Statistical diagnostics for nonlinear regression models based on scale mixtures of skew-normal distributions

Aldo M. Garay, Filidor V. Labra, Víctor H. Lachos*

Departamento de Estatística, Universidade Estatual de Campinas, Campinas, São Paulo, SP-Brazil

Edwin M. M. Ortega

Departamento de Ciências Exatas, Universidade de São Paulo, Piracicaba, SP-Brazil

---

Abstract

The purpose of this paper is to develop diagnostics analysis for nonlinear regression models under scale mixtures of skew-normal distributions (Branco and Dey, 2001). This novel class of models provides a useful generalization of the symmetrical nonlinear regression models (Vanegas and Cysneiros, 2010) since the random terms distributions cover both symmetric as well as asymmetric and heavy-tailed distributions such as skew-t, skew-slash, skew-contaminated normal, among others. The main virtue of considering the nonlinear regression model under the class of scale mixtures of skew-normal distributions is that they have a nice hierarchical representation which allows an easy implementation of inference procedures. A simple EM-type algorithm for iteratively computing maximum likelihood estimates is presented and the observed information matrix is derived analytically. We discuss a score test for testing the homogeneity of the scale parameter and its properties are investigated through Monte Carlo simulations. Furthermore, local influence measures and the one-step approximations of the estimates in the case-deletion model are obtained. The newly developed procedures are illustrated considering a real data previously analyzed under normal and skew-normal nonlinear regression models.

Keywords: Case-deletion model, EM algorithm, Homogeneity, Nonlinear regression models, Scale mixtures of skew-normal distributions, Score test.

---

*Corresponding author: Departamento de Estatística, Universidade Estatual de Campinas, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil. CEP 13083-859 – Caixa Postal 6065

Email addresses: amedina@ime.unicamp.br (Aldo M. Garay), fily@ime.unicamp.br (Filidor V. Labra), hlachos@ime.unicamp.br (Víctor H. Lachos*), edwin@esalq.usp.br (Edwin M. M. Ortega)

1. Introduction

Normal nonlinear regression models (N-NLM) are usually applied in the sciences and engineering to model symmetrical data for which mathematical nonlinear functions of unknown parameter are postulated with the aim of explaining or describing the phenomena under study. But N-NLM suffers from the same lack of robustness against departures from distributional assumptions as other statistical models based on the Gaussian distribution and may be too restrictive to provide an accurate representation of the structure that is present in the data (Azzalini and Capitanio, 1999). To overcome the aforementioned deficiency, some proposals have been made in the literature by replacing the normality assumption by for more flexible classes of distributions. For instance, Cysneiros and Vanegas (2008) study the symmetrical nonlinear regression model and performed an analytical and empirical study to describe the behavior of the standardized residuals. Vanegas and Cysneiros (2010) propose diagnostic procedures based on case-deletion model for symmetrical nonlinear regression models. Cancho et al. (2009) introduce the skew-normal nonlinear regression models (SN-NLM) and they present a complete likelihood based analysis, including an efficient EM algorithm to maximum likelihood estimation. Xie et al. (2009a) and Xie et al. (2009b) develop score test statistics for testing homogeneity in the SN-NLM proposed by Cancho et al. (2009). A common feature of these classes of NLM is that the N-NLM is also a member of the same class.

However, it is known that the parameter estimates of a skew-normal based model are also sensitive to atypical observations (Montenegro et al., 2009). A solution to the problem of atypical data in a asymmetrical context was postulated by Branco and Dey (2001), who proposed to use scale mixtures of skew-normal distributions (SMSN) in order to deal simultaneously with skewness and heavy-tails. Interestingly, this rich class contains the entire family of scale mixtures of normal distributions (Lange and Sinsheimer, 1993) and some skewed versions of classical symmetric distributions such as the skew-t (ST), skew-slash (SSL) and the skew contaminated normal (SCN) distributions. In this article, we extend the SN-NLM by assuming that the model errors follows SMSN distributions, so that the SMSN-NLM is defined. The hierarchical representation of the proposed model makes possible the implementation of an EM–type algorithm, which yields computationally attractive expressions for the E and M–steps.

The assessment of robustness aspects of the parameter estimates in statistical models has been an important concern of various researchers in recent decades. The deletion methodology (CDM), which consists of studying the impact on the parameter estimates after dropping individual observations, is probably the most employed technique to detect influential observations (Cook and Weisberg, 1982). Nevertheless, research on the influence of small perturbations in the model/data on the parameter estimates has received increasing attention in recent years. This can be achieved performing the local influence analysis (Cook, 1986), a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs. Several authors have applied these methods to nonlinear regression models different to normal case; see for instance, Galea et al. (2005), Cysneiros and Vanegas (2008) and Lin et al. (2009). However, to the best of our knowledge, there are neither studies on the SMSN family and nor on influence diagnostics related this topic. Thus, we believe that the research to develop statistical tools with nonstandard assumptions in NLM is a significant contribution to this field.

Another interesting problem is that in NLM a standard assumption is that all the observations have equal variances and failure to comply with this assumption will affect the efficiency of the estimators, so it is important to develop tests that allow us to determine the presence or absence of such homogeneity. In recent years several authors have proposed tests for heterogeneity of variance in different models. Cook and Weisberg (1983) provided a score test for heteroscedasticity in regression models. Lin and Wei (2003) considered heteroscedasticity tests in nonlinear models. Cysneiros et al. (2007) developed diagnostic tests for detecting heteroscedasticity in symmetrical linear regression models and more recently Lin et al. (2009) developed a score test for testing the homogeneity of the scalar parameter in the ST-normal-NLM, introduced by Gómez et al. (2007). Following these ideas, in this paper we propose a score test for testing homogeneity of the scale parameter in the SMSN-NLM.

The paper is organized as follows. In Section 2 we present the asymmetric model as well as some inferential results, additionally an EM-type algorithm for maximum likelihood estimation is developed. In Section 3, we discuss the score test for testing homogeneity of scale parameter in SMSN-NLM. The properties of score test statistics are investigated through Monte Carlo simulations. In Section 4, we derive global influence measures for SMSN-NLM and we study the local influence of two perturbation schemes. Finally, in Section 5 we illustrate the methodology considering an application with a real data set.

2. The model and maximum likelihood estimation

In order to introduce some notations, we start with the definition of SMSM distributions. Details of the next subsection are provided in Basso et al. (2009).

2.1. SMSN distributions and main notation

A random variable Y is in the SMSN family if it can be written as

$$Y = \mu + \kappa^{1/2}(U)Z, \tag{1}$$

where μ is a location parameter, Z is skew-normal random variable with location 0, scale σ^2, skewness λ ($Z \sim SN(0, \sigma^2, \lambda)$), $\kappa(u)$ is a positive function of u, U is a random variable with distribution function $H(\cdot; \boldsymbol{\nu})$ and density $h(\cdot; \boldsymbol{\nu})$ and $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of U. Although we can deal with any κ function, in this paper we restrict our attention to the case in that $\kappa(u) = 1/u$, since it leads to good mathematical properties. Given $U = u$, we have that $Y|U = u \sim SN(\mu, u^{-1}\sigma^2, \lambda)$. Thus, the density of Y is given by

$$f(y) = 2\int_0^\infty \phi(y; \mu, u^{-1}\sigma^2)\Phi\left(\frac{u^{1/2}\lambda(y-\mu)}{\sigma}\right) dH(u; \boldsymbol{\nu}), \tag{2}$$

where $\phi(\cdot; \mu, \sigma^2)$ denotes the density of the univariate normal distribution with mean μ and variance $\sigma^2 > 0$ and $\Phi(\cdot)$ is the distribution function of the standard univariate normal distribution. The notation $Y \sim SMSN(\mu, \sigma^2, \lambda; H)$ will be used when Y has pdf (2). When H is degenerate, with $u = 1$, we obtain the $SN(\mu, \sigma^2, \lambda)$ distribution. When $\lambda = 0$, the SMSN distributions reduces to the class of scale–mixtures of the normal (SMN) distribution represented by the pdf $f_0(\mathbf{y}) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})dH(u; \boldsymbol{\nu})$.

For a random variable $Y \sim SMSN(\mu, \sigma^2, \lambda; H)$, we have that the mean and the variance are given, respectively, by

$$E[Y] = \mu + \sqrt{\frac{2}{\pi}}k_1\Delta, \; Var[Y] = \sigma^2 k_2 - \frac{2}{\pi}k_1^2\Delta^2,$$

where $\Delta = \sigma\delta$, $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$ and $k_m = E[U^{-m/2}]$. The distributions in the SMSN class that will be considered in this work are:

- *The skew–t distribution with ν degrees of freedom.* In this case we consider $U \sim Gamma(\nu/2, \nu/2)$, $\nu > 0$, in definition (2) – where $Gamma(a, b)$ denotes the gamma distribution with mean a/b. The density of Y takes the form

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1 + \frac{d}{\nu}\right)^{-\frac{\nu+1}{2}} T\left(\sqrt{\frac{\nu+1}{d+\nu}}A; \nu+1\right), \quad y \in \mathbb{R}, \tag{3}$$

  where $d = (y-\mu)^2/\sigma^2$ and $T(\cdot; \nu)$ denotes the distribution function of the standard Student–t distribution, with location zero, scale one and ν degrees of freedom, namely $t(0, 1, \nu)$. We use the notation $Y \sim ST(\mu, \sigma^2, \lambda; \nu)$. A particular case of the skew-t distribution is the skew–Cauchy distribution, when $\nu = 1$. Also, when $\nu \to \infty$, we get the skew-normal distribution as the limiting case. Applications of the skew–t distribution in robust estimation can be found in Lin et al. (2007) and Azzalini and Genton (2008).

- *The skew–slash distribution.* In this case we have $U \sim Beta(\nu, 1)$ with positive shape parameter ν, where $Beta(a, b)$ denotes the beta distribution with parameters a and b, and we use the notation $Y \sim SSL(\mu, \sigma^2, \lambda; \nu)$. The density of Y is given by

$$f(y) = 2\nu \int_0^1 u^{\nu-1}\phi(y; \mu, u^{-1}\sigma^2)\Phi(u^{1/2}A)du, \quad y \in \mathbb{R}. \tag{4}$$

  The skew-slash is a heavy-tailed distribution having as limiting distribution the skew-normal one (when $\nu \to \infty$). Applications can be found in Wang and Genton (2006).

- *The skew contaminated normal distribution.* Here U is a discrete random variable taking one of two states. The probability function of U is given by

$$h(u; \boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \;\; 0 < \nu < 1, \, 0 < \gamma \leq 1,$$

  where $\boldsymbol{\nu} = (\nu, \gamma)^\top$. We denote it by $Y \sim SCN(\mu, \sigma^2, \lambda; \nu, \gamma)$. It follows immediately that

$$f(y) \;\; = \;\; 2\{\nu\phi(y; \mu, \gamma^{-1}\sigma^2)\Phi(\gamma^{1/2}A) + (1-\nu)\phi(y; \mu, \sigma^2)\Phi(A)\}.$$

  The parameters ν and γ can be interpreted as the proportion of outliers and a scale factor, respectively. The skew contaminated normal distribution reduces to the skew-normal distribution when $\gamma = 1$.

2.2. *The SMSN nonlinear regression model*

The nonlinear regression model based on SMSN distributions (hereafter SMSN-NLM) is defined as

$$Y_i = \eta(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{5}$$

where the Y_i are responses, $\eta(.)$ is an injective and twice continuously differentiable function with respect to the parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\mathbf{x}_i$ is a vector of explanatory variable values and the random errors $\varepsilon_i \sim SMSN(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda; H)$, which corresponds to the regression model where the error distribution has mean zero and hence the regression parameters are all comparable. From Lemma 2 given in Basso et al. (2009) we have that

$$E[Y_i] = \eta(\boldsymbol{\beta}, \mathbf{x}_i), \; Var[Y_i] = k_2\sigma^2 - b^2\Delta^2,$$

where $b = -\sqrt{\frac{2}{\pi}}k_1$, k_1 and k_2 are as defined in Subsection 2.1 and hence $Y_i \sim SMSN(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + b\Delta, \sigma^2, \lambda; H)$, for $i = 1, \ldots, n$. As recommended by Lange et al. (1989) and Berkane et al. (1994), who pointed out difficulties in estimating $\boldsymbol{\nu}$ due to problems of unbounded and local maximum in the likelihood function, we taken the value of $\boldsymbol{\nu}$ to be known. Thus, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$ given the observed sample $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is given by $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$, where

$$\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 + \log K_i, \tag{6}$$

with $K_i = \int_0^\infty u_i^{1/2} \exp\{-\frac{1}{2}u_i d_i\}\Phi(u_i^{1/2}A_i)dH(u_i)$, $d_i = (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta)^2/\sigma^2$ is the Mahalanobish distance, and $A_i = \lambda(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta)/\sigma$. The score function $\mathbf{U} = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and the observed information matrix $\mathbf{J} = -\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ can be obtained easily as a byproduct from the results given in Section 3.

Since one has a closed-form expression for the observed information matrix for $\boldsymbol{\theta}$, the Newton-Raphson method can be easily applied to get the ML estimates. Starting from an initial point $\widehat{\boldsymbol{\theta}}^{(0)}$, the NR procedure proceeds according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \widehat{\mathbf{J}}^{(k)-1}\widehat{\mathbf{U}}^{(k)}, \tag{7}$$

where $\widehat{\mathbf{U}}^{(k)}$ and $\widehat{\mathbf{J}}^{(k)}$ are the score vector and the observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}^{(k)}$, respectively. An oft-voiced complaint of the NR algorithm is that it may not converge unless good starting values are used. In the next section we discuss a technique more elaborate to find the ML estimates of the parameters vector $\boldsymbol{\theta}$, based on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977)

2.3. *Parameter estimation via the EM-algorithm*

In this subsection we develop an EM-type algorithm to get the ML estimates. In order to do this, we first represent the SMSN-NLM in an incomplete data framework by using the Lemma 2 given in Basso et al. (2009). We consider the following hierarchical representation for Y_i

$$Y_i | T_i = t_i \sim N_1(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + \Delta t_i, U_i^{-1}\Gamma), \tag{8}$$

$$T_i | U_i \sim TN_1(b, u_i^{-1})I(b, \infty), \tag{9}$$

$$U_i \sim H(.; \boldsymbol{\nu}) \tag{10}$$

where $\Gamma = (1-\delta^2)\sigma^2$, $\quad \Delta = \sigma\delta$ and $TN_1(r,s)I(b,\infty)$ denotes the truncated univariate normal distribution on (b,∞) with mean r and variance s before truncation. An useful straightforward result is that the conditional distribution of T_i given y_i and u_i is $TN_1(\mu_{Ti}+b, u_i^{-1}M_T^2)I(b,\infty)$, with

$$M_T^2 = \frac{\Gamma}{\Delta^2+\Gamma}, \; \mu_{T_i} = \frac{\Delta}{\Delta^2+\Gamma}(y_i - \eta(\boldsymbol{\beta},\mathbf{x}_i) - \Delta b).$$

Now we proceed for the E-step of the algorithm. To represent the estimator of the parameter $\xi = g(\boldsymbol{\theta})$, we will use the general notation $\widehat{\xi} = g(\widehat{\boldsymbol{\theta}})$, where $g(\cdot)$ is a generic function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$. Thus, let $\mathbf{y} = (y_1,\ldots,y_n)^\top$, $\mathbf{t} = (t_1,\ldots,t_n)^\top$ and $\mathbf{u} = (u_1,\ldots,u_n)^\top$. It follows that the complete log-likelihood function associated with $(\mathbf{y},\mathbf{t},\mathbf{u})$ is given by

$$\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{t},\mathbf{u}) = c - \frac{n}{2}\log\Gamma - \frac{1}{2\Gamma}\sum_{i=1}^n u_i(y_i - \eta(\boldsymbol{\beta},\mathbf{x}_i) - \Delta t_i)^2, \tag{11}$$

where c is a constant that is independent of $\boldsymbol{\theta}$. Letting $\widehat{u}_i = E[U_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_i]$, $\widehat{ut}_i = E[U_i t_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_i]$, $\widehat{ut_i^2} = E[U_i t_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_i]$ and using known properties of conditional expectation we obtain

$$\widehat{ut}_i = \widehat{u}_i(\widehat{\mu}_{Ti} + b) + \widehat{M}_T\widehat{\tau}_{1_i}, \quad \widehat{ut_i^2} \;=\; \widehat{u}_i(\widehat{\mu}_{T_i}+b)^2 + \widehat{M_T^2} + \widehat{M}_T(\widehat{\mu}_{T_i} + 2b)\widehat{\tau}_{1_i}, \tag{12}$$

where

$$\widehat{\tau}_{1_i} = E\left[U_i^{1/2} W_\Phi\left(\frac{U_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T}\right)|\widehat{\boldsymbol{\theta}}, y_i\right].$$

In each step, the conditional expectations $\widehat{u}_i = \widehat{u}_{1_i}$ and $\widehat{\tau}_{1_i}$ can be easily derived from the results given in Basso et al. (2009). For the skew–t, skew-slash and skew contaminated normal distribution we have computationally attractive expressions that can be easily implemented.

These expressions are quite useful in implementing the M-step, which consists in maximizing the expected complete data function or the $Q-$function over $\boldsymbol{\theta}$, given by

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= E[\ell_c(\boldsymbol{\theta})|\mathbf{y},\widehat{\boldsymbol{\theta}}^{(k)}] = c - \frac{n}{2}\log(\Gamma) - \frac{1}{2\Gamma}\sum_{i=1}^n \Big[\widehat{u}_i^{(k)}\,(y_i - \eta(\boldsymbol{\beta},\mathbf{x}_i))^2 \\ &- 2\Delta(y_i - \eta(\boldsymbol{\beta},\mathbf{x}_i))\widehat{ut}_i^{(k)} + \Delta^2\widehat{ut_i^2}^{(k)}]\Big], \end{aligned}$$

where $\widehat{\boldsymbol{\theta}}^{(k)}$ is an updated value of $\widehat{\boldsymbol{\theta}}$.

When the M-step turns out to be analytically intractable, it can be replaced with a sequence of conditional maximization (CM) steps. The resulting procedure is known as *ECM algorithm* (Meng and Rubin, 1993). Next, we describe this EM-type algorithm (ECM) for maximum likelihood estimation of the parameters of the SMSN-NLM defined in (5).

E-step: Given a current estimate $\widehat{\boldsymbol{\theta}}^{(k)}$, compute $\widehat{u}_i^{(k)}$ $\widehat{ut}_i^{(k)}$, $\widehat{ut_i^2}^{(k)}$, for $i = 1,\ldots,n$.

CM-step: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following nice expressions

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \text{argmin}_{\boldsymbol{\beta}}(\mathbf{z}^{(k)} - \boldsymbol{\eta}(\boldsymbol{\beta},\mathbf{x}))^\top \widehat{\mathbf{U}}^{(\mathbf{k})}(\mathbf{z}^{(k)} - \boldsymbol{\eta}(\boldsymbol{\beta},\mathbf{x})), && (13)\\
\widehat{\Delta}^{(k+1)} &= \frac{\sum_{i=1}^n \widehat{ut}_i^{(k)}(y_i - \eta(\boldsymbol{\beta}^{(k+1)},\mathbf{x}_i))}{\sum_{i=1}^n \widehat{ut_i^2}^{(k)}}, && (14)\\
\widehat{\Gamma}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n \Big((y_i - \eta(\boldsymbol{\beta}^{(k+1)},\mathbf{x}_i))^2\widehat{u}_i^{(k)} - 2\Delta^{(k+1)}(y_i - \eta(\boldsymbol{\beta}^{(k+1)},\mathbf{x}_i))\widehat{ut}_i^{(k)}\\
&\quad + (\Delta^2)^{(k+1)}\widehat{ut^2}_i^{(k)}\Big), && (15)
\end{aligned}$$

where $\widehat{\mathbf{U}}^{(k)} = diag(\widehat{u}_1^{(k)},\ldots,\widehat{u}_n^{(k)})$, $\mathbf{z}^{(k)}$ is the corrected observed response given by $\mathbf{z}^{(k)} = \mathbf{y} - \widehat{\Delta}^{(k)}\widehat{\boldsymbol{\tau}}^{(k)}$, with $\widehat{\boldsymbol{\tau}}^{(k)} = (\widehat{\tau}_1^{(k)},\ldots,\widehat{\tau}_n^{(k)})^\top$, $\widehat{\tau}_i^{(k)} = \widehat{ut}_i^{(k)}/\widehat{u}_i^{(k)}$ and $\boldsymbol{\eta}(\boldsymbol{\beta},\mathbf{x}) = (\eta(\boldsymbol{\beta},\mathbf{x}_1),\ldots,\eta(\boldsymbol{\beta},\mathbf{x}_n))^\top$. This process is iterated until a suitable convergence rule is satisfied, e.g. if $||\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}||$ is sufficiently small, or until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta})$, like $||\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})||$ or $||\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1||$, is small enough. An interesting observation is that the M–step to estimate $\boldsymbol{\beta}$ is equivalent to the weighted nonlinear least squares in the NLM, $\mathbf{z} = \boldsymbol{\eta}(\boldsymbol{\beta},\mathbf{x}) + \boldsymbol{\epsilon}$, in which reliable and efficient implementation of algorithms are available in softwares as SAS, R, Ox and Matlab. To recover σ^2 and λ, we observe that

$$\sigma^2 = \Delta^2 + \Gamma \text{ and } \lambda = \Delta/\sqrt{\Gamma}.$$

3. Score Test for Homogeneity of Variance

The model defined in (5) assumes that $Var(Y_i) = \sigma^2\left[k_2 - b\delta^2\right]$ is constant, in which the scale parameter σ^2 is constant. However, similar to the dispersion parameter mentioned by Lin et al. (2009), the actual scalar parameter may be related to the ith observation. Then, one cannot make any inference for the model without further assumptions, since there are too many unknown parameters involved. Hence, it is necessary to test homogeneity of the scalar parameter. This section concentrates on this problem in SMSN-NLM.

Following Lin et al. (2009), we generalized the scale parameter σ^2 by σ_i^2, where σ_i^2 is modeled by

$$\sigma_i^2 = \sigma^2 m(\mathbf{z}_i,\boldsymbol{\rho}), \tag{16}$$

where $m(.)$ is an injective and twice continuously differentiable function with respect to the parameter $\boldsymbol{\rho} = (\rho_1,\ldots,\rho_{q_1})^\top$, and $\mathbf{z}_i$ is a vector of explanatory variable values, which constitute in general, although not necessary, a subset of $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}$. Now let $m_i = m(\mathbf{z}_i,\boldsymbol{\rho})$, we assume that exists a unique value $\boldsymbol{\rho}_0$ of $\boldsymbol{\rho}$ such that $m(\mathbf{z}_i,\boldsymbol{\rho}_0) = 1$, for all i. Obviously, if $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ then $\sigma_i^2 = \sigma^2$ and Y_i have constant variance. With this consideration, the test for homogeneity of scalar parameter in model defined in (5), under the above assumptions, can be expressed by

$$H_0 : \boldsymbol{\rho} = \boldsymbol{\rho}_0 \text{ vs } H_0 : \boldsymbol{\rho} \neq \boldsymbol{\rho}_0. \tag{17}$$

In the remainder of the this section, let $\boldsymbol{\theta}_1 = \boldsymbol{\rho}$ be the parameter of interest and the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$ is considered a nuisance parameter, then of (5) and (16), the log-likelihood function for $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}^\top)^\top$ given the observed sample $\mathbf{y} = (y_1, \ldots, y_n)^\top$ can be written as

$$\ell(\boldsymbol{\theta}_2) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}_2) = \sum_{i=1}^n \left[\log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log m_i + \log K_i\right], \tag{18}$$

where $K_i = \int_0^\infty u^{1/2}\exp\{-\frac{1}{2}ud_i\}\Phi(u^{1/2}A_i)dH(u)$, $d_i = (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta m_i^{1/2})^2/(m_i\sigma^2)$ is the Mahalanobish distance, $A_i = \lambda(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta m_i^{1/2})/(m_i\sigma^2)$. The score function is given by $U(\boldsymbol{\theta}_2) = \dfrac{\partial\ell(\boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_2} = \sum_{i=1}^n U_i(\boldsymbol{\theta}_2)$ where

$$U_i(\boldsymbol{\theta}_2) = \frac{\partial\ell_i(\boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_2} = -\frac{1}{2}\frac{\partial\log\sigma^2}{\partial\boldsymbol{\theta}_2} - \frac{1}{2}\frac{\partial\log m_i}{\partial\boldsymbol{\theta}_2} + \frac{1}{K_i}\frac{\partial K_i}{\partial\boldsymbol{\theta}_2}, \tag{19}$$

with

$$\frac{\partial K_i}{\partial\boldsymbol{\theta}_2} = I_i^\phi(1)\frac{\partial A_i}{\partial\boldsymbol{\theta}_2} - \frac{1}{2}I_i^\Phi\left(\frac{3}{2}\right)\frac{\partial d_i}{\partial\boldsymbol{\theta}_2}$$

and

$$I_i^\Phi(w) = \int_0^\infty u^w e^{(-ud_i/2)}\Phi_1(u^{1/2}A_i)dH(u) \quad \text{and} \quad I_i^\phi(w) = \frac{1}{\sqrt{2\pi}}\int_0^\infty u^w e^{\left(-u(d_i+A_i^2)/2\right)}dH(u).$$

The observed information matrix is given by

$$\mathbf{J}(\boldsymbol{\theta}_2) = \begin{bmatrix} \mathbf{J}_{\rho\rho} & \mathbf{J}_{\rho\theta} \\ \mathbf{J}_{\rho\theta} & \mathbf{J}(\boldsymbol{\theta}) \end{bmatrix} = -\sum_{i=1}^n \frac{\partial^2\ell_i(\boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top}, \tag{20}$$

where

$$\frac{\partial^2\ell_i(\boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} = -\frac{1}{2}\frac{\partial^2\log\sigma^2}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} - \frac{1}{2}\frac{\partial^2\log m_i}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} - \frac{1}{(K_i)^2}\frac{\partial K_i}{\partial\boldsymbol{\theta}_2}\frac{\partial K_i}{\partial\boldsymbol{\theta}_2^\top} + \frac{1}{K_i}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top},$$

with

$$\begin{aligned}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} &= \frac{1}{4}I_i^\Phi\left(\frac{5}{2}\right)\frac{\partial d_i}{\partial\boldsymbol{\theta}_2}\frac{\partial d_i}{\partial\boldsymbol{\theta}_2^\top} - \frac{1}{2}I_i^\Phi\left(\frac{3}{2}\right)\frac{\partial^2 d_i}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} - \frac{1}{2}I_i^\phi(2)\left(\frac{\partial A_i}{\partial\boldsymbol{\theta}_2}\frac{\partial d_i}{\partial\boldsymbol{\theta}_2^\top} + \frac{\partial d_i}{\partial\boldsymbol{\theta}_2}\frac{\partial A_i}{\partial\boldsymbol{\theta}_2^\top}\right) \\ &\quad - I_i^\phi(2)\,A_i\frac{\partial A_i}{\partial\boldsymbol{\theta}_2}\frac{\partial A_i}{\partial\boldsymbol{\theta}_2^\top} + I_i^\Phi(1)\frac{\partial^2 A_i}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top}.\end{aligned}$$

The derivatives of d_i and A_i involve standard algebraic manipulations and are given in the Appendix A. It is important to note when $\boldsymbol{\rho} = \boldsymbol{\rho}_0$, the submatrix $\mathbf{J}(\boldsymbol{\theta})$ in (20) represents the observed information matrix for the model with constant variance.

From Basso et al. (2009), we have the following results related with $I_i^\phi(w)$ and $I_i^\Phi(w)$ for each element of the SMSN class that we are considering:

- *Skew-t:* Letting $\nu_w = 2w + \nu$, we have

$$\begin{aligned} I_i^\Phi(w) &= \frac{2^w\nu^{\nu/2}\Gamma(\nu_w/2)}{\Gamma(\nu/2)(\nu + d_i)^{\nu_w/2}}T\left(\sqrt{\frac{\nu_w}{d_i + \nu}}\mathrm{A}_i; \nu_w\right) \quad \text{and} \\ I_i^\phi(w) &= \frac{2^w\nu^{\nu/2}\Gamma(\nu_w/2)}{\sqrt{2\pi}\Gamma(\nu/2)(d_i + \mathrm{A}_i^2 + \nu)^{\nu_w/2}};\end{aligned}$$

- *Skew–slash:* Letting $\nu_w = w + \nu$, we have

$$I_i^{\Phi}(w) = \frac{2^{2+w}\Gamma(\nu_w)}{d_i^{\nu_w}} P_1(\nu_w, d_i/2)\, E\{\Phi(S_i^{1/2}\mathrm{A}_i)\} \quad \text{and}$$
$$I_i^{\phi}(w) = \frac{\nu 2^{\nu_w}\Gamma(\nu_w)}{\sqrt{2\pi}(d_i + \mathrm{A}_i^2)^{\nu_w}} P_1\big(\nu_w, \frac{d_i + \mathrm{A}_i^2}{2}\big),\ S_i \sim Gamma(\nu_w, \frac{d_i}{2})\mathbb{I}_{(0,1)},$$

where $P_x(a,b)$ denotes the distribution function of the $Gamma(a,b)$ distribution evaluated at x.

- *Skew contaminated normal:*

$$I_i^{\Phi}(w) = \sqrt{2\pi}\{\nu\gamma^{w-1/2}\phi_1\left(\sqrt{d_i}; 0, \frac{1}{\gamma}\right)\Phi(\gamma^{1/2}A_i) + (1-\nu)\phi_1(\sqrt{d_i}; 0, 1)\Phi(A_i)\} \quad \text{and}$$
$$I_i^{\phi}(w) = \nu\gamma^{w-1/2}\phi_1\left(\sqrt{d_i + A_i^2}; 0, \frac{1}{\gamma}\right) + (1-\nu)\phi_1\left(\sqrt{d_i + A_i^2}; 0, 1\right).$$

Furthermore, the score test statistic for H_0 is of the form (Cox and Hinkley, 1974)

$$SR = U(\widehat{\boldsymbol{\theta}}_2^0)^\top \mathbf{J}_\theta^{\rho\rho}(\hat{\boldsymbol{\theta}}_2^0) U(\widehat{\boldsymbol{\theta}}_2^0),$$

where $U(\widehat{\boldsymbol{\theta}}_2^0) = \frac{\partial \ell(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\rho}}|_{\widehat{\boldsymbol{\theta}}_2 = \widehat{\boldsymbol{\theta}}_2^0}$, $\mathbf{J}_\theta^{\rho\rho}(\hat{\boldsymbol{\theta}}_2^0)$ is the upper left corner block of $\mathbf{J}^{-1}(\boldsymbol{\theta}_2)|_{\boldsymbol{\theta}_2 = \widehat{\boldsymbol{\theta}}_2^0}$ corresponding to $\boldsymbol{\rho}$, and $\hat{\boldsymbol{\theta}}_2^0 = (\boldsymbol{\rho}_0, \widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma}^2, \widehat{\lambda})^\top$ is the ML estimate of $\boldsymbol{\theta}_2$ under the null hypothesis H_0. When H_0 is true, the statistic SR is asymptotically distributed as $\chi^2_{q_1}$.

3.1. Simulation Studies

In this subsection, we study the performance of the asymptotic distribution and power of the score test statistic.

3.1.1. The empirical distributions of the score statistics

The performance of the asymptotic distribution of the score statistic SR is examined in order to compare the empirical distribution with the theoretical distribution via Monte Carlo simulations. As in Xie et al. (2009b), the model used for the simulation study is

$$Y_i = \exp(\beta x_i) + \epsilon_i,\ i = 1, \ldots, n, \tag{21}$$

where $\varepsilon_i \sim SMSN(-\sqrt{\frac{2}{\pi}}k_1\sigma_i\delta, \sigma_i^2, \lambda; H)$, with $\sigma_i^2 = \sigma^2 m(x_i, \rho) = \sigma^2 \exp(\rho x_i)$. The simulation is performed for test the homogeneity of scalar parameter as discussed above. The true values of the parameters are set as $\beta = 2, \sigma^2 = 0.5, \lambda = 1$. In our analysis we use SN, ST with $\nu = 3$ and the SSL distribution with $\nu = 3$. The explanatory variable x, was generated following uniform distribution in the interval (0.2, 2) and their values were held fixed throughout the simulations. To get values of Y_i, a random variable is drawn from the model (21) with the true values of parameters, the values de x_i and $\rho = 0$ (H_0), repeating this procedure 2,000 times. Then the empirical distribution functions of the score statistic are obtained and recorded. For $n = 30, 50, 70, 90,$ and 150 the comparisons between the empirical distribution functions of the score statistics an the distribution function of $\chi^2_{(1)}$ are depicted in Figure 1. These figures shows that the empirical distributions functions of the score statistic are very close to the theoretical distribution $\chi^2_{(1)}$ for all the SMSN models considered in our study.

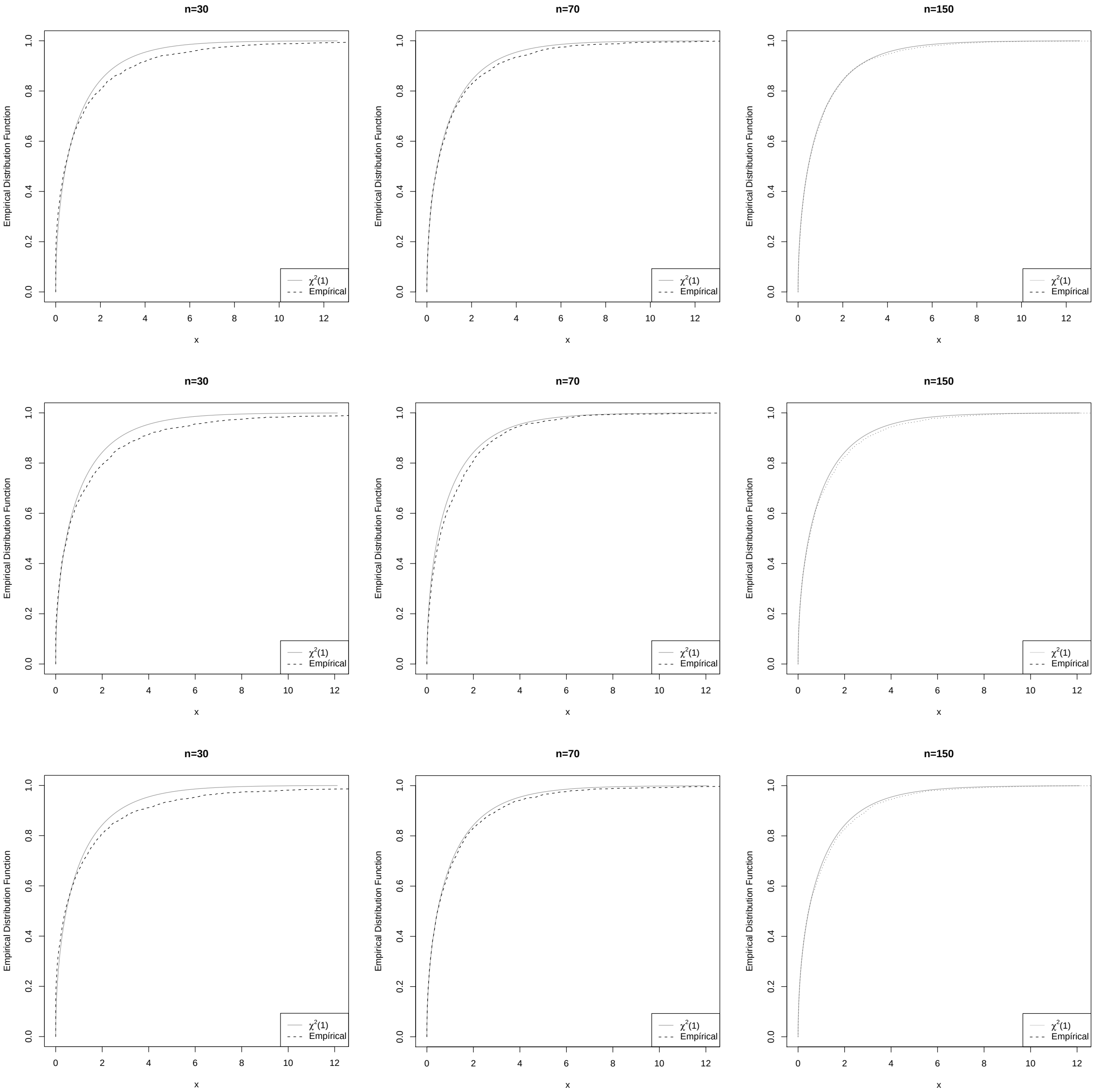

Figure 1: Simulated comparisons between the empirical distribution functions of the score statistic and $\chi^2_{(1)}$, using SN in the first row, ST in the second row and SSL in the last row.

3.1.2. Power of score test

The design considered in this simulation study is the same as in the previous subsection and the simulation was performed for different values of n and ρ to get the simulated sizes and powers for the test statistic. We take $\rho = 0, 0.2, 0.4, 0.6, 0.8$ and 1, and $n = 30$, 50, 70, 90 and 150. Each simulated case was replied 2,000 times, then the proportion of times which rejected the null hypothesis was just the simulated valued of power. Here all the statistics are compared with the χ^2_1 critical value at $\alpha = 0.05$ level.

Table 1 present the simulated sizes and powers for the score test statistic SC. From this table, we can see that for $n = 30$ the sizes of the test statistics SR are 0.082 for the SN, 0.0875 for the ST and

0.0945 for the SSL distribution at the 0.05 level and hence this test statistic is very conservative in all cases. However, when $n \geq 50$, the actual size of the test is close to 0.05. As ρ and n increase, the power of this test statistic SR approach 1 quickly as depicted in Figure 2. This figure shows that if the sample size is moderate or large, the proposed score test statistic SR can detect heteroscedasticity of the scalar parameter and consequently of the variance very well.

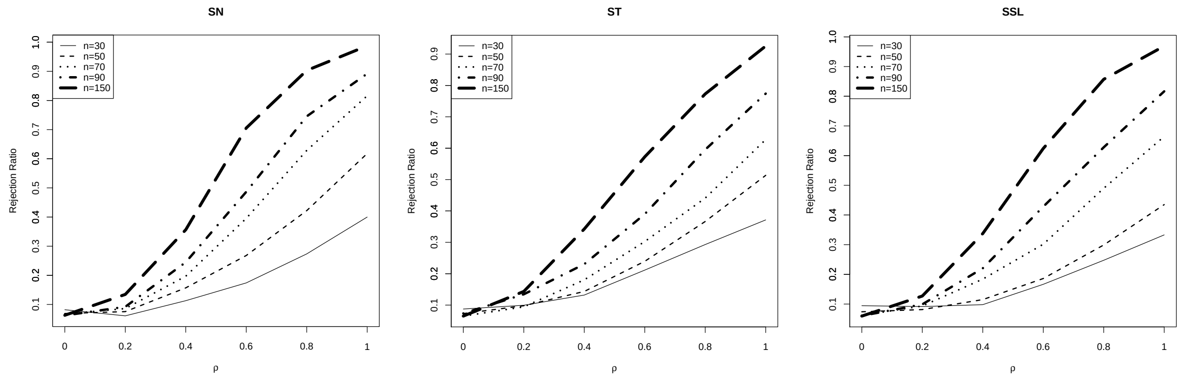

Figure 2: Power of the analysis to detect significant heteroscedasticity over a range of possible ρ_0 values, sample sizes (n), and for three errors distributions in the model (21).

Table 1: Rejection rates of the hypothesis $H_0 : \rho = 0$ at nominal level of 5% from the statistic test SR for three errors distributions in the model (21).

SN-NLM						
n	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
30	0.0820	0.0610	0.1130	0.1740	0.2735	0.4000
50	0.0690	0.0755	0.1570	0.2680	0.4220	0.6180
70	0.0670	0.0860	0.1970	0.3955	0.6285	0.8160
90	0.0625	0.0920	0.2450	0.4865	0.7455	0.8920
150	0.0630	0.1345	0.3565	0.7055	0.9030	0.9875

ST-NLM						
n	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
30	0.0875	0.0995	0.1320	0.2115	0.2930	0.3715
50	0.0730	0.0980	0.1435	0.2395	0.3665	0.5135
70	0.0650	0.0945	0.1805	0.3025	0.4410	0.6275
90	0.0735	0.1345	0.2305	0.3900	0.5955	0.7740
150	0.0650	0.1435	0.3425	0.5725	0.7730	0.9255

SSL-NLM						
n	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
30	0.0945	0.0915	0.0980	0.1665	0.2475	0.3330
50	0.0740	0.0815	0.1150	0.1860	0.2990	0.4350
70	0.0620	0.0935	0.1840	0.3015	0.4900	0.6645
90	0.0605	0.0990	0.2205	0.4280	0.6280	0.8165
150	0.0595	0.1270	0.3375	0.6240	0.8570	0.9690

4. Influence diagnostics

There are basically two approaches to detecting observations that seriously influence the results of a statistical analysis. One approach is the case-deletion approach, in which the impact of deleting an observation on the estimates is directly assessed by measures such as the likelihood distance and Cook's distance (see, Cook, 1977). The second approach is one in which the stability of the estimated outputs with respect to the model inputs is studied via various minor model perturbation schemes such as the local influence approach developed in Cook (1986). In the following subsections we describe the background and details of the classical diagnostics methods to the detection of influential observations, as well as two types of perturbation schemes.

4.1. Case deletion model

The identification of observations with a disproportionate influence in the estimates of the parameters is a fundamental component of the process of model validation. The presence of these types of observations can become inadequate inference. An important approach for the identification of influential observations can be based on the methodology known as case-deletion model (CDM), proposed by Cook (1977) for the normal linear regression models. To study the influence of i-th observation in the maximum likelihood estimate of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$, it is usual to compare the estimate with all observations, $\widehat{\boldsymbol{\theta}}$, and the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_{(i)}$ obtained when the i-th observation has been excluded from the data set. This approach corresponds to the case-deletion model, which can be expressed as

$$Y_j = \eta(\boldsymbol{\beta}, \mathbf{x}_j) + \varepsilon_j, \quad j \neq i,$$

where the log-likelihood function of $\boldsymbol{\theta}$ is denoted by $\ell_{(i)}(\boldsymbol{\theta}) = \sum_{j\neq i} \ell_j(\boldsymbol{\theta})$. However, to compute $\widehat{\boldsymbol{\theta}}_{(i)} = (\boldsymbol{\beta}_{(i)}^\top, \sigma^2_{(i)}, \lambda_{(i)})^\top$ for all i and to compare them with $\widehat{\boldsymbol{\theta}}$ would be very time-consuming when the total sample size n is large. Fortunately, the following result due to Cook and Weisberg (1982) gives an updating formulae under case deletion to avoid direct model estimation for each of the n cases. This result is essential for our case-deletion diagnostics.

$$\widehat{\boldsymbol{\theta}}_{(i)} = \widehat{\boldsymbol{\theta}} + \{\mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1} \dot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}}), \tag{22}$$

where $\dot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}}) = \partial \ell_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\partial \ell_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$. From this result, we can see the difference between the estimates with and without a case deleted and can obtain the case-deletion measures for assessing the influential observations in SMSN-NLM.

- **Generalized Cook's distance**

The generalized Cook's distance is defined as a standardized norm of $\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}$, i.e.,

$$GD_i = (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top \mathbf{M} (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}) \tag{23}$$

where $\mathbf{M}$ is a non-negative definite matrix, which measures the weighted combination of the elements for the difference $\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}$. Cook and Weisberg (1982) considered several choices for $\mathbf{M}$. A commonly used choice is the observed Fisher information matrix $\mathbf{M} = \mathbf{J}(\boldsymbol{\theta})$. Substituting Equation (22) into Equation (23), we obtain the following approximation:

$$GD_i^l = \dot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}})^\top \{\mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1} \dot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}}),\ i = 1, \ldots, n.$$

- **Likelihood distance**

The likelihood distance (Cook and Weisberg, 1982) is defined as

$$LD_i(\boldsymbol{\theta}) = 2\{\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_{(i)})\}, \tag{24}$$

Substituting (22) into (24), we obtain the following approximation:

$$LD_i^l = 2\{\ell(\boldsymbol{\theta}) - \ell(\widehat{\boldsymbol{\theta}} + \{\mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1} \dot{\ell}_{(i)}(\boldsymbol{\theta}))\},\ i = 1, \ldots, n.$$

4.2. Local influence

Case deletion is a common way to assess the effect of an observation on the estimation process. This is a global influence analysis, since the effect of the observation is evaluated by eliminating it from the data set. The work of Cook (1986), laid the foundation for assessing local influence of a group of observations when a minor perturbation is made in the statistical model or in the data set. Based on his proposal many papers have been written on the subject. In his seminal paper, Cook (1986) shows that the normal curvature for $\boldsymbol{\theta} \in \mathbb{R}^{p+2}$ in the direction of $\mathbf{d} \in \mathbb{R}^q$, $||\mathbf{d}|| = 1$ is given by $C_d(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^{*\top} \mathbf{J}^{-1} \boldsymbol{\Delta}^* \mathbf{d}|$, where $\mathbf{J}$ is the observed information matrix and $\boldsymbol{\Delta}^*$ is the $(p+2) \times q$ matrix with elements $\boldsymbol{\Delta}^*_{rs} = \partial^2 \ell(\boldsymbol{\theta}) / \partial\theta_r \partial\omega_s$, for $r = 1, \ldots, (p+2)$ and $r = 1, \ldots, q$, both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_o$ (postulated model). The suggestion here to examine the elements of the eigenvector associated with the largest eigenvalue of the matrix $\ddot{\mathbf{T}} = \boldsymbol{\Delta}^{*\top} \{\mathbf{J}\}^{-1} \boldsymbol{\Delta}^*$. Alternatively, one may also examine the total local influence $C_i = C_{d_i}(\boldsymbol{\theta})$, where $\mathbf{d}_i$ is an $q \times 1$ vector of zeros with one at the ith position.

Since $C_d(\boldsymbol{\theta})$ is not invariant under uniform change of scale, Poon and Poon (1999) proposed the conformal normal curvature $B_d(\boldsymbol{\theta}) = C_d(\boldsymbol{\theta}) / tr(2\ddot{\mathbf{T}})$. An interesting property of the conformal normal curvature is that for any unitary direction $\mathbf{d}$ one has $0 \leq B_d(\boldsymbol{\theta}) \leq 1$, which allows comparison of curvatures among different scale mixtures of normal models. In order to determine if the ith observation is possible influential, Poon and Poon (1999) proposed classify the ith observation as possible influential if $M(0)_i = B_{d_i}$, where $\mathbf{d}_i$ is an $q \times 1$ vector of zeros with one at the ith position, is greater than the benchmark

$$\bar{M(0)} + c^* SM(0),$$

where $\bar{M(0)} = 1/q$ and $SM(0)$ is the sample standard error of $\{M(0)_k,\ k = 1 \ldots, q\}$ and c^* is a selected constant. Depending on the real application, c^* may be taken to be any value. We will evaluate in the sequel the matrix $\boldsymbol{\Delta}^*$ under two perturbation schemes for the SMSN-NLM given in (5).

- **Case weight perturbation**

First, consider the following arbitrary attribution of weights for the experimental units in the log-likelihood function, which can be defined by

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \Big[\log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 + \log K_i \Big],$$

where and K_i is defined in equation (6). Note that, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is the vector of weights of the contributions from each observation to the likelihood and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ is the non perturbation point, that is, $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$. This perturbation scheme is intended to evaluate whether the contribution of the observations with differing weights affects the maximum likelihood estimator of $\boldsymbol{\theta}$. It follows after some algebraic manipulation that the delta matrix is given by

$$\boldsymbol{\Delta}^* = (\boldsymbol{\Delta}_1^*, \ldots, \boldsymbol{\Delta}_n^*),$$

where $\boldsymbol{\Delta}_i^* = \dfrac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is as given in (19) after dropping the element corresponding to $\boldsymbol{\rho}$.

- **Scale parameter perturbation**

To study the effects from departures from the homogeneity assumption regarding the scale parameter σ^2, we consider the following perturbation $\sigma^2_{\omega_i} = \sigma^2/\omega_i$. This perturbation corresponds to considering that the distribution of Y_i is heteroscedastic, once

$$Var(Y_i) = \sigma^2_{\omega_i}\left(k_2 - b\delta^2\right),$$

where k_2, δ and b as denoted in Subsection 2.2. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top$. Moreover, the perturbed log-likelihood function has the form

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \Big[\log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log\omega_i + \log K_{\omega i}, \Big],$$

where $d_{\omega i} = \omega_i^{1/2}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta\omega_i^{-1/2})^2/\sigma^2$, $A_{\omega i} = \lambda\omega_i^{1/2}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta\omega_i^{-1/2})/\sigma$ and $K_{\omega i}$ is as in Section 3, switching $d_{\omega i}$ and $A_{\omega i}$ with d_i and A_i, respectively. The matrix $\boldsymbol{\Delta}^* = (\boldsymbol{\Delta}_\beta^{*\top}, \boldsymbol{\Delta}_{\sigma^2}^{*\top}, \boldsymbol{\Delta}_\lambda^{*\top},)^\top$ is given in Appendix B.

4.3. Residuals

Residual analysis aims at identifying atypical observations and/or model misspecification once residuals are measures of agreement between the data and the fitted model. Most residuals are based on the differences between the observed responses and the fitted conditional mean. We defined the following standardized ordinary residual (Pearson residuals):

$$r_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{Var}(y_i)}}, \; i = 1, \ldots, n,$$

where $\widehat{Var}(y_i) = k_2\widehat{\sigma}^2 - \frac{2}{\pi}k_1^2\widehat{\sigma}^2\widehat{\delta}^2$. Here, $\widehat{\mu}_i = \eta(\widehat{\boldsymbol{\beta}}, \mathbf{x}_i)$, and $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$ and $\widehat{\delta}$ denoting the maximum likelihood estimators of $\boldsymbol{\beta}$, σ^2 and δ, respectively. We also generate envelopes, as suggested by Atkinson (1981), to detect incorrect specification of the error distribution and the systematic component $\eta(\boldsymbol{\beta}, \mathbf{x}_i)$ as well as the presence of outlying observations.

5. Application: Oil palm yield data.

The oil palm yield data data have been analyzed by Cancho et al. (2009) using the SN-NLM and assuming a nonlinear growth-curve model - see Figure 7(b). We illustrate our methods, replacing the SN assumption by the SMSN class of distributions as follows:

$$Y_i = \frac{\beta_1}{1+\beta_2\exp(-\beta_3 x_i)} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} SMSN(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda; H), \tag{25}$$

for $i = 1, \ldots, 19$, where H denote the distribution function for the mixture variable U. In our analysis we will assume the SN, ST and the SSL distributions from the SMSN class for comparative purposes.

5.1. *Estimation models:*

We choose the value of $\boldsymbol{\nu}$ by maximizing the the likelihood function as illustrated in Figure 3; for the ST model we found $\nu = 3$ and for the SSL we found $\nu = 2$. Actually, with $\nu = 3$ and $\nu = 2$ the variances of the slash and skew-t distributions are finite. Table 2 contains the ML estimates of the parameters from the three models, together with their corresponding standard errors calculated via the observed information matrix. The AIC model selection criterion indicate that the ST distribution present the best fit. Although the regression estimates parameters are similar in all the three fitted models (see Table 2) the standard errors of the SMSN-NLM with heavy tails are smaller than those in the SN-NLM. This suggests that the two models with longer tails than the SN model seem to produce more accurate maximum likelihood estimates. The estimates for the variance components (σ^2 and λ) are not comparable since they are on different scale.

5.2. *Influence diagnostic analysis:*

- *Case deletion model:*

  Here case-deletion measures GD_i^l and LD_i^1, as presented in Subsection 4.1, are computed. The results are displayed in Figure 4. We observe that cases (♯10, ♯13, ♯15, ♯16 and ♯18) are identified as

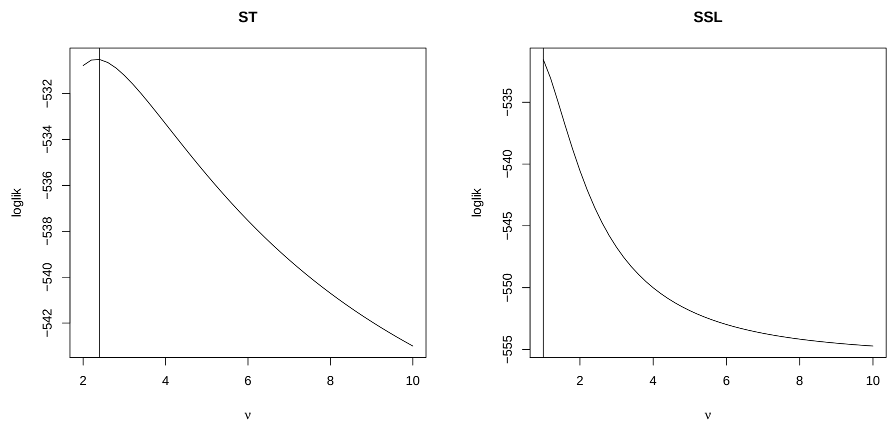

Figure 3: Plot of the profile log-likelihood of the parameter ν for fitting a ST-NLM and SSL-NLM for the oil palm yield data.

Table 2: ML estimation results for fitting various mixture models on the oil palm yield data set. SE are the asymptotic standard errors based on the observed information matrix.

	SN-NLM		ST-NLM		SSL-NLM	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
β_1	37.351	0.462	37.529	0.441	37.463	0.486
β_2	44.576	17.039	43.483	10.364	43.373	14.982
β_3	0.731	0.070	0.732	0.045	0.728	0.063
σ^2	6.919	2.655	1.644	1.152	3.105	1.708
λ	-4.453	3.125	-1.871	1.332	-3.489	2.481
ν	-	-	3	-	2	-
log-likelihood	-35.037		-33.829		-34.781	
AIC	80.074		79.659		81.562	

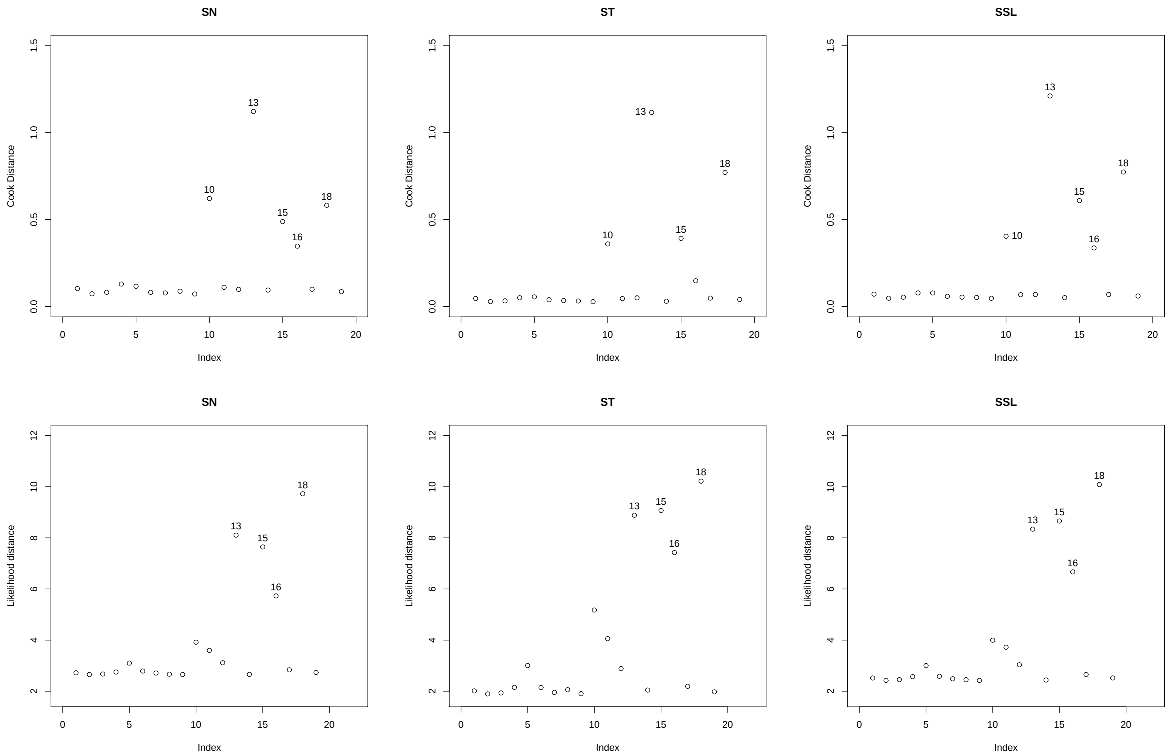

Figure 4: Index plots of a) Generalized Cook's distance GD_i^l and b) Likelihood Distance LD_i^l using SN-NLM, ST-NLM and SSL-NLM, on the oil palm yield data.

the most influential in the estimation of the parameters under the SN and SSL cases. Meanwhile, only cases (♯10, ♯13, ♯15, and ♯18) are influential under the ST-NLM.

Note however from Figure 5 that when we use distributions with tails heavier than the SN one, the EM algorithm allows to accommodate such observations attributing to them small weights in the estimation procedure. The estimated weights for the skew–normal distribution ($\widehat{u}_i, i = 1, \ldots, 19$) are indicated in Figure 5 as a continuous line. Therefore, this rich class of distributions may naturally attribute different weights to each observation and consequently control the influence of a single observation on the parameter estimates. These results agree with similar considerations, presented in Osorio et al. (2007), in a symmetric context.

Next we conduct a local influence study with interest focussing on $\boldsymbol{\theta}$. The perturbation schemes described in the Section 4.2 are considered and in all cases we consider the benchmark for M(0) with $c^* = 1.96$

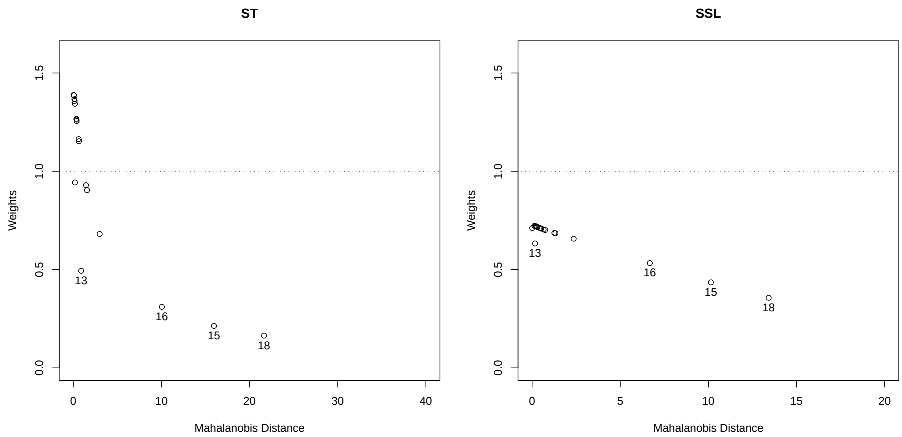

Figure 5: Estimated u_i for the ST-NLM and the SSL-NLM, on the oil palm yield data.

- *Case weight perturbation*

  Under this perturbation scheme, we obtain $C_{\mathbf{d}_{maxSN}} = 3.55$, $C_{\mathbf{d}_{maxST}} = 2.28$ and $C_{\mathbf{d}_{maxSSL}} = 2.13$, as values of maximum curvature. From Figure 6 (first row), it is noted that under the ST-NLM and SSL-NLM, the observation 13 is identified as influential. In addition, the observation 18 is also identified as influential under the SN-NLM.

- *Scale perturbation*

  In this case, the value of the maximum curvature are $C_{\mathbf{d}_{maxSN}} = 6.40$, $C_{\mathbf{d}_{maxST}} = 2.10$ and $C_{\mathbf{d}_{maxSSL}} = 4.82$. The ML estimators are quite stable with respect to this perturbation scheme in the ST case, as displayed in Figure 6 (second row). However, it is appreciated some influence of the observation 13 under the SN-NLM and SSL-NLM.

It is important to note that as expected, the influence of the observations is reduced when we consider distributions with heavier tails than the SN one. For this data set, the ST model accommodates slightly better the influential observations.

5.3. Residuals Analysis:

We first perform residual analysis for the ST-NLM fit by plotting the Pearson residuals r_i against the explanatory variable x_i (see Figure 7a). This plot also shows that the residuals of the observations ♯13, ♯15, ♯16 and ♯18 are possible outliers. These observations are the same as detected by the case deletion analysis.

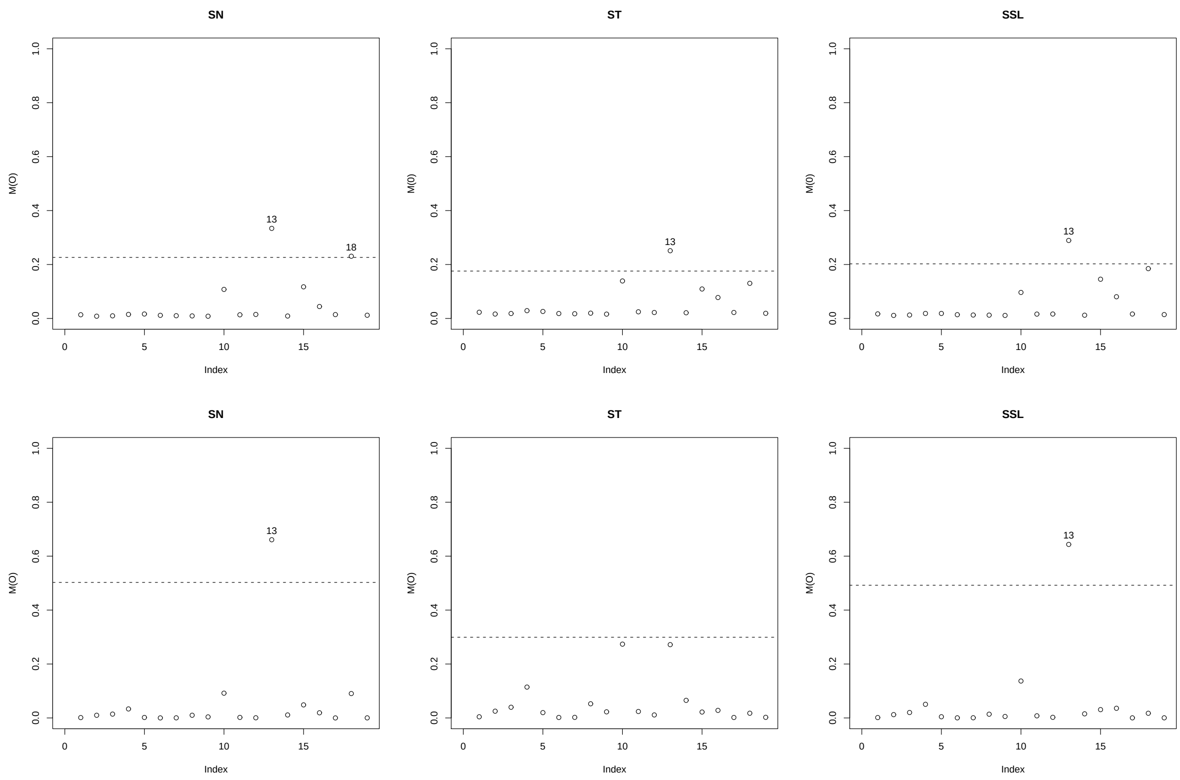

Figure 6: Oil palm data. Index plot of M(0) using SN, ST and SSL models. In the first row case weights perturbation and in the second row scale parameter perturbation.

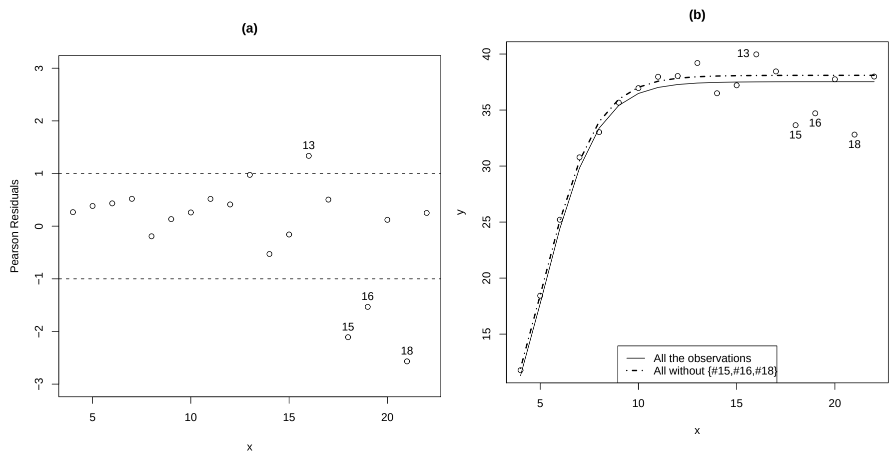

Figure 7: Oil palm yield data. (a) Index plot of residuals versus explanatory variable x_i for the ST-NLM. (b) Predicted values for the ST-NLM.

In order to detect incorrect specification of the error distribution and the systematic component (25), in Figure 8 we show the QQ-plots and simulated envelopes for the Pearson residuals. This Figure clearly indicate that the ST-NLM is more suitable for modeling the current data than the SN-NLM and SSL-NLM, since there are no observations falling outside the envelope. Moreover, there is evidence of lack of fit for the SN-NLM.

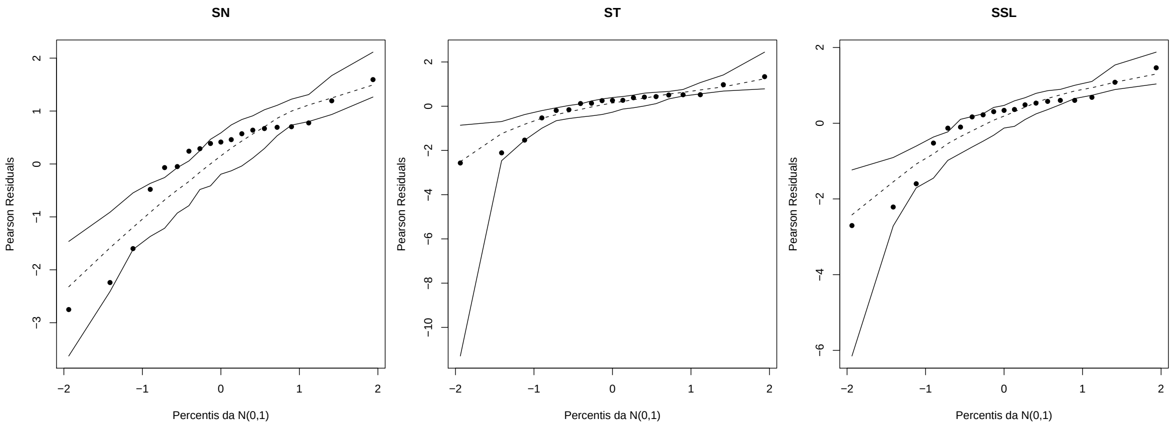

Figure 8: Oil palm yield data. Plots of the Pearson residuals against the order statistics of the normal distribution to the SN-NLM, ST-NLM and SSL-NLM.

5.4. Influence of a single outlier

The robustness of the ST-NLM and SSL-NLM can be also studied through the influence of a single outlying observation on the ML estimate of $\boldsymbol{\beta}$. In particular, we can asses how much the ML estimates of $\boldsymbol{\theta}$ influences by a change of ∇ units in a single observation Y_k. We replace a single observation y_k by $y_k(\nabla) = y_k - \nabla$, and record the relative change in the estimates $\Big((\widehat{\theta}(\nabla) - \widehat{\theta})/\widehat{\theta}\Big)$, where $\widehat{\theta}$ denotes the original estimate and $\widehat{\theta}(\nabla)$ the estimate for the contaminated data. In this example, we contaminated the observation on subject 5, and varied ∇ between 0 and 5 in increments of 0.4. In Figure 9, we have presented the results of relative changes of the estimate β_2 and, β_3 for different contaminations of ∇, under SN-LMM, ST-LMM and SSL-NLM. As expected, the ST and SSL models are less adversely affected by variations of ∇ than the SN model.

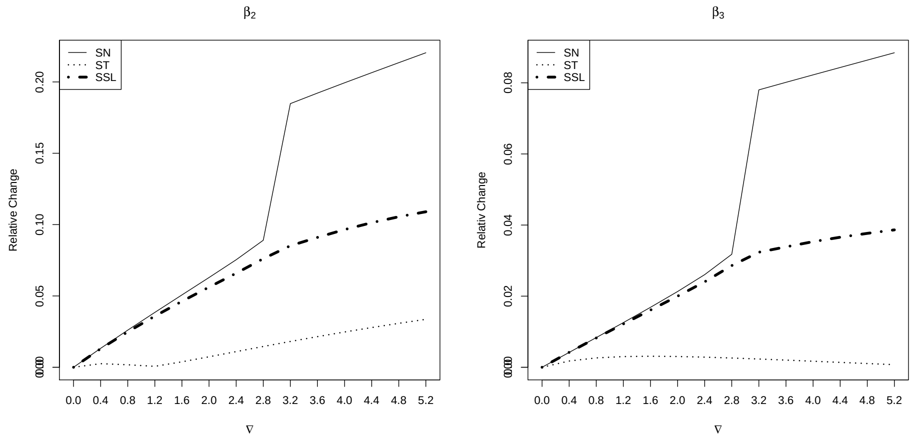

Figure 9: Oil palm yield data set. Relative changes on the ML estimates of β_2 and β_3 when fitting a SN-NLM, ST-NLM and SSL-NLM for different contaminations of ∇ in the fifth observation on subject 5. Relative change= $\Big((\widehat{\theta}(\nabla) - \widehat{\theta})/\widehat{\theta}\Big)$, where $\widehat{\theta}$ denotes the original estimate and $\widehat{\theta}(\nabla)$ the estimate for the contaminated data.

5.5. *Heteroscedasticity diagnostic*

We now consider the test for heteroscedasticity for the oil palm data. In the previous analysis, we saw that the ST-NLM is the most appropriate for these data, so our analysis will be based on this distribution. As suggested by Cook and Weisberg (1982), we take the exponential function as the weight function, that is, we consider $m_i = \exp(\rho x_i)$. It is easily seen that when $\rho = 0$, then $w_i = 1$ and $\sigma_i^2 = \sigma^2$ for all i. Hence, the test for the homogeneity of scalar parameter becomes the test of hypothesis $H_0 : \rho = 0$. Based on the statistic SR given in Section 3 and a little computation, we get $SR_{ST} = 19.79010$ and the corresponding p-value is about 0. Thus, we should reject the hypothesis H_0 and therefore the assumption of homogeneity of variance is not suitable for the oil palm data.

We believe that the proposed score test is very sensitive to the presence of influential observation, so we eliminate now these observations from the full data (15, 16 and 18) and by similar computation we get $SR_{ST} = 3.028285$ (p-value $= 0.0818$), which indicates that when the influential observations are deleted, we cannot reject the hypothesis H_0. This result agrees with the graphical analysis depicted in Figure 7(a), where we can see a clear constant pattern of the residuals on the interval $(-1, 1)$. Table 3 shows the values of the SR statistics for the SN-NLM, SSL-NLM and the ST-NLM. We note that, for the SN-NLM and SSL-NLM always should reject the hypothesis H_0. As expected, the presence of influential observations can affect significantly our decision about the heteroscedasticity, and this decision can be changed depending of the model we are used. Similar conclusions emerged when we chose $m_i(x_i, \rho) = \mathbf{x_i}^{\rho}$.

Finally, in Figure 7(b) we show the predicted values, where the full data and the data without the influential observations are considered for the ST-NLM. We note that, when the influential observations are deleted, a slight modification on the curve can be seen related to consider the full data.

Table 3: Oil palm yield data set. Score statistics and the corresponding p-values for some SMSN-NLM.

	Full data		All - ♯15, ♯16 and ♯18	
Model	SR	*p-value*	SR	*p-value*
SN-NLM	38.87257	4.5239e-10	5.324405	0.02102
ST-NLM	19.79010	8.6429e-06	3.028285	0.08182
SSL-NLM	32.12216	1.4477e-08	3.910459	0.04798

6. Conclusions

In this paper, we have proposed the application of a new class of asymmetric distributions, called SMSN distributions, to nonlinear regression models. An EM-type algorithm is developed by exploring the statistical properties of the SMSN class that can be implemented efficiently in softwares as SAS, R, Ox and Matlab. The observed information matrix is derived analytically which allows direct implementation of inference on this class of models. In order to examine the performance and properties of the score test for heteroscedasticity of the scalar parameter in the framework of SMSN-NLM, some simulation studies are carried out under different situations. These simulation studies indicates that the test is effective for all the models. Furthermore, influence diagnostics analyses are discussed for SMSN-NLM where it is noted that the influence of the observations is reduced when we consider distributions with heavier tails than the SN one. For the Oil Palm data set, the ST model accommodates slightly better the influential observations.

Due to recent advances in computational technology, it is worthwhile to carry out Bayesian treatments via Markov chain Monte Carlo (MCMC) sampling methods in the context of SMSN-NLM. Bayesian influence diagnostics can be treated via the Kullback-Leibler divergence as proposed by Cho et al. (2009). Other extensions of the current work include, for example, a generalization of SMSN-NLM to multivariate settings and nonlinear mixed effects models.

Acknowledgment: This research work was supported in part by grants 2008/02159-9 from FAPESP-Brazil. Grants from CNPq-Brazil are also Acknowledged.

Appendix A: First and second order derivatives to the heteroscedastic model

In this Appendix the first and second order derivatives of $d_i = B_i^2$ and $A_i = \lambda B_i$ are obtained, where $B_i = (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}) - b\sigma_i\delta)/\sigma_i = C_i - b\delta$, with $C_i = (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}))/\sigma_i$ and $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$. Here $\sigma_i^2 = \sigma^2 m(\mathbf{x}_i, \boldsymbol{\rho}) = \sigma^2 m_i$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\rho}^\top, \boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$.

- d_i:

$$
\begin{aligned}
\frac{\partial d_i}{\partial \boldsymbol{\beta}} &= -2\frac{B_i}{\sigma_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial d_i}{\partial \sigma^2} = -\frac{B_i}{\sigma^2}C_i, \quad \frac{\partial d_i}{\partial \lambda} = -2bB_i\delta', \quad \frac{\partial d_i}{\partial \boldsymbol{\rho}} = -\frac{B_i}{m_i}C_i\frac{\partial m_i}{\partial \boldsymbol{\rho}},\\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= 2\left[\frac{1}{\sigma_i^2}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}^\top} + \frac{B_i}{\sigma_i}\frac{\partial^2 \eta_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\right],\\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}\partial \sigma^2} &= \frac{1}{\sigma^2\sigma_i}\left[2B_i + b\delta\right]\frac{\partial \eta_i}{\partial \boldsymbol{\beta}},\\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}\partial \lambda} &= \frac{2b}{\sigma_i}\delta'\frac{\partial \eta_i}{\partial \boldsymbol{\beta}},\\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\rho}^\top} &= \frac{1}{\sigma_i m_i}\left[2B_i + b\delta\right]\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top},\\
\frac{\partial^2 d_i}{\partial \sigma^2\partial \sigma^2} &= \frac{1}{2\sigma^4}(2B_i + \frac{1}{2}b\delta)C_i, \quad \frac{\partial^2 d_i}{\partial \sigma^2\partial \lambda} = \frac{b\delta'}{\sigma^2}C_i, \quad \frac{\partial^2 B_i}{\partial \sigma^2\partial \boldsymbol{\rho}^\top} = \frac{1}{2\sigma^2 m_i}(2B_i + b\delta)C_i\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top},\\
\frac{\partial^2 d_i}{\partial \lambda\partial \lambda} &= -2b[\delta'' B_i - b(\delta')^2], \quad \frac{\partial^2 d_i}{\partial \lambda\partial \boldsymbol{\rho}} = \frac{b\delta'}{m_i}(2B_i + b\delta)\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top},\\
\frac{\partial^2 d_i}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}^\top} &= \left[\frac{1}{2m_i^2}(4B_i + b\delta)\frac{\partial m_i}{\partial \boldsymbol{\rho}}\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top} - \frac{B_i}{m_i}\frac{\partial^2 m_i}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}^\top}\right]C_i,
\end{aligned}
$$

where δ' and δ'' are the first and second order derivatives of δ.

- A_i:

$$
\begin{aligned}
\frac{\partial A_i}{\partial \boldsymbol{\beta}} &= -\frac{\lambda}{\sigma_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial A_i}{\partial \sigma^2} = -\frac{\lambda}{2\sigma^2}C_i, \quad \frac{\partial A_i}{\partial \lambda} = B_i - b\lambda\delta', \quad \frac{\partial A_i}{\partial \boldsymbol{\rho}} = -\frac{\lambda}{2m_i}C_i\frac{\partial m_i}{\partial \boldsymbol{\rho}},\\
\frac{\partial^2 A_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= -\frac{\lambda}{\sigma_i}\frac{\partial^2 \eta_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}, \quad \frac{\partial^2 A_i}{\partial \boldsymbol{\beta}\partial \sigma^2} = \frac{\lambda}{\sigma^2\sigma_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial^2 A_i}{\partial \boldsymbol{\beta}\partial \lambda} = -\frac{1}{\sigma_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial^2 A_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\rho}^\top} = \frac{\lambda}{2m_i\sigma_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top},\\
\frac{\partial^2 A_i}{\partial \sigma^2\partial \sigma^2} &= \frac{3\lambda}{4\sigma^4}C_i, \quad \frac{\partial^2 A_i}{\partial \sigma^2\partial \lambda} = -\frac{1}{2\sigma^2}C_i, \quad \frac{\partial^2 A_i}{\partial \sigma^2\partial \boldsymbol{\rho}^\top} = \frac{\lambda}{4m_i\sigma^2}C_i\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top},\\
\frac{\partial^2 A_i}{\partial \lambda\partial \lambda} &= -b[2\delta' + \lambda\delta''], \quad \frac{\partial^2 A_i}{\partial \lambda\partial \boldsymbol{\rho}^\top} = -\frac{1}{2m_i}C_i\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top},\\
\frac{\partial^2 A_i}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}^\top} &= \frac{\lambda}{2}\left[\frac{3}{2m_i^2}\frac{\partial m_i}{\partial \boldsymbol{\rho}}\frac{\partial m_i}{\partial \boldsymbol{\rho}^\top} - \frac{1}{m_i}\frac{\partial^2 m_i}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}^\top}\right]C_i.
\end{aligned}
$$

Appendix B: Elements of $\boldsymbol{\Delta}^*$

In order to asses the perturbation scheme $\sigma^2(\omega_i) = \sigma^2/\omega_i$, we go to present the elements of $\boldsymbol{\Delta}^*$ for this the perturbation scheme. In this case the perturbed log-likelihood function given by $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\omega_i)$, where

$$
\ell_i(\boldsymbol{\theta}|\omega_i) = \log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log\omega_i + \log K_{\omega_i},
$$

where $K_{\omega i}$ is as in Section 2.2 with $\sigma^2(\omega_i)$ instead of σ^2 and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$. In this case $d_{\omega_i} = B^2(\omega_i)$ and $A_{\omega i} = \lambda B(\omega_i)$, where $B(w_i) = \omega_i^{1/2}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i))/\sigma - b\delta\sigma$. So, under this perturbation scheme we have $\boldsymbol{\Delta}^*_{\boldsymbol{\omega}_0} = (\boldsymbol{\Delta}^{*\top}_{\beta}, \boldsymbol{\Delta}^{*\top}_{\sigma^2}, \boldsymbol{\Delta}^{*\top}_{\lambda})^\top$, where

$$
\boldsymbol{\Delta}^*_{i\boldsymbol{\gamma}} = -\frac{1}{K^2_{\omega_i}}\frac{\partial}{\partial \omega_i}K_{\omega i}\frac{\partial}{\partial \boldsymbol{\gamma}}K_{\omega i} + \frac{1}{K_{\omega_i}}\frac{\partial^2}{\partial \omega_i\partial \boldsymbol{\gamma}}K_{\omega_i}, \ \boldsymbol{\gamma} = \boldsymbol{\beta}, \sigma^2, \lambda,
$$

with

$$
\begin{aligned}
\frac{\partial}{\partial \omega_i}K_{\omega_i} &= -\frac{1}{2}I_i^\Phi(3/2)\frac{\partial}{\partial \omega_i}d_{\omega_i} + I_i^\phi(1)\frac{\partial}{\partial \omega_i}A_{\omega_i},\\
\frac{\partial}{\partial \boldsymbol{\gamma}}K_{\omega_i} &= -\frac{1}{2}I_i^\Phi(3/2)\frac{\partial}{\partial \boldsymbol{\gamma}}d_{\omega_i} + I_i^\phi(1)\frac{\partial}{\partial \boldsymbol{\gamma}}A_{\omega_i},\\
\frac{\partial^2}{\partial \omega_i\partial \boldsymbol{\gamma}}K_{\omega_i} &= -\frac{1}{2}\left[-\frac{1}{2}I_i^\Phi(5/2)\frac{\partial}{\partial \boldsymbol{\gamma}}d_{\omega_i} + I_i^\phi(2)\frac{\partial}{\partial \omega_i}A_{\omega_i}\right]\frac{\partial}{\partial \boldsymbol{\gamma}}d_{\omega_i} - \frac{1}{2}I_i^\Phi(3/2)\frac{\partial^2}{\partial \omega_i\partial \boldsymbol{\gamma}}d_{\omega_i}\\
&\quad -\frac{1}{2}I_i^\phi(2)\left[\frac{\partial}{\partial \boldsymbol{\gamma}}d_{\omega_i} + +2A_{\omega_i}\frac{\partial}{\partial \omega_i}A_{\omega_i}\right]\frac{\partial}{\partial \boldsymbol{\gamma}}A_{\omega_i} + I_i^\phi(1)\frac{\partial^2}{\partial \omega_i\partial \boldsymbol{\gamma}}A_{\omega_i}.
\end{aligned}
$$

Next we present the derivatives of d_{ω_i} and A_{ω_i} with respect to $\boldsymbol{\theta}$ and ω. To simplify the notation let $\eta_i = \eta(\boldsymbol{\beta}, \mathbf{x}_i)$. So evaluating at $\omega_i = \omega_{i0} = 1$ we obtain

- d_i:

$$
\begin{aligned}
\frac{\partial d_{\omega i}}{\partial \omega i} &= C_i B_i,\\
\frac{\partial d_{\omega i}}{\partial \boldsymbol{\beta}} &= -\frac{2}{\sigma}B_i\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial d_{\omega i}}{\partial \sigma^2} = -\frac{1}{\sigma^2}B_iC_i, \quad \frac{\partial d_{\omega i}}{\partial \lambda} = -2b\delta' B_i,\\
\frac{\partial^2 d_{wi}}{\partial \boldsymbol{\beta}\partial \omega_i} &= -\frac{1}{\sigma}\left(2B_i + b\delta\right)\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial^2 d_{\omega i}}{\partial \omega\partial \sigma^2} = -\frac{1}{2\sigma^2}\left(2B_i + b\delta\right), \quad \frac{\partial^2 d_{\omega i}}{\partial \lambda\partial \omega_i} = -b\delta' C_i,
\end{aligned}
$$

- A_i:

$$
\begin{aligned}
\frac{\partial A_{\omega i}}{\partial \omega_i} &= \frac{\lambda}{2}C_i,\\
\frac{\partial A_{\omega i}}{\partial \boldsymbol{\beta}} &= -\frac{\lambda}{\sigma}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial A_{\omega i}}{\partial \sigma^2} = -\frac{\lambda}{2\sigma^2}C_i, \quad \frac{\partial A_{\omega i}}{\partial \lambda} = B_i - b\lambda\delta',\\
\frac{\partial^2 A_{\omega i}}{\partial \boldsymbol{\beta}\partial \omega_i} &= -\frac{\lambda}{2\sigma}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial^2 A_{\omega i}}{\partial \sigma^2\partial \omega_i} = -\frac{\lambda}{4\sigma^2}C_i, \quad \frac{\partial^2 A_{wi}}{\partial \lambda\partial \omega_i} = \frac{1}{2}C_i.
\end{aligned}
$$

References

Atkinson, A., 1981. Two graphical displays for outlying and influential observations in regression. Biometrika 68, 13.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew-normal distribution. Journal of the Royal Statistical Society 61, 579–602.

Azzalini, A., Genton, M., 2008. Robust likelihood methods based on the skew-t and related distributions. International Statistical Review 76, 1490–1507.

Basso, R. . M., Lachos, V. H., Cabral, C. R., Ghosh, P., 2009. Robust mixture modeling based on scale mixtures of skew-normal distributions. Computational Statistics & Data Analysis (In press).

Berkane, M., Kano, Y., Bentler, P. M., 1994. Pseudo maximum likelihood estimation in elliptical theory: effects of misspecification. Computational Statistics & Data Analysis 18, 255–267.

Branco, M. D., Dey, D. K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis 79, 99–113.

Cancho, V. C., Lachos, V. H., Ortega, E. M. M., 2009. A nonlinear regression model with skew-normal errors. Statistical Papers doi:10.1007/s00362-008-0139-y.

Cho, H., Ibrahim, J. G., Sinha, D., Zhu, H., 2009. Bayesian case influence diagnostics for survival models. Biometrics 65, 116–124.

Cook, R., 1977. Detection of influential observation in linear regression. Technometrics 19 (1), 15–18.

Cook, R., Weisberg, S., 1983. Diagnostics for heteroscedasticity in regression. Biometrika, 1–10.

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

Cook, R. D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman & Hall/CRC, Boca Raton, FL.

Cox, D., Hinkley, D., 1974. Theoretical Statistics. Chapman & Hall/CRC, New York, NY.

Cysneiros, F., Paula, G., Galea, M., 2007. Heteroscedastic symmetrical linear models. Statistics and Probability Letters 77 (11), 1084–1090.

Cysneiros, F. J. A., Vanegas, L. H., 2008. Residuals and their statistical properties in symmetrical nonlinear models. Statistics & Probability Letters 78, 3269–3273.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39, 1–38.

Galea, M., Paula, G. A., Cysneiros, F. J. A., 2005. On diagnostics in symmetrical nonlinear models. Statistics & Probability Letters, 459–467.

Gómez, H., Venegas, O., Bolfarine, H., 2007. Skew-symmetric distributions generated by the distribution function of the normal distribution. Environmetrics 18 (4), 395–408.

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

Lange, K. L., Sinsheimer, J. S., 1993. Normal/independent distributions and their applications in robust regression. J. Comput. Graph. Stat 2, 175–198.

Lin, J., Wei, B., 2003. Testing for heteroscedasticity in nonlinear regression models. Communications in statistics–Theory and methods 32 (1), 171–192.

Lin, J., Xie, F., Wei, B., 2009. Statistical Diagnostics for Skew-t-Normal Nonlinear Models. Communications in Statistics-Simulation and Computation 38 (10), 2096–2110.

Lin, T. I., Lee, J. C., Hsieh, W. J., 2007. Robust mixture modelling using the skew t distribution. Statistics and Computing 17, 81–92.

Meng, X., Rubin, D. B., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika 81, 633–648.

Montenegro, L. C., Lachos, V., Bolfarine, H., 2009. Local influence analysis of skew-normal linear mixed models. Communication in Statistics–Theory and Methods 38, 484–496.

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

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

Vanegas, L. H., Cysneiros, F. J. A., 2010. Assesment of diagnostic procedures in symmetrical nonlinear regression models. Computational Statistics & Data Analysis 54, 1002–1016.

Wang, J., Genton, M. G., 2006. The multivariate skew-slash distribution. Journal of Statistical Planning and Inference 136, 209–220.

Xie, F., Lin, J., Wei, B., 2009a. Diagnostics for skew-normal nonlinear regression models with AR (1) errors. Computational Statistics and Data Analysis 53 (12), 4403–4416.

Xie, F. C., Wei, B. C., Lin, J. G., 2009b. Homogeneity diagnostics for skew-normal nonlinear regression models. Statistics and Probability Letters 79, 821–827.