{22}(\delta/\sigma)\right],\\
I_{\sigma^2\delta} &= \frac{\delta}{(\sigma^2+\delta^2)^2}\left[1-\frac{(2\sigma^2+\delta^2)}{2\sigma^2}a_{22}(\delta/\sigma)\right],\\
I_{\delta\delta} &= \frac{1}{(\sigma^2+\delta^2)^2}\left[2\delta^2+\frac{\sigma^4}{\delta^2}a_{22}(\delta/\sigma)\right],
\end{aligned}$$

where c is as in (.) and $a_{hk}(x) = E[Z^h W_\Phi^k(xZ)]$. The values of the $a_{01}(\cdot)$ and $a_{21}(\cdot)$ are given by

$$a_{01}(x) = c(x^2+1)^{-1/2} \quad \text{and } a_{21}(x) = c(x^2+1)^{-3/2}.$$

The other values $a_{12}(\cdot)$, $a_{22}(\cdot)$ are obtained by using the approximation given in Rodríguez (2005).

2 THE SKEW-NORMAL LINEAR REGRESSION MODEL

In this section we present the linear regression model under skew-normal distribution, which is an extension of the ordinary normal regression model. In this paper we consider the regression models under Sahu's et al. (2003) skew-normal distributions. Hence, we consider that relating response and covariates we have the model

$$Y_j = \beta_1 + \sum_{i=2}^{p}\beta_i x_{ji} + \delta z_j + \epsilon_j, \tag{2.1}$$

where $\epsilon_j \sim N(0, \sigma^2)$ and $z_j \sim HN_1(0,1)$, $j = 1, \ldots, n$ all independent, where $HN_1(0,1)$ denotes the univariate standardized half-normal distribution. By Sahu et al.(2003), it follows that $\delta z_j + \epsilon_j \sim SN(0, \sigma^2; \delta)$, $j = 1, ...n$. From the properties of the Skew-normal distributions, it follows that $Y_j \sim SN(\boldsymbol{x}_j^\top\boldsymbol{\beta}, \sigma^2, \delta)$. Thus, we have n independent observed one-dimensional response variables Y_j with $Y_j \sim SN(\boldsymbol{x}_j^\top\boldsymbol{\beta}, \sigma^2, \delta)$, where $\boldsymbol{x}_j^\top = (1, x_{j2}, ..., x_{jp})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^\top$. Thus, the density function of Y_j is given by

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

where ϕ and Φ are as in Section 1, that is, the density function and the cumulative distribution function, respectively, of the $N(0,1)$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \delta)$. The mean and variance of Y_j are given by

$$E(Y_j) = \boldsymbol{x}_j^\top\boldsymbol{\beta} + c\delta \quad \text{and} \quad Var(Y_j) = \sigma^2 + (1-c^2)\delta^2, \tag{2.3}$$

where c is as (1.2). The log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $Y_1, ..., Y_n$ is given by $l(\boldsymbol{\theta}) = \sum_{j=1}^n l_j(\boldsymbol{\theta})$, where

$$l_j(\boldsymbol{\theta}) = \log 2 - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2+\delta^2) - \frac{1}{2}A_j + \log\Phi(B_j), \tag{2.4}$$

with $A_j = \frac{1}{\sigma^2+\delta^2}(y_j - \mathbf{X}_j^\top\boldsymbol{\beta})^2$ and $B_j = \frac{\delta}{\sigma\sqrt{\sigma^2+\delta^2}}(y_j - \mathbf{X}_j^\top\boldsymbol{\beta})$. The log-likelihood above may be written as

$$l(\boldsymbol{\theta}) = n\log 2 - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2+\delta^2) - \frac{1}{2(\sigma^2+\delta^2)}Q(\boldsymbol{\beta}) + \sum_{j=1}^n \log\Phi(B_j), \tag{2.5}$$

where $Q(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$, $\boldsymbol{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$ and B_j is as in (2.4).

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

$$U(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = (U(\boldsymbol{\beta})^\top, U(\sigma^2), U(\delta))^\top, \tag{2.6}$$

where the elements of the $U(\boldsymbol{\theta})$ are given by

$$U(\boldsymbol{\beta}) = \frac{1}{\sigma^2+\delta^2}\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) - \frac{\delta}{\sigma(\sigma^2+\delta^2)^{1/2}}\boldsymbol{X}^\top\boldsymbol{a}, \tag{2.7}$$

$$U(\sigma^2) = -\frac{1}{2}\frac{n}{\sigma^2+\delta^2} + \frac{1}{2(\sigma^2+\delta^2)^2}Q(\boldsymbol{\beta}) - \frac{1}{2}\frac{2\sigma^2+\delta^2}{\sigma^2(\sigma^2+\delta^2)^{3/2}}\frac{\delta}{\sigma}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{a}, \tag{2.8}$$

$$U(\delta) = -\frac{n\delta}{\sigma^2+\delta^2} + \frac{\delta}{(\sigma^2+\delta^2)^2}Q(\boldsymbol{\beta}) + \frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{a}, \tag{2.9}$$

where $Q(\boldsymbol{\beta})$ is as in (2.5) and $\boldsymbol{a} = (a_1, ..., a_n)^\top$, with $a_j = W_\Phi(B_j) = \phi(B_j)/\Phi(B_j)$, $j = 1, ..., n$. A joint iterative procedure to obtain the maximum likelihood estimates of $\boldsymbol{\beta}$, σ^2 and δ is given by

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top[\boldsymbol{Y} - \frac{\delta^{(m)}}{\sigma^{(m)}}(\sigma^{2(m)} + \delta^{2(m)})^{1/2}\boldsymbol{a}(\boldsymbol{\theta}^{(m)})] \tag{2.10}$$

and

$$(\sigma^{2(m+1)}, \delta^{(m+1)})^\top = \text{argmax}_{\sigma^2,\delta}\left[l(\boldsymbol{\beta}^{(m+1)}, \sigma^2, \delta)\right], \tag{2.11}$$

for $m = 0, 1, 2....$ To perform the maximization with the last equation above we consider a multivariate secant method (see Dennis and Schnabel, 1996), so that the score functions used with this maximization algorithm are given by (2.8) and (2.9). Initial values $\boldsymbol{\beta}^{(0)}$, ${\sigma^2}^{(0)}$ and $\delta^{(0)}$ are required to the procedure given by (2.10) and (2.11). Alternatively, since the parametrization of Sahu et al.

(2003) is equivalent to the parametrization used in Azzalini and Capitanio (1999), then we also can use the EM algorithm in this model.

2.1 The observed information matrix

Considering $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \delta)^\top$, the observed information matrix is given by

$$\boldsymbol{L} = -\left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\psi}^\top}\right],\ \boldsymbol{\gamma}, \boldsymbol{\psi} = \boldsymbol{\beta}, \sigma^2, \delta, \tag{2.12}$$

where

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2+\delta^2}\boldsymbol{X}^\top\boldsymbol{X} + \frac{\delta^2}{\sigma^2(\sigma^2+\delta^2)}\boldsymbol{X}^\top D(\boldsymbol{b})\boldsymbol{X} \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma^2} &= \frac{1}{(\sigma^2+\delta^2)^{3/2}}\left[-\frac{\mathbf{X}^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{(\sigma^2+\delta^2)^{1/2}} + \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3}\left(\frac{\delta}{\sigma}\frac{\mathbf{X}^\top D(\mathbf{b})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{(\sigma^2+\delta^2)^{1/2}} + \mathbf{X}^\top\mathbf{a}\right)\right] \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\delta} &= -\frac{1}{(\sigma^2+\delta^2)^{3/2}}\mathbf{X}^\top\left[\frac{2\delta}{(\sigma^2+\delta^2)^{1/2}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) + \sigma\mathbf{a} + \frac{\delta}{(\sigma^2+\delta^2)^{1/2}}D(\mathbf{b})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right] \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2} &= \frac{1}{(\sigma^2+\delta^2)^2}\left[\frac{n}{2} - \frac{Q(\boldsymbol{\beta})}{\sigma^2+\delta^2} + \frac{\delta(8\sigma^4+8\sigma^2\delta^2+3\delta^4)}{4\sigma^5(\sigma^2+\delta^2)^{1/2}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{a} + \frac{\delta^2(2\sigma^2+\delta^2)^2}{4\sigma^6(\sigma^2+\delta^2)}Q_b(\boldsymbol{\beta})\right] \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\sigma^2\partial\delta} &= \frac{1}{(\sigma^2+\delta^2)^2}\left[n\delta - \frac{2\delta}{\sigma^2+\delta^2}Q(\boldsymbol{\beta}) + \frac{\delta^2-2\sigma^2}{2\sigma(\sigma^2+\delta^2)^{1/2}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{a} - \frac{\delta(2\sigma^2+\delta^2)}{2\sigma^2(\sigma^2+\delta^2)}Q_b(\boldsymbol{\beta})\right] \\
\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\delta^2} &= \frac{1}{(\sigma^2+\delta^2)^2}\left[n(\delta^2-\sigma^2) + \frac{\sigma^2-3\delta^2}{\sigma^2+\delta^2}Q(\boldsymbol{\beta}) - \frac{3\sigma\delta}{(\sigma^2+\delta^2)^{1/2}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{a} + \frac{\sigma^2}{\sigma^2+\delta^2}Q_b(\boldsymbol{\beta})\right]
\end{aligned}$$

where $\mathbf{a}$ is as in (2.6), $Q_b(\boldsymbol{\beta}) = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top D(\mathbf{b})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ and $\mathbf{b}^\top = (b_1, \cdots, b_n)$, with $b_j = W'_\Phi(B_j) = -W_\Phi(B_j)(B_j + W_\Phi(B_j))$. In particular, for the normal model, the observed information matrix for $\boldsymbol{\beta}$ and σ^2 coincides with the expression given in Cook (1986).

Furthermore, from (2.6) and after some algebraic manipulations it follows that the information matrix denoted by $I_F(\boldsymbol{\theta})$ is given by

$$I_F(\boldsymbol{\theta}) = \begin{pmatrix} I_{\boldsymbol{\beta}\boldsymbol{\beta}} & I_{\boldsymbol{\beta}\sigma^2} & I_{\boldsymbol{\beta}\delta} \\ . & I_{\sigma^2\sigma^2} & I_{\sigma^2\delta} \\ . & . & I_{\delta\delta} \end{pmatrix}, \tag{2.13}$$

where the components are given by

$$
\begin{aligned}
I_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{1}{\sigma^2+\delta^2}\left[1+\frac{\delta^2}{\sigma^2}a_{02}(\delta/\sigma)\right]\boldsymbol{X}^\top\boldsymbol{X},\\
I_{\boldsymbol{\beta}\sigma^2} &= \frac{\delta}{(\sigma^2+\delta^2)^{3/2}}\left[\frac{c}{(\sigma^2+\delta^2)^{1/2}}-\frac{2\sigma^2+\delta^2}{2\sigma^3}\left(a_{01}(\delta/\sigma)-\frac{\delta}{\sigma}a_{12}(\delta/\sigma)-\frac{\delta^2}{\sigma^2}a_{21}(\delta/\sigma)\right)\right]\boldsymbol{X},\\
I_{\boldsymbol{\beta}\delta} &= \frac{\sigma}{(\sigma^2+\delta^2)^{3/2}}\left[2c\frac{\delta^2}{(\sigma^2+\delta^2)^{1/2}}-\delta\left(\frac{\delta}{\sigma}a_{21}(\delta/\sigma)+a_{12}(\delta/\sigma)\right)+\sigma a_{01}(\delta/\sigma)\right]\boldsymbol{X},\\
I_{\sigma^2\sigma^2} &= \frac{n}{(\sigma^2+\delta^2)^2}\left[\frac{1}{2}+\frac{(2\sigma^2+\delta^2)^2\delta^2}{4\sigma^6}a_{22}(\delta/\sigma)\right],\\
I_{\sigma^2\delta} &= \frac{n\delta}{(\sigma^2+\delta^2)^2}\left[1-\frac{(2\sigma^2+\delta^2)}{2\sigma^2}a_{22}(\delta/\sigma)\right],\\
I_{\delta\delta} &= \frac{n}{(\sigma^2+\delta^2)^2}\left[2\delta^2+\sigma^2a_{22}(\delta/\sigma)\right],
\end{aligned}
$$

with $c=\sqrt{2/\pi}$ and $a_{hk}(.)$ are as in (1.4).

In particular under the normal distribution, $\delta=0$, so that the information matrix for $\boldsymbol{\beta}$ and σ^2 reduces to the matrix

$$
I_F(\boldsymbol{\beta},\sigma^2)=\begin{pmatrix}\frac{1}{\sigma^2}\boldsymbol{X}^\top\boldsymbol{X} & \mathbf{0}\\ \mathbf{0} & \frac{n}{2\sigma^4}\end{pmatrix}, \tag{2.14}
$$

which has the same form as in Cook (1986).

3 Local influence diagnostics

Case deletion is a popular way to asses the individual impact of cases on the estimation process. This approach can be regarded as a global measure of influence. An alternative methodology for the identification of groups of cases which may require some concern is local influence which is based on differential geometry instead of complete deletion. It employs a differential comparison of parameter estimates before and after perturbation to data values or model assumptions. This first approach was proposed by Cook (1986) there the likelihood displacement is used as the metric to assess the local influence. Recently, inspired by the basic idea of the EM-algorithm, Zhu and Lee (2001) proposed a unified method for local influence analysis of general statistical models with missing data on the basis of the Q- displacement function that we define further on.

Cook's approach:

Let $l(\boldsymbol{\theta})$ denote the log-likelihood function given in (2.5), $\boldsymbol{\omega}$, $q\times 1$, the perturbation introduced in the model, where $\boldsymbol{\omega}\in\boldsymbol{\Omega}\subseteq\mathbb{R}^q$, $\boldsymbol{\Omega}$ is an open subset and $l(\boldsymbol{\theta}|\boldsymbol{\omega})$ the log-likelihood function corresponding to the perturbed data or model. Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_\omega$ denote the maximum likelihood estimates under the model defined by $l(\boldsymbol{\theta})$ and $l(\boldsymbol{\theta}|\boldsymbol{\omega})$, respectively, and assume that there is a $\boldsymbol{\omega}_0\in\boldsymbol{\Omega}$ representing no perturbation, such that $l(\boldsymbol{\theta})$=$l(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$ for all $\boldsymbol{\theta}$. The influence of $\boldsymbol{\omega}$ can be assessed by the log-likelihood displacement

$$LD(\boldsymbol{\omega}) = 2[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{\omega})], \tag{3.1}$$

where $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\omega_0}$. Because evaluation of $LD(\boldsymbol{\omega})$ for all $\boldsymbol{\omega}$ is practically unfeasible, Cook (1986) proposed to study the local behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$, which can be performed by evaluating the normal curvature $C_{\boldsymbol{l}}$ of $LD(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\boldsymbol{l}$.

Cook (1986) showed that the normal curvature in the direction $\boldsymbol{l}$ takes the form

$$C_{\boldsymbol{l}} = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|, \tag{3.2}$$

where $\|\boldsymbol{l}\| = 1$, $\ddot{\boldsymbol{L}} = -\dfrac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$ is a $(p+2) \times (p+2)$ observed information matrix, and

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta}/\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \tag{3.3}$$

are both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega_0}$.

Let $\boldsymbol{l}_{\max}$ be the direction of the maximum normal curvature (C_{max}), which is the perturbation that produces the greatest local change in $\widehat{\boldsymbol{\theta}}$. The most influential elements of the data may be identified by looking at the components of the vector $\boldsymbol{l}_{\max}$, which are relatively large. Furthermore, $\boldsymbol{l}_{\max}$ is just the eigenvector corresponding to the largest eigenvalue, (C_{max}), of $\boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta}$. Other important direction is $\boldsymbol{l} = \boldsymbol{e}_j$, denoting that the element of the j-th position is one. In that case, the normal curvature, called the total local influence of individual j, is given by $C_j = 2\boldsymbol{\Delta}_j^\top \ddot{\boldsymbol{L}}^{-1} \boldsymbol{\Delta}_j$, where $\boldsymbol{\Delta}_j$ is the j-th column of $\boldsymbol{\Delta}$, $j = 1, ..., n$. We use $\boldsymbol{l}_{\max}$ and C_{max} as diagnostics for local influence. When a subset $\boldsymbol{\theta}_1$ from the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is of interest, influence diagnostics can be based on (Cook, 1986) $\boldsymbol{\Delta}^\top (\ddot{\boldsymbol{L}}^{-1} - \boldsymbol{B}_{22}) \boldsymbol{\Delta}$, with $\boldsymbol{B}_{22} = diag(\boldsymbol{0}, \ddot{\boldsymbol{L}}_{22}^{-1})$ and $\ddot{\boldsymbol{L}}_{22}$ is determined by the partition of $\ddot{\boldsymbol{L}}$ accordingly with the partition of $\boldsymbol{\theta}$.

Zhu and Lee's approach:

Consider a perturbation vector $\boldsymbol{w}$ varying in an open region $\boldsymbol{\Omega} \in \mathbb{R}^q$. Let $l_c(\boldsymbol{\theta}, \boldsymbol{w}|\mathbf{y}_c)$, $\boldsymbol{\theta} \in \mathbb{R}^p$ be the complete-data log-likelihood of the perturbed model. We assume that there is a $\boldsymbol{w}_0$ such that $l_c(\boldsymbol{\theta}, \boldsymbol{w}_0|\mathbf{Y}_c) = l_c(\boldsymbol{\theta}|\mathbf{Y}_c)$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\boldsymbol{w})$ denotes the maximum $\boldsymbol{\theta}$ of the function $Q(\boldsymbol{\theta}, \boldsymbol{w}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[l_c(\boldsymbol{\theta}, \boldsymbol{w}|\mathbf{Y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$. The graph of influence is defined as $\boldsymbol{\alpha}(\boldsymbol{w}) = (\boldsymbol{w}^\top, f_Q(\boldsymbol{w}))^\top$, where $f_Q(\boldsymbol{w})$ is the Q-displacement function defined as

$$f_Q(\boldsymbol{w}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{w})|\widehat{\boldsymbol{\theta}}\right)\right].$$

Following the approach developed in Cook (1986) and Zhu and Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$, of $\boldsymbol{\alpha}(\boldsymbol{w})$ at $\boldsymbol{w}_0$ in the direction of some unit vector $\mathbf{d}$ can be used to summarize the local behavior of the Q-displacement function. It can be shown that (see, Zhu and Lee, 2001)

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{w}_0} \mathbf{d} = 2\mathbf{d}^\top \Delta_{\boldsymbol{w}_0}^\top \left\{-\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})\right\}^{-1} \Delta_{\boldsymbol{w}_0} \mathbf{d}^\top$$

where $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}}) = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\Delta_{\boldsymbol{w}} = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{w}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{w}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\mathbf{w})}$.

As in Cook (1986), the expression $-\ddot{Q}_{\boldsymbol{w}_0}$ is the fundamental equation for detecting influential observations. A clear picture of $-\ddot{Q}_{\boldsymbol{w}_0}$ (a symmetric matrix), is given by its spectral decomposition

$$-2\ddot{Q}_{\boldsymbol{w}_o} = \sum_{k=1}^{n} \lambda_k \mathbf{e}_k \mathbf{e}_k',$$

where $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_n, \mathbf{e}_n)$ are the eigenvalue-eigenvector pairs of the matrix $-2\ddot{Q}_{\mathbf{w}_o}$ with $\lambda_1 \geq \ldots \geq \lambda_q, \lambda_{q+1} = \ldots = \lambda_n = 0$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are elements of the associated orthonormal basis. Lesaffre and Verbeke (1998), Poon and Poon (1999) and Zhu and Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for more revealing information but it 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 $\widetilde{\lambda}_k = \lambda_k/(\lambda_1 + \ldots + \lambda_q)$ and $\mathbf{e}_k^2 = (e_{k1}^2, \ldots, e_{kn}^2)$, and

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

Hence, the assessment of influential cases is based on $\{M(0)_l, l = 1, \ldots, n\}$ and one can obtain $M(0)_l$ via $B_{f_Q,\mathbf{u}_l} = -2\mathbf{u}_l^\top \ddot{Q}_{\mathbf{w}_0} \mathbf{u}_l / tr[-2\ddot{Q}_{\mathbf{w}_0}]$, 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 $B_{f_Q,\mathbf{u}_l}$, such as invariance under reparameterization of $\boldsymbol{\theta}$. Additionally, Lee and Xu (2004) propose to use $1/n + c^* SM(0)$ as a bench-mark 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 $\{M(0)_l, l = 1, \ldots, n\}$.

3.1 Cook's approach

3.1.1 Case weight Perturbation

Consider the vector $\boldsymbol{w} = (w_1, ..., w_n)^\top$ of case-weights, so that the perturbed log-likelihood function is given by

$$L(\boldsymbol{\theta}/\boldsymbol{w}) = \sum_{j=1}^{n} w_j l_j(\boldsymbol{\theta}),$$

where $l_j(\boldsymbol{\theta})$ is as in (2.4). Here, we are interested in pinpointing influential data points among all observations. This perturbation scheme is the most common perturbation scheme in the literature. In this case, the vector of no perturbations is given by $\boldsymbol{w}_0 = \mathbf{1}_n$, where $\mathbf{1}_n$ is an $n \times 1$ vector with all its elements equal to one. Under this perturbation scheme the matrix $\boldsymbol{\Delta}$ defined in (3.3) is a

$(p+2) \times n$ matrix and given by $\mathbf{\Delta} = (\mathbf{\Delta}^\top(\boldsymbol{\beta}), \mathbf{\Delta}^\top(\sigma^2), \mathbf{\Delta}^\top(\delta))^\top$, where

$$
\begin{aligned}
\mathbf{\Delta}(\boldsymbol{\beta}) &= \frac{1}{\widetilde{\sigma}^2 + \widetilde{\delta}^2}\mathbf{X}^\top D(\mathbf{e}) - \frac{\widetilde{\delta}}{\widetilde{\sigma}(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^{1/2}}\mathbf{X}^\top D(\widetilde{\mathbf{a}}), \\
\mathbf{\Delta}(\sigma^2) &= -\frac{1}{2(\widetilde{\sigma}^2 + \widetilde{\delta}^2)}\mathbf{1}_n^\top + \frac{1}{2(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^2}\mathbf{e}^\top D(\mathbf{e}) - \frac{\widetilde{\delta}(2\widetilde{\sigma}^2 + \widetilde{\delta}^2)}{2\widetilde{\sigma}^3(\widetilde{\sigma}^2 + \widetilde{\delta})^{3/2}}\mathbf{e}^\top D(\widetilde{\mathbf{a}}) \\
\mathbf{\Delta}(\delta) &= -\frac{\widetilde{\delta}}{\widetilde{\sigma}^2 + \widetilde{\delta}^2}\mathbf{1}_n^\top + \frac{\widetilde{\delta}}{(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^2}\mathbf{e}^\top D(\mathbf{e}) + \frac{\widetilde{\sigma}}{(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^{3/2}}\mathbf{e}^\top D(\widetilde{\mathbf{a}}),
\end{aligned}
$$

with $\mathbf{a}$ as in (2.6) and

$$
\mathbf{e} = (e_1, \cdots, e_n)^\top = \mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}} \tag{3.4}
$$

is the vector of residual and $e_j = Y_j - \boldsymbol{x}_j^\top \widetilde{\boldsymbol{\beta}}$ denotes the component of the vector $\mathbf{e}$.

Now, letting $\boldsymbol{\beta}_* = (\beta_2, \cdots, \beta_p)^\top = \mathbf{0}$ in (2.1), we have the skew normal model $SN(\mu, \sigma^2, \delta)$, with $\mu = \beta_1$, defined in Sahu et al. (2003). In this case the matrix $\mathbf{\Delta} = (\mathbf{\Delta}^\top(\mu), \mathbf{\Delta}^\top(\sigma^2), \mathbf{\Delta}^\top(\delta))^\top$, is given by

$$
\begin{aligned}
\mathbf{\Delta}(\mu) &= \frac{\mathbf{e}^\top}{\widetilde{\sigma}^2 + \widetilde{\delta}^2} - \frac{\widetilde{\delta}}{\widetilde{\sigma}(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^{1/2}}\widetilde{\mathbf{a}}^\top, \\
\mathbf{\Delta}(\sigma^2) &= -\frac{\mathbf{1}_n^\top}{2(\widetilde{\sigma}^2 + \widetilde{\delta}^2)} + \frac{1}{2(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^2}\mathbf{e}^\top D(\mathbf{e}) - \frac{\widetilde{\delta}(2\widetilde{\sigma}^2 + \widetilde{\delta}^2)}{2\widetilde{\sigma}^3(\widetilde{\sigma}^2 + \widetilde{\delta})^{3/2}}\mathbf{e}^\top D(\widetilde{\mathbf{a}}), \\
\mathbf{\Delta}(\delta) &= -\frac{\widetilde{\delta}}{\widetilde{\sigma}^2 + \widetilde{\delta}^2}\mathbf{1}_n^\top + \frac{\widetilde{\delta}}{(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^2}\mathbf{e}^\top D(\mathbf{e}) + \frac{\widetilde{\sigma}}{(\widetilde{\sigma}^2 + \widetilde{\delta}^2)^{3/2}}\mathbf{e}^\top D(\widetilde{\mathbf{a}}),
\end{aligned}
$$

where

$$
\mathbf{e} = (e_1, \cdots, e_n)^\top \quad \text{and} \quad \widetilde{\mathbf{a}} = (\widetilde{a}_1, \cdots, \widetilde{a}_n)^\top, \tag{3.5}
$$

with $e_j = y_j - \widetilde{\mu}$ and $\widetilde{a_j} = W_\Phi \left(\frac{\widetilde{\delta}}{\widetilde{\sigma}} \frac{y_j - \widetilde{\mu}}{(\widetilde{\delta}^2 + \widetilde{\sigma}^2)^{1/2}} \right)$.

3.1.2 Response variable perturbation

One way of perturbing the response variable, when our interest is to detect the sensitivity of the model when this kind of perturbation happens, the perturbation is introduced by considering $Y_{wj} = Y_j + S_j w_j$. The scale factor S_j can be taken as $S = S_y$, where S_y denotes for example, the sample standard deviation of $Y_1, ..., Y_n$. The perturbed log-likelihood function is given by $L(\boldsymbol{\theta}/\boldsymbol{w}) = \sum_{j=1}^{n} l_j(\boldsymbol{\theta}/w_j)$, where $l_j(\boldsymbol{\theta}/w_j)$ is as given in (2.4), switching Y_{wj} with Y_j and $\boldsymbol{w} = (w_1, ..., w_n)^\top$. Under this perturbation scheme the vector $\boldsymbol{w}_0$ is given by $\boldsymbol{w}_0 = \mathbf{0}$ and the $(p+2) \times n$ matrix $\mathbf{\Delta} = (\mathbf{\Delta}^\top(\boldsymbol{\beta}), \mathbf{\Delta}^\top(\sigma^2), \mathbf{\Delta}^\top(\delta))^\top$, can be expressed as

$$
\mathbf{\Delta} = \frac{S_y}{\widetilde{\sigma}^2 + \widetilde{\delta}^2} \begin{pmatrix} \boldsymbol{X}^\top \\ \frac{1}{\widetilde{\sigma}^2 + \widetilde{\delta}^2}\boldsymbol{e}^\top \\ \frac{2\widetilde{\delta}}{\widetilde{\sigma}^2 + \widetilde{\delta}^2}\boldsymbol{e}^\top \end{pmatrix} - S_y \begin{pmatrix} a_0^2 \boldsymbol{X}^\top D(\widetilde{\boldsymbol{b}}) \\ a_{0\sigma^2}\left[\widetilde{\mathbf{a}}^\top + a_0 \mathbf{e}^\top D(\widetilde{\mathbf{b}})\right] \\ -a_{0\delta}\left[\widetilde{\mathbf{a}}^\top + a_0 \mathbf{e}^\top D(\widetilde{\mathbf{b}})\right] \end{pmatrix}, \tag{3.6}
$$

where

$$a_0 = \frac{\widetilde{\delta}}{\widetilde{\sigma}(\widetilde{\sigma}^2+\widetilde{\delta}^2)^{1/2}}, \quad a_{0\sigma^2} = \frac{\widetilde{\delta}(2\widetilde{\sigma}^2+\widetilde{\delta}^2)}{2\widetilde{\sigma}^3(\widetilde{\sigma}^2+\widetilde{\delta}^2)^{3/2}} \text{ and } a_{0\delta} = \frac{\widetilde{\sigma}}{(\widetilde{\sigma}^2+\widetilde{\delta}^2)^{3/2}}.$$

In the spacial case, when $Y_j \sim SN(\mu, \sigma^2, \delta)$, with $\mu = \beta_1$ and $\boldsymbol{\beta}_* = \mathbf{0}$, the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}^\top(\mu), \boldsymbol{\Delta}^\top(\sigma^2), \boldsymbol{\Delta}^\top(\delta))^\top$, is given by

$$\boldsymbol{\Delta} = \frac{S_y}{\widetilde{\sigma}^2+\widetilde{\delta}^2}\begin{pmatrix} \mathbf{1}_n^\top \\ \frac{1}{\widetilde{\sigma}^2+\widetilde{\delta}^2}\boldsymbol{e}^\top \\ \frac{2\widetilde{\delta}}{\widetilde{\sigma}^2+\widetilde{\delta}^2}\boldsymbol{e}^\top \end{pmatrix} - S_y \begin{pmatrix} a_0^2\widetilde{\boldsymbol{b}}^\top \\ a_{0\sigma^2}\left[\widetilde{\mathbf{a}}^\top + a_0\mathbf{e}^\top D(\widetilde{\mathbf{b}})\right] \\ -a_{0\delta}\left[\widetilde{\mathbf{a}}^\top + a_0\mathbf{e}^\top D(\widetilde{\mathbf{b}})\right] \end{pmatrix},$$

where $\mathbf{e}$ and $\widetilde{\mathbf{a}}$ are as in (3.5), with $\boldsymbol{\beta}_* = \mathbf{0}$ and $\widetilde{\mathbf{b}} = (\widetilde{b}_1, \cdots, \widetilde{b}_n)^\top$, with

$$\widetilde{b}_j = -W_\Phi\left(\frac{\widetilde{\delta}}{\widetilde{\sigma}}\frac{y_j-\widetilde{\mu}}{(\widetilde{\delta}^2+\widetilde{\sigma}^2)^{1/2}}\right)\left[\frac{\widetilde{\delta}}{\widetilde{\sigma}}\frac{y_j-\widetilde{\mu}}{(\widetilde{\delta}^2+\widetilde{\sigma}^2)^{1/2}} + W_\Phi\left(\frac{\widetilde{\delta}}{\widetilde{\sigma}}\frac{y_j-\widetilde{\mu}}{(\widetilde{\delta}^2+\widetilde{\sigma}^2)^{1/2}}\right)\right], \quad j = 1, ..., n.$$

3.1.3 Perturbation of the explanatory variable

If we are interested in investigating the sensitivity of minor perturbation in the explanatory variable, we can define the following perturbation scheme for the explanatory variable in the same way that was defined in Cook (1986) or Galea et al. (2003), namely

$$\mathbf{x}_{tw} = \mathbf{x}_t + S_t\boldsymbol{w}, \quad t = 1, ..., p.$$

Here, $\mathbf{x}_t$ is the t-th column of the matrix $\mathbf{X}$, $\boldsymbol{w}_j$ and denotes the $n \times 1$ perturbation vector and S_t is a scale factor that can be considered as the sample standard deviation of the elements of $\mathbf{x}_t = (x_{1t}, ..., x_{nt})^\top$. The log-likelihood function for the perturbed model $L(\boldsymbol{\theta}/\boldsymbol{w})$ is as defined in (2.5), switching $\mathbf{x}_{tw}$ with $\mathbf{x}_t$. The vector $\boldsymbol{w}_0$ representing no perturbation is given by $\boldsymbol{w}_0 = \mathbf{0}$. The $(p+2) \times n$ matrix $\boldsymbol{\Delta}$ defined in (3.3) is given by $\boldsymbol{\Delta} = (\boldsymbol{\Delta}^\top(\boldsymbol{\beta}), \boldsymbol{\Delta}(\sigma^2), \boldsymbol{\Delta}(\delta))^\top$ with

$$\boldsymbol{\Delta} = \frac{S_t}{\widetilde{\sigma}^2+\widetilde{\delta}^2}\begin{pmatrix} \boldsymbol{e}_{pt}\mathbf{e}^\top - \widetilde{\beta}_t\boldsymbol{X}^\top \\ -\frac{\widetilde{\beta}_t}{\widetilde{\sigma}^2+\widetilde{\delta}^2}\boldsymbol{e}^\top \\ -\frac{2\widetilde{\delta}\widetilde{\beta}_t}{\widetilde{\sigma}^2+\widetilde{\delta}^2}\boldsymbol{e}^\top \end{pmatrix} + S_t\begin{pmatrix} a_0^2\widetilde{\beta}_t\boldsymbol{X}^\top D(\widetilde{\boldsymbol{b}}) - a_0\boldsymbol{e}_{pt}\widetilde{\mathbf{a}}^\top \\ a_{0\sigma^2}\widetilde{\beta}_t\left[\widetilde{\mathbf{a}}^\top + a_0\mathbf{e}^\top D(\widetilde{\mathbf{b}})\right] \\ -a_{0\delta}\widetilde{\beta}_t\left[\widetilde{\mathbf{a}}^\top + a_0\mathbf{e}^\top D(\widetilde{\mathbf{b}})\right] \end{pmatrix}, \tag{3.7}$$

where $\boldsymbol{e}_p(t)$ is a $p \times 1$ vector with 1 in the t-th position and zeros elsewhere.

3.1.4 Generalized leverage

Other 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_i on its own predicted value (see, Emerson et al., 1989; Wei et al., 1998). This influence may be represented by the derivative $\partial\widehat{y}_i/\partial y_i$. Under normal linear case, $\partial\widehat{y}_i/\partial y_i = h_{ii}$ that is the i-th principal diagonal element of the projection matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$.

In the linear regression models under the skew-normal distribution, the expectation of the $\boldsymbol{Y}$ is given by $E(\boldsymbol{Y}) = \mu(\boldsymbol{\theta}) = \boldsymbol{X\beta} + c\delta\mathbf{1}_n$, then $\widehat{\boldsymbol{Y}} = \mu(\widehat{\boldsymbol{\theta}})$ is the predicted response vector. The generalized leverage proposed by Wei et al. (1998) is defined as

$$GL(\widehat{\boldsymbol{\theta}}) = D_{\boldsymbol{\theta}}(-\ddot{L}(\boldsymbol{\theta}))^{-1}\ddot{L}_{\boldsymbol{\theta}\mathbf{Y}}, \tag{3.8}$$

where $-\ddot{L}$ is observed information matrix given in (2.12), $D_{\boldsymbol{\theta}} = \dfrac{\partial\mu}{\partial\boldsymbol{\theta}^\top}$ and $\ddot{L}_{\boldsymbol{\theta}\mathbf{Y}} = \dfrac{\partial^2 L_{\mathbf{Y}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\mathbf{Y}^\top}$. In the skew-normal linear case, we have that

$$D_{\boldsymbol{\theta}} = [\mathbf{X},\ \mathbf{0}_n,\ c\mathbf{1}_n,\ ] \tag{3.9}$$

and

$$\ddot{L}_{\boldsymbol{\theta}\mathbf{Y}} = \frac{1}{\sigma^2+\delta^2}\begin{bmatrix} \mathbf{X}^\top\left(-\frac{\delta^2}{\sigma^2}D(\mathbf{b}) + \mathbb{I}_n\right) \\ -\frac{\delta(2\sigma^2+\delta^2)}{2\sigma^3(\sigma^2+\delta^2)^{1/2}}\left(\mathbf{a}^\top + \frac{\delta^2}{\sigma(\sigma^2+\delta^2)^{1/2}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top D(\mathbf{b})\right) + \frac{1}{\sigma^2+\delta^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top \\ \frac{\sigma}{(\sigma^2+\delta^2)^{1/2}}\left(\mathbf{a}^\top + \frac{\delta}{\sigma^2(\sigma^2+\delta^2)^{1/2}}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top D(\mathbf{b}) + \frac{2\delta}{(\sigma^2+\delta^2)^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top\right) \end{bmatrix},$$

where $\mathbf{a}$ and $\mathbf{b}$ are as in (2.6) and (2.12), respectively.

3.2 Zhu and Lee's approach

The model (2.1) can be obtained by stochastic representation

$$\begin{aligned} Y_j|u_j &\sim N(\mathbf{x}_j^\top \boldsymbol{\beta} + \delta u_j, \sigma^2) \\ u_j &\sim HN(0,1). \end{aligned} \tag{3.10}$$

It follows that

$$u_j|Y_j \sim HN\left(\frac{\delta}{\sigma^2+\delta^2}(y_j - \mathbf{x}_j^\top \boldsymbol{\beta}), \frac{\sigma^2}{\sigma^2+\delta^2}\right). \tag{3.11}$$

Now, let $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{u} = (u_1, \dots, u_n)^\top$ and treating $\mathbf{u}$ as missing data, it follows that the complete log-likelihood function associated with $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ is given by

$$l_c(\boldsymbol{\theta}|\mathbf{y}_c) \;\propto\; -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) - \frac{\delta}{\sigma^2}\mathbf{u}^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{\delta^2}{2\sigma^2}\sum_{j=1}^n u_j^2.$$

Letting $\widehat{u}_j = \mathrm{E}[U_j|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_j]$ and $\widehat{u^2}_j = \mathrm{E}[U_j^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_j]$, we obtain, using the moments of the truncated normal distribution, it follows that

$$\begin{aligned} \widehat{u}_j &= \frac{\widehat{\delta}}{\widehat{\sigma}^2+\widehat{\delta}^2}e_j + \frac{\widehat{\sigma}}{(\widehat{\sigma}^2+\widehat{\delta}^2)^{1/2}}W_{\Phi_1}\left(\frac{\widehat{\delta}e_j}{\widehat{\sigma}\sqrt{\widehat{\sigma}^2+\widehat{\delta}^2}}\right) \\ \widehat{u^2}_j &= \frac{\widehat{\sigma}^2}{\widehat{\sigma}^2+\widehat{\delta}^2} + \frac{\widehat{\delta}^2}{(\widehat{\sigma}^2+\widehat{\delta}^2)^2}e_j^2 + \frac{3\widehat{\delta}\widehat{\sigma}}{(\widehat{\sigma}^2+\widehat{\delta}^2)^{3/2}}W_{\Phi_1}\left(\frac{\widehat{\delta}e_j}{\widehat{\sigma}\sqrt{\widehat{\sigma}^2+\widehat{\delta}^2}}\right)e_i \end{aligned} \tag{3.12}$$

where $e_j = y_j - \mathbf{x}_j^\top\widehat{\boldsymbol{\beta}}$ and $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$. It follows that the conditional expectation of the complete log-likelihood function has the form

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) &= \mathrm{E}[l_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}] \\ &\propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) - \frac{\delta}{\sigma^2}\widehat{\mathbf{u}}^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{\delta^2}{2\sigma^2}\sum_{j=1}^n \widehat{u_j^2}. \end{aligned} \tag{3.13}$$

Thus, we have the following EM-algorithm:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{u}_j$, $\widehat{u^2}_j$ for $j = 1, .., n$ using (3.12).

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

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} - \widehat{\delta}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{u}}, \qquad (3.14) \\ \widehat{\sigma^2} &= \frac{1}{n}\big[\mathbf{e}^\top\mathbf{e} - 2\widehat{\delta}\widehat{\mathbf{u}}^\top\mathbf{e} + \widehat{\delta}^2\sum_{j=1}^n \widehat{u^2}_j\big], \\ \widehat{\delta} &= \frac{\widehat{\mathbf{u}}^\top\mathbf{e}}{\sum_{j=1}^n \widehat{u^2}_j}, \end{aligned}$$

where $\boldsymbol{e} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}} = (\widehat{u}_1, .., \widehat{u}_n)^\top$. When $\delta = 0$, the M-step equations reduces to the equations assuming the symmetric normal distribution. Note that if $\widehat{\boldsymbol{\beta}}_N$ and $\widehat{\sigma^2}_N$ denote the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2, respectively under normal distribution, then

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_N - \widehat{\delta}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\widehat{\mathbf{u}} \;\text{ and }\; \widehat{\sigma^2} = \widehat{\sigma^2}_N - \frac{\widehat{\delta}}{n}(2\mathbf{e}_N^\top(\mathbf{I}-\mathbf{H})\mathbf{u} + \widehat{\delta}\mathbf{u}^\top\mathbf{H}\mathbf{u} + n\widehat{\delta}\,\overline{\widehat{u^2}}), \tag{3.15}$$

where $\mathbf{e}_N = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_N$, $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ and $\overline{\widehat{u^2}} = \frac{1}{n}\sum_{j=1}^{n}\widehat{u^2}_j$.

Now, to obtain the diagnostic measures for local influence of a particular perturbation scheme, it is necessary to compute $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})$. So, after some algebraic manipulations it follows that that the matrix $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})$ can be written as

$$\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}}) = -\begin{pmatrix} \ddot{Q}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2}) & \ddot{\boldsymbol{Q}}_{12}(\widehat{\boldsymbol{\theta}}) \\ \ddot{\boldsymbol{Q}}_{12}^\top(\widehat{\boldsymbol{\theta}}) & \frac{n}{\widehat{\sigma^2}}\overline{\widehat{u^2}} \end{pmatrix}, \tag{3.16}$$

$$\ddot{Q}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2}) = \begin{pmatrix} \frac{1}{\widehat{\sigma^2}}\mathbf{X}^\top\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{n}{\widehat{\sigma^4}} \end{pmatrix} \quad \text{and} \quad \ddot{Q}_{12}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{1}{\widehat{\sigma^2}}\mathbf{X}^\top\widehat{\mathbf{u}} \\ \mathbf{0} \end{pmatrix}.$$

3.2.1 Case weight Perturbation

Let $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ a $n \times 1$ dimensional vector with $\boldsymbol{w}_0 = (1, \ldots, 1)^\top$, then the expected value of the perturbed complete-data log-likelihood function (perturbed Q-function), can be written as

$$Q(\boldsymbol{\theta}, \boldsymbol{w}|\widehat{\boldsymbol{\theta}}) \;=\; \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{w}|\mathbf{y}_c)] = \sum_{j=1}^{n}\omega_i \mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)] = \sum_{j=1}^{n} w_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

In this case, the matrix $\boldsymbol{\Delta}_{\boldsymbol{w}_0}$ is given by

$$\boldsymbol{\Delta}_{\boldsymbol{w}_0} = \frac{1}{\widehat{\sigma^2}}\begin{pmatrix} \mathbf{X}^\top\mathbf{D}(\mathbf{e}) \\ \frac{1}{2\widehat{\sigma^2}}\mathbf{e}^\top\mathbf{D}(\mathbf{e}) - \frac{1}{2}\mathbf{1}_n^\top \\ \mathbf{0} \end{pmatrix} + \frac{1}{\widehat{\sigma^2}}\begin{pmatrix} -\widehat{\delta}\mathbf{X}^\top\mathbf{D}(\widehat{\mathbf{u}}) \\ -\frac{\widehat{\delta}}{\widehat{\sigma^2}}\mathbf{e}^\top\mathbf{D}(\widehat{\mathbf{u}}) + \frac{\widehat{\delta^2}}{2\widehat{\sigma^2}}\widehat{\mathbf{u}^2}^\top \\ \mathbf{e}^\top\mathbf{D}(\widehat{\mathbf{u}}) - \widehat{\delta}\widehat{\mathbf{u}^2}^\top \end{pmatrix}, \tag{3.17}$$

$\widehat{\mathbf{u}^2} = (\widehat{u_1^2}, ..., \widehat{u_n^2})^\top$. Note that when $\delta = 0$, the matrix $\boldsymbol{\Delta}_{\boldsymbol{w}_0}$ reduces to normal distribution case (for $\boldsymbol{\beta}$ and σ^2, see for instance, Cook, 1986).

3.2.2 Perturbation of the response variable

A perturbation of the response variables $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ is introduced by replacing Y_i by $Y_{iw} = Y_i + w_i S_y$, where S_y is the standard deviation of $\mathbf{Y}$. The perturbed Q-function is given by $Q(\boldsymbol{\theta}, \boldsymbol{w}|\widehat{\boldsymbol{\theta}})$, switching Y_{wj} with Y_j and $\boldsymbol{w} = (w_1, ..., w_n)^\top$. Under this perturbation scheme the vector $\boldsymbol{w}_0$ is given by $\boldsymbol{w}_0 = \mathbf{0}$ and the matrix $\boldsymbol{\Delta}_{\boldsymbol{w}_0}$ is given by

$$\boldsymbol{\Delta}_{\boldsymbol{w}_0} = \frac{S_y}{\widehat{\sigma^2}}\begin{pmatrix} \mathbf{X}^\top \\ \frac{1}{\widehat{\sigma^2}}\mathbf{e}^\top \\ \mathbf{0} \end{pmatrix} + \frac{S_y}{\widehat{\sigma^2}}\begin{pmatrix} \mathbf{0} \\ -\frac{\widehat{\delta}}{\widehat{\sigma^2}}\widehat{\mathbf{u}}^\top \\ \widehat{\mathbf{u}}^\top \end{pmatrix}. \tag{3.18}$$

3.2.3 Perturbation of the explanatory variable

In this case we are interested in perturbing a specific explanatory variable. Under this condition we have the following perturbed explanatory variable $\mathbf{x}_{tw} = \mathbf{x}_t + S_t\boldsymbol{w}$, $t = 1, \ldots, p$, where S_t is the standard deviation of the explanatory variable $\mathbf{x}_t$. In this case, $\boldsymbol{w}_0 = \mathbf{0}$ and the perturbed Q-function is given by $Q(\boldsymbol{\theta}, \boldsymbol{w}|\widehat{\boldsymbol{\theta}})$, switching $\mathbf{x}_{tw}$ with $\mathbf{x}_t$. The $(p+2) \times n$ matrix $\boldsymbol{\Delta}_{\boldsymbol{w}_0}$ is given by

$$\boldsymbol{\Delta}_{\boldsymbol{w}_0} = \frac{S_t}{\widehat{\sigma^2}} \begin{pmatrix} \boldsymbol{e}_p(t)\mathbf{e}^\top - \widehat{\beta}_t\mathbf{X}^\top \\ -\dfrac{\widehat{\beta}_t}{\widehat{\sigma^2}}\mathbf{e}^\top \\ \mathbf{0} \end{pmatrix} + \frac{S_t}{\widehat{\sigma^2}} \begin{pmatrix} -\widehat{\delta}\boldsymbol{e}_p(t)\widehat{\mathbf{u}}^\top \\ \dfrac{\widehat{\beta}_t}{\widehat{\sigma^2}}\widehat{\delta}\widehat{\mathbf{u}}^\top \\ -\widehat{\beta}_t\widehat{\mathbf{u}}^\top \end{pmatrix}. \tag{3.19}$$

In the equation above, $\boldsymbol{e}_p(t)$ is, as before, a $p \times 1$ vector with 1 in the t-th position and zeros elsewhere.

3.2.4 Generalized Leverage

According with Wei, Fu e Fung (1998) as well as in the work of Zhu and Lee (2001), Salgado (2006) defines the *generalized leverage matrix for models with incomplete data* by

$$GL(\widehat{\boldsymbol{\theta}}) = D_{\boldsymbol{\theta}}[-\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})]^{-1}\ddot{Q}_{\boldsymbol{\theta},\mathbf{y}}(\widehat{\boldsymbol{\theta}}), \tag{3.20}$$

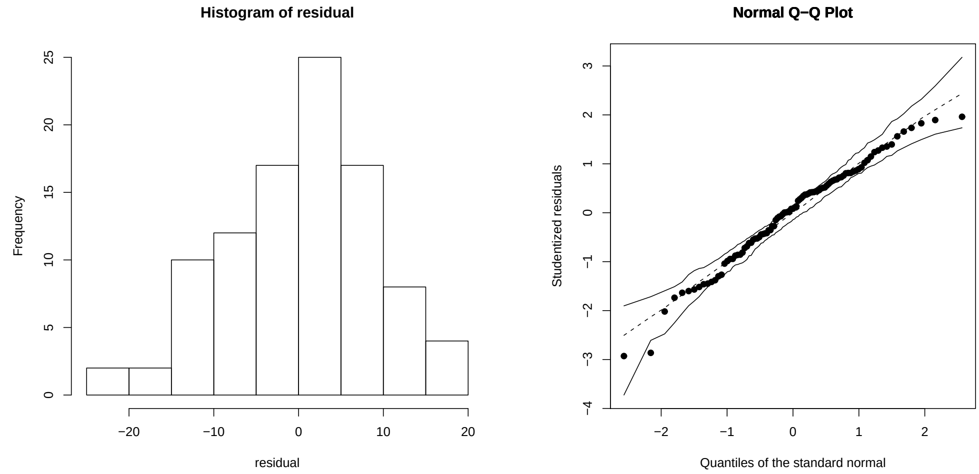
where $D_{\boldsymbol{\theta}} = \dfrac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\theta}^\top}$, $\ddot{Q}_{\boldsymbol{\theta},\mathbf{y}}(\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\mathbf{y}^\top}$, $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \delta)^\top$.

The matrix $D_{\boldsymbol{\theta}}$ is as in (3.9). The matrix $\ddot{Q}_{\boldsymbol{\theta},\mathbf{y}}(\widehat{\boldsymbol{\theta}})$ is well as in (3.18) without the term S_y and $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$ is given in the Section 3.16.

Salgado (2006) propose to use c_0p_0/n, $p_0 = \sum_{j=1}^{n}\mathbf{GL}_{jj}(\widehat{\boldsymbol{\theta}}) = \mathrm{tr}(\mathbf{GL}(\widehat{\boldsymbol{\theta}}))$, as a bench-mark to regard the l-th case as leverage point, where c_0 is a selected constant (depending on the real application).

4 APPLICATION

A data set about life's quality of women with breast cancer realized by Center of Integral Attention to the Woman's Health (State University from Campinas, Brazil) was adjusted to verify the application of the skew normal regression model and the diagnostics analysis. The index of life's quality was obtained trough of a 36 items questionnaire (SF-36), largely utilized in sources of the health's area. The same index is composed by two components, a physical and a mental. Conde et al. (2005) studied this data set, evaluating the associated factors to life's quality of women with breast cancer. The histogram of the index *pcs* is plotted in the Figure 2(a), showing a moderate asymmetry.



(a) **(b)**

Figure 1: Data set about Life's Quality. Adjust under normal linear model. (a) Histogram of the residuals (b) Simulated envelope.

The regression model was built using as dependent variable the the physical component summary of the index of life's quality (*pcs*) and, as explanatory variables, the indicator variable *dizziness* and the body mass index (*bmi*) of the individual. So, it has the skew normal linear regression model

$$\begin{aligned} pcs_i &= \beta_0 + \beta_1 * dizziness_i + \beta_2 * bmi_i + e_i, \quad i = 1, \ldots, 97, \\ e_i &\sim SN(0, \sigma^2, \lambda). \end{aligned} \tag{4.1}$$

After adjusting the model, the estimated parameters, jointly with with their corresponding estimated standard errors (calculated using the observed information matrix), are described in the Table 1. The residual in the skew normal linear regression model is calculated by using $y_i - \mathbf{x}_i^\top \beta - \sqrt{\frac{2}{\pi}}\delta$.

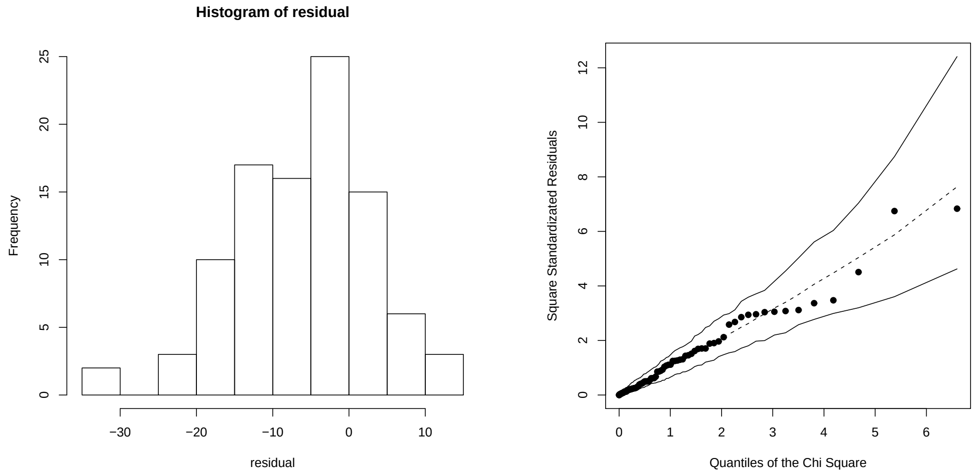
Table 1: MLE of the data set of life's quality (estimated standard error asymptotically in parentheses) for normal linear and skew normal linear models.

Modelo	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\sigma^2}$	$\widehat{\delta}$	$\ell(\boldsymbol{\theta})$
SN	68.57 (4.64)	-7.62(1.91)	-0.35(0.17)	25.47(9.35)	-4.65(1.09)	-346.20
Normal	61.29(4.75)	-7.92(1.91)	-0.43(0.16)	79.13(11.73)	-	-348.14

The simulated envelope graphic built to validate the skew normal regression model, using the stochastic representation given in (1.3), does not indicates points outside the confidence bounds (Figure 2b). On the other hand, the simulated envelope for the normal linear regression model (Figure 1b) indicate various points outside.

According to Table 1, the estimates of $\boldsymbol{\beta}$ in the two models are similar. On the other hand, the estimates of the scale σ^2 parameters are different. The likelihood-ratio test (LR) for $H_0 : \delta = 0 \times H_1 : \delta \neq 0$ is $LR = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_0)) = 3.88$, with a p-value of 0.05. Or rather, a skew normal model is more suitable than normal linear model.

The Cook and the Zhu and Lee's approaches present similar results in the local influence analysis, although the last presents more influence points. Both methodologies shows as influential under the perturbation schemes the points on the upper part of the data (2, 6, 28, 52, 55 and 65). Another way, the methodologies are similar in presenting the points 15, 21 and 23 as leverages (this last only the Zhu and Lee's approach), every on the outside of the data mass.



(a) **(b)**

Figure 2: Data set about Life's Quality. Adjust under skew normal linear model. (a) Histogram of the residuals and (b) Simulated envelope.

5 CONCLUSION

In this paper we have developed diagnostics analysis for skew normal regression models, based on Zhu and Lee's methodology, using the likelihood augmented of the EM algorithm, which has analytical form on the M step. The diagnostics analysis is based in local influence and generalized leverage. Case weight, response variable and explanatory variable perturbations were considered. The Cook's approach is also calculated and comparated with Zhu and Lee's approach. The two methodologies resulted similar conclusions, however, the Zhu and Lee's approach appointed more influence points. The schemes of local influence studied showed influence points on the top of data, while the points outside data mass are appointed as leverage.

REFERENCES

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. Statistica. 46, 199-208.

Azzalini, A. (2005). The Skew-normal distribution and related multivariate families. Scandinavian Journal of Statistics. 32, 159-188.

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

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. American Statistical Association, Journal of Business and Economic Statistics. 23(3), 346-354.

Capitanio, A., Azzalini, A. and Stanghellini, E. (2003). Graphical models for skew-normal variates. Scandinavian Journal of Statistics. 30, 129-144.

Conde, D.M., Pinto-Neto, A.M., Cabello, C., Santos-Sá, D., Costa-Paiva, C., e Martinez, E.Z. (2005). Quality of life in brazilian breast cancer survivors age 45-65 years: associated factors. The Breast Journal. 11, 425-432.

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

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

Cook, R,D. and Weisberg, S. (1982). Residuals and Influence in Regression, Chapman and Hall, London.

Dennis, J.E. and Schnabel, R.B. (1996). Numerical methods for unconstrained optimization and nonlinear equations. Society for Industrial and Applied Mathematics (SIAM). Philadelphia, Pa.

Emerson, J.D., Hoaglin, D.C. and Kempthorne, P.J. (1984). Leverage in least squares additive-plus-multiplicative fits for two-way tables. Journal of the American Statistical Association. 79, 329-335.

Fang, K.-T.; Kotz, S. and Ng, K.-W. (1990). Symmetric Multivariate and Related Distributions, Chapman and Hall, New York.

Fang, K.-T. and Zhang, Y.-T. (1990). Generalized Multivariate Analysis, Springer-Verlag, New York.

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

Galea, M., Paula, G.A. and Uribe, M. (2003). Influence diagnostics in univariate linear elliptical regression model. Statistical Paper. 44, 23-45.

Genton, M.G., ed. (2004). Skew - elliptical distributions and their applications: a journey beyond normality, Chapman and Hall, London.

Genton, M.G., He, L. and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. Statistics and Probability Letters. 4, 319-325.

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., Bolfarine H. and Montenegro, L.C (2008). Inference and Assessment of Local Influence in Skew-Normal Null Intercept Measurement Error Models. Journal of Statistical Computation and Simulation. 78(3), 395-419.

Lachos, V.H., Vilca, L.F. and Galea, M. (2007). Influence Diagnostics for Grubbs's Model. Statistical Papers. 48(3), 419-436.

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

Lee, S. and Xu, L. (2004). Influence analysis of nonlinear mixed-effects models. Computational Statistics and Data Analysis. 45, 321-341.

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

Liseo, B. and Loperfido N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. Statistics and Probability Letters. 61, 395-401.

Liu, S.Z. (2000). On local influence for elliptical linear models. Statistical Papers. 41, 211224.

Lu, B. and Song, X.Y. (2006). Local influence of multivariate probit latent variable models. Journal of Multivariate Analysis. 97, 1783-1798.

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

Rodríguez, C.L.B. (2005). Inferência Bayesiana no Modelo Normal Assimétrico. Master's Thesis. Universidade de Sao Paulo, Brasil.

Sahu, S.K., Dey, D.K. and Branco, M. (2003). A new class of multivariate skew distributions with application to Bayesian regression models. Canadian Journal of Statistics. 31, 129-150.

Salgado, F.A.O. (2006). Diagnóstico de influência em modelos elípticos com efeitos mistos. Doctoral's Thesys. Universidade de Sao Paulo, Brasil.

Vilca-Labra, F.E. and Leiva-Sánchez, V. (2006). A new fatigue life model based on the family of skew-elliptical distributions. Communications in Statistics - Theory and Methods. 35(2), 229-244.

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

APPENDIX

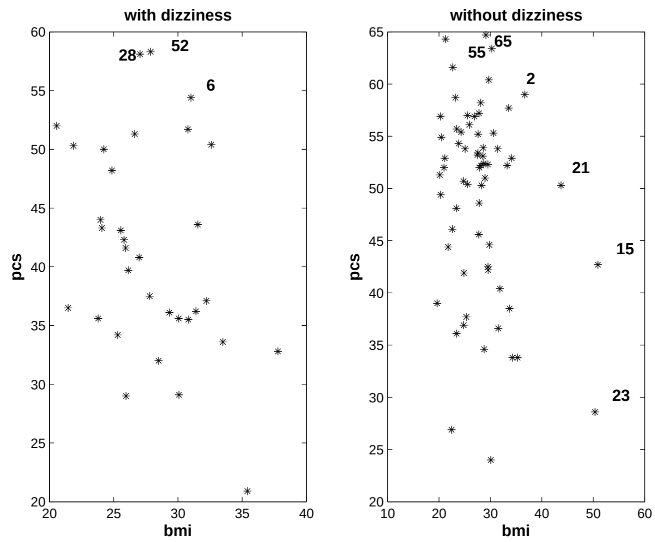


Figure 3: Data set about Life's Quality. Scatter plot of *bmi* by *pcs*, with and without dizziness.

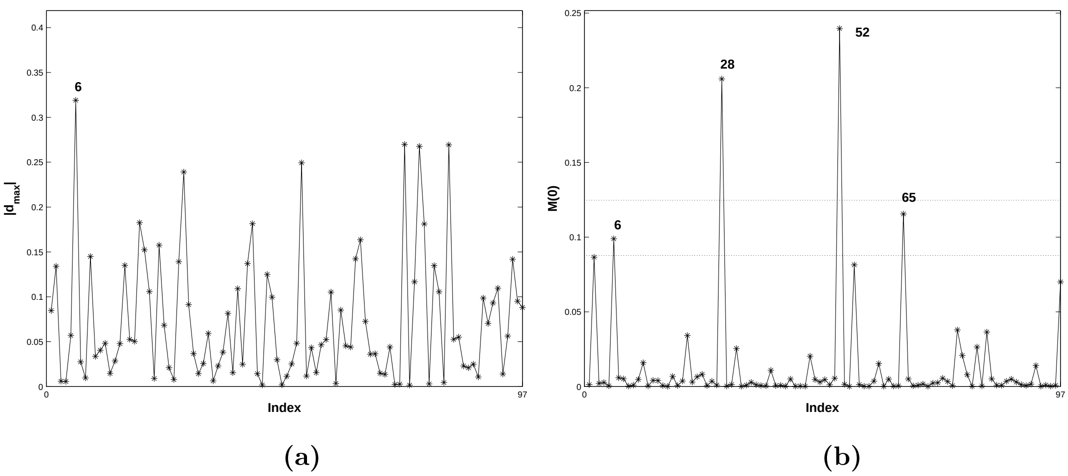
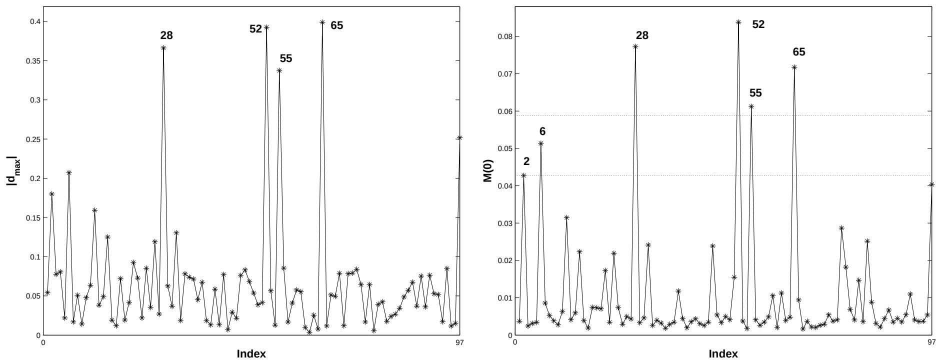
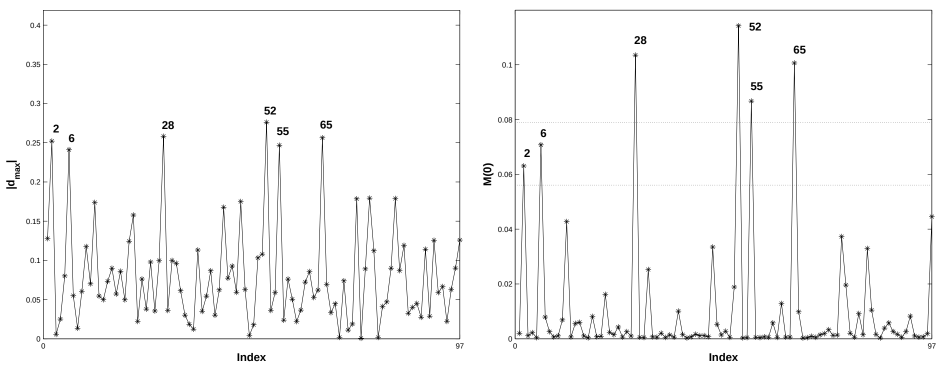


Figure 4: Data set about Life's Quality. Diagnostics of case weight perturbation for (a) Cook's approach and (b) Zhu and Lee's approach (benchmark with $c^* = 2$ and 3).



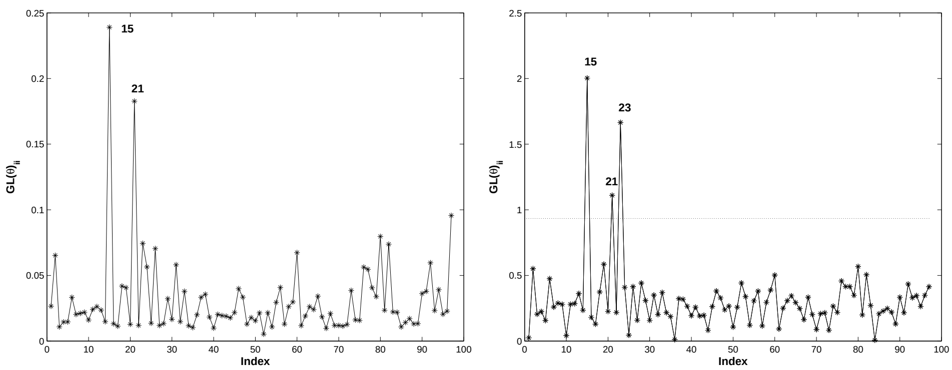
(a) **(b)**

Figure 5: Data set about Life's Quality. Diagnostics of response variable perturbation for (a) Cook's approach and (b) Zhu and Lee's approach (benchmark with $c^* = 2$ and 3).



(a) **(b)**

Figure 6: Data set about Life's Quality. Diagnostics of the explanatory variable perturbation for (a) Cook's approach and (b) Zhu and Lee's approach (benchmark with $c^* = 2$ and 3).



(a) **(b)**

Figure 7: Data set about Life's Quality. Generalized leverage for (a) Cook's approach and (b) Zhu and Lee's approach (benchmark with $c_0 = 3$).