Estimation and diagnostics in heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions

Victor H. Lachos, Aldo M. Garay, Filidor V. Labra

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

An extension of some standard likelihood based procedures to heteroscedastic nonlinear regression models under scale mixtures of skew-normal (SMSN) distributions is developed. This novel class of models provides a useful generalization of the heteroscedastic symmetrical nonlinear regression models (Cysneiros et al., 2009) 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. We derive a simple EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters and the observed information matrix is derived analytically. In order to examine the performance of the proposed methods, some simulation studies are presented to show the robust aspect of this flexible class against outlying and influential observations and that the maximum likelihood estimates based on the EM-type algorithm do provide good asymptotic properties. Furthermore, local influence measures and the one-step approximations of the estimates in the case-deletion model are obtained. Finally, an illustration of the methodology is given considering a data set previously analyzed under the homoscedastic skew-t nonlinear regression model.

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

1. Introduction

Normal nonlinear regression models (N-NLM) are usually applied in sciences and engineering to model symmetrical data for which nonlinear functions of unknown parameters are used in order to explaining or describing the phenomena under study. However, it is well-known that several phenomena are not always in agreement with the normal model due to lack of symmetry in the distribution or the presence of heavy-and-lightly tailed distributions related the normal law in the data. Particularly, it is well-known

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

that the parameter estimates of the normal model based on maximum likelihood (ML) methods are often sensitive to atypical observations. To deal with this problem, some proposals have been made in the literature by replacing the normality assumption 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 model (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). Although these models are attractive, there is a need to check the distributional assumptions of the model errors because these can present simultaneously skewness and heavy-tailed. To overcome the problem of atypical data in an asymmetrical context, Branco and Dey (2001) proposed the class of scale mixtures of skew-normal (SMSN) distributions. This rich class of distributions contains the entire family of scale mixtures of normal distributions (Lange and Sinsheimer, 1993), and skewed versions of classical symmetric distributions such as the skew–t (ST), the skew–slash (SSL) and the skew contaminated normal (SCN) distribution.

In regression analysis, it is a standard assumption that the error terms all have equal variances and the violation of this assumption can have adverse consequences for the efficiency of estimators (Cysneiros et al., 2009). The problem of modelling variances has been largely discussed in the statistical literature. For instance, Barroso and Cordeiro (2005) considered Bartlett corrections in heteroscedastic regression models with Student-t distributions; Cysneiros et al. (2007) discussed diagnostic methods in heteroscedastic linear models with symmetrical errors and Cysneiros et al. (2009) introduced the class of heteroscedastic symmetrical nonlinear regression models where a joint iterative process for estimating the mean and dispersion parameters is proposed. In this article, we extend the heteroscedastic symmetric nonlinear regression model (Cysneiros et al., 2009), where both mean and dispersion parameters vary across observations through nonlinear regression structures, by assuming that the model errors follow a mean-zero SMSN distribution, so that the SMSN-HNLM 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. In our approach, the ML estimates of the location and scale parameters are less sensitive under the presence of outliers than the skew-normal counterpart and, moreover, it is shown through a simulation study that the ML estimates do provide good asymptotic properties.

On the other hand, 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. (2009a). 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 heteroscedastic NLMs is a significant contribution to this field.

The rest of the paper is organized as follows. The heteroscedastic asymmetric nonlinear model (SMSN-HNLM) is presented in Section 2, including a brief introduction to univariate SMSN distributions and likelihood-based methodology for inference. In Section 3 an EM-type algorithm for ML estimation is developed. In Section 4 we present the results of a simulation study in which the characteristics of our proposed model are evaluated in the presence of outliers. Further we show that the maximum likelihood estimates based on the EM-type algorithm do provide reasonable accuracy. In Section 5, we derive global influence measures for SMSN-HNLM and we study the local influence approach under perturbation of case weight. The methodology proposed is illustrated in Section 6 by analyzing a real data set and some conclusive remarks are presented in Section 7.

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

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

where μ is a location parameter, Z is skew-normal random variable with location 0, scale σ^2, skewness λ. That is $Z \sim SN(0, \sigma^2, \lambda)$, U is a positive random variable with cumulative distribution function (cdf) $H(\cdot; \boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of U. Given $U = u$, we have that $Y|U = u \sim SN(\mu, u^{-1}\sigma^2, \lambda)$. Thus, the probability density function (pdf) 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 $N(\mu, \sigma^2)$ and $\Phi(\cdot)$ is the cumulative distribution function(cdf) 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 $f_0(y) = \int_0^\infty \phi(y; \boldsymbol{\mu}, u^{-1}\sigma^2)dH(u; \boldsymbol{\nu})$.

For a random variable Y com pdf (2), a convenient stochastic representation is given by

$$Y = \mu + \Delta T + U^{-1/2}\Gamma^{1/2}T_1, \tag{3}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $\Delta = \sigma\delta$, $\Gamma = (1-\delta^2)\sigma^2$, $T = U^{-1/2}|T_0|$, T_0 and T_1 are independent standard normal random variables and $|\cdot|$ denotes absolute value. Hence,

$$E[Y] = \mu + \sqrt{\frac{2}{\pi}}k_1\Delta \quad \text{and} \quad Var[Y] = \sigma^2(k_2 - b^2\delta^2), \tag{4}$$

where $b = -\sqrt{\frac{2}{\pi}}k_1$ and $k_m = E[U^{-m/2}]$.

We refer to Basso et al. (2009) for details and additional properties related to this class of distributions. Some of these distributions are described subsequently.

- *The skew–t distribution.* 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

$$f(y|\mu, \sigma^2, \lambda; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1 + \frac{d}{\nu}\right)^{-\frac{\nu+1}{2}} T\left(\sqrt{\frac{\nu+1}{d+\nu}}A; \nu+1\right), \quad y \in \mathbb{R}, \tag{5}$$

  where $d = (y-\mu)^2/\sigma^2$, $A = \lambda(y-\mu)/\sigma$ and $T(\cdot; \nu)$ denotes the cdf of the standard Student–t distribution with ν degrees of freedom, namely $t(0, 1, \nu)$. We use the notation $Y \sim ST(\mu, \sigma^2, \lambda; \nu)$.

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

$$f(y|\mu, \sigma^2, \lambda; \nu) = 2\nu \int_0^1 u^{\nu-1}\phi(y; \mu, u^{-1}\sigma^2)\Phi(u^{1/2}A)du, \quad y \in \mathbb{R}. \tag{6}$$

- *The skew-contaminated normal distribution.* Denoted by $Y \sim SCN(\mu, \sigma^2, \lambda; \nu, \gamma)$. Here U is a discrete random variable taking one of two states. In this case the pdf of U is given by

$$h(u|\boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \quad \nu, \gamma \in (0, 1),$$

  where $\boldsymbol{\nu} = (\nu, \gamma)^\top$. It follows immediately that

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

3. The SMSN heteroscedastic nonlinear regression model

We propose a heteroscedastic nonlinear regression model based on SMSN distributions (SMSN-HNLM), which is defined by

$$Y_i = \eta(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{7}$$

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 $\varepsilon_i \sim SMSN(-\sqrt{\frac{2}{\pi}}k_1\Delta m_i^{1/2},\ m_i\sigma^2,\ \lambda; H)$, where $m_i = m(\boldsymbol{\rho}, \mathbf{z}_i)$ is a known positive continuously differentiable function and $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq})^\top$ containing values of explanatory variables and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)^\top$ a vector of unknown parameters. If the variances depend on the values of some explanatory variables $\mathbf{z}_i$, one specific form of m is the log-linear model given by $m_i(\boldsymbol{\rho}, \mathbf{z}_i) = \exp\left(\sum_{j=1}^p \rho_j z_{ij}\right)$.

Following the properties from SMNS distribution Basso et al. (2009), from (7) we have that marginally $Y_i \sim SMSN(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + b_1\Delta w_i^{1/2}, \sigma^2 w_i, \lambda; H)$, for $i = 1, \ldots, n$. Thus the SMSN-HNLM corresponds to the regression model where the error distribution has mean zero and hence the regression parameters are all comparable with

$$E[Y_i] = \eta(\boldsymbol{\beta}, \mathbf{x}_i) \quad \text{and} \quad Var[Y_i] = \sigma^2 m_i(k_2 - b^2\delta^2).$$

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 from now on will take the value of $\boldsymbol{\nu}$ to be known. Thus, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\rho}^\top, \boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$ 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

$$\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log m_i + \log K_i, \tag{8}$$

with $K_i = \int_0^\infty u_i^{1/2}\exp\{-\frac{1}{2}u_i d_i\}\Phi(u_i^{1/2}A_i)dH(u_i)$, $d_i = (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta m_i^{1/2})^2/(m_i\sigma^2)$ is the Mahalanobish distance, and $A_i = \lambda(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta m_i^{1/2})/(m_i\sigma^2)$.

3.1. The observed information matrix

Suppose that we have observations on n independent individuals, $Y_1, ..., Y_n$, we define $I_i^\Phi(w) = E_H[U^w e^{-Ud_i/2}\Phi_1(U^{1/2}A_i)]$ and $I_i^\phi(w) = \frac{1}{\sqrt{2\pi}}E_H[U^w e^{-U(d_i+A_i^2)/2}]$, $i = 1, .., n$. So the score function and the observed information matrix are respectively given by $U(\boldsymbol{\theta}) = \sum_{i=1}^n U_i(\boldsymbol{\theta})$ where

$$U_i(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{\partial \text{log}\sigma^2}{\partial \boldsymbol{\theta}} - \frac{1}{2}\frac{\partial \text{log}m_i}{\partial \boldsymbol{\theta}} + \frac{1}{K_i}\frac{\partial K_i}{\partial \boldsymbol{\theta}}, \tag{9}$$

with $\frac{\partial K_i}{\partial \boldsymbol{\theta}} = I_i^\phi(1)\frac{\partial A_i}{\partial \boldsymbol{\theta}} - \frac{1}{2}I_i^\Phi(3/2)\frac{\partial d_i}{\partial \boldsymbol{\theta}}$, and the observed information matrix is given by

$$\mathbf{J}(\boldsymbol{\theta}) \quad = \quad -\sum_{i=1}^{n}\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, \tag{10}$$

where

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

with

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

The derivatives of d_i and A_i involve standard algebraic manipulations which are given in the Appendix.

Note that 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. 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)

3.2. Parameter estimation via the EM-algorithm

In this section we develop an EM-type algorithm for maximum likelihood estimation of the parameters of the SMSN-HNLM. In order to do this, we first represent the SMSN-HNLM in an incomplete data framework using the stochastic representation given in (3). Thus, we consider the following hierarchical representation for Y_i

$$
\begin{aligned}
Y_i|T_i = t_i &\sim N_1(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + \Delta m_i^{1/2} t_i, U_i^{-1} m_i \Gamma), && (11)\\
T_i|U_i &\sim TN_1(b, u_i^{-1}; (b, \infty)), && (12)\\
U_i &\sim H(.; \boldsymbol{\nu}) && (13)
\end{aligned}
$$

where $TN_1(r, s; (a, b))$ denotes the univariate normal distribution $(N(r, s))$, truncated on the interval (a, b). An useful straightforward result is that the conditional distribution of T_i given y_i and u_i is $TN_1(\mu_{Ti} + b, u_i^{-1} M_T^2; (b, \infty))$, with $M_T^2 = \dfrac{\Gamma}{\Delta^2 + \Gamma}$, $\mu_{T_i} = \dfrac{\Delta}{m_i^{1/2}(\Delta^2 + \Gamma)}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \Delta b m_i^{1/2})$.

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 $\boldsymbol{\theta} = (\boldsymbol{\rho}^\top, \boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$. 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

$$
\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{t}, \mathbf{u}) = c - \frac{n}{2}\log\Gamma - \frac{1}{2}\sum_{i=1}^{n}\log m_i - \frac{1}{2\Gamma}\sum_{i=1}^{n}\frac{u_i}{m_i}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \Delta m_i^{1/2} t_i)^2, \qquad (14)
$$

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

$$
\widehat{ut}_i = \widehat{u}_i(\widehat{\mu}_{Ti} + b) + \widehat{M}_T \widehat{\tau}_{1_i}, \quad \widehat{ut_i^2} = \widehat{u}_i(\widehat{\mu}_{T_i} + b)^2 + \widehat{M_T^2} + \widehat{M}_T(\widehat{\mu}_{T_i} + 2b)\widehat{\tau}_{1_i}, \qquad (15)
$$

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

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 of Subsection 2.1 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

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

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). The *ECME algorithm* (Liu and Rubin, 1994), a faster extension of EM and ECM, is obtained by maximizing the constrained Q-function with some CM steps that maximize the corresponding constrained actual marginal likelihood function, called *CML steps*. Next, we describe this EM-type algorithm (ECME) for maximum likelihood estimation of the parameters of the SMSN distributions.

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

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \mathrm{argmin}_{\boldsymbol{\beta}}(\mathbf{z}^{(k)}-\boldsymbol{\eta}(\boldsymbol{\beta},\mathbf{x}))^\top\widehat{\mathbf{U}}^{(\mathbf{k})}(\mathbf{z}^{(k)}-\boldsymbol{\eta}(\boldsymbol{\beta},\mathbf{x})), \tag{16}$$

$$\widehat{\Delta}^{(k+1)} = \frac{\sum_{i=1}^{n}\widehat{ut}_i^{*(k)}(y_i-\eta(\boldsymbol{\beta}^{(k+1)},\mathbf{x}_i))}{\sum_{i=1}^{n}\widehat{ut_i^2}^{(k)}}, \tag{17}$$

$$
\begin{aligned}
\widehat{\Gamma}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\Big((y_i-\eta(\boldsymbol{\beta}^{(k+1)},\mathbf{x}_i))^2\widehat{u}_i^{*(k)} - 2\Delta^{(k+1)}(y_i-\eta(\boldsymbol{\beta}^{(k+1)},\mathbf{x}_i))\widehat{ut}_i^{*(k)}\\
&+ (\Delta^2)^{(k+1)}\widehat{ut^2}_i^{(k)}\Big)
\end{aligned} \tag{18}
$$

$$\widehat{\boldsymbol{\rho}}^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\rho}}\sum_{i=1}^{n}\log(f(y_i|\boldsymbol{\beta}^{(k+1)},\sigma^{2(k+1)},\lambda^{(k+1)},\boldsymbol{\nu})). \tag{19}$$

where $\widehat{ut}_i^{*(k)} = \widehat{u}_i^{(k)}/\widehat{m}_i^{m(k)}$, $\widehat{ut}_i^{*(k)} = \widehat{ut}_i^{(k)}/\widehat{m}_i^{1/2(k)}$, $\widehat{\mathbf{U}}^{(k)} = diag(\widehat{u}_1^{*(k)},\ldots,\widehat{u}_n^{*(k)})$, $\mathbf{z}^{(k)}$ 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. Note that $\widehat{\sigma}^{2(k+1)}$ and $\widehat{\lambda}^{(k+1)}$ can be recovered using that $\lambda = \Delta/\sqrt{\Gamma}$ and $\sigma^2 = \Delta^2+\Gamma$.

4. Simulation study

In order to study the performance of our proposed model and algorithm, we present some simulation studies. The first part of this simulation study to show the necessity of heavy-tails asymmetric models to

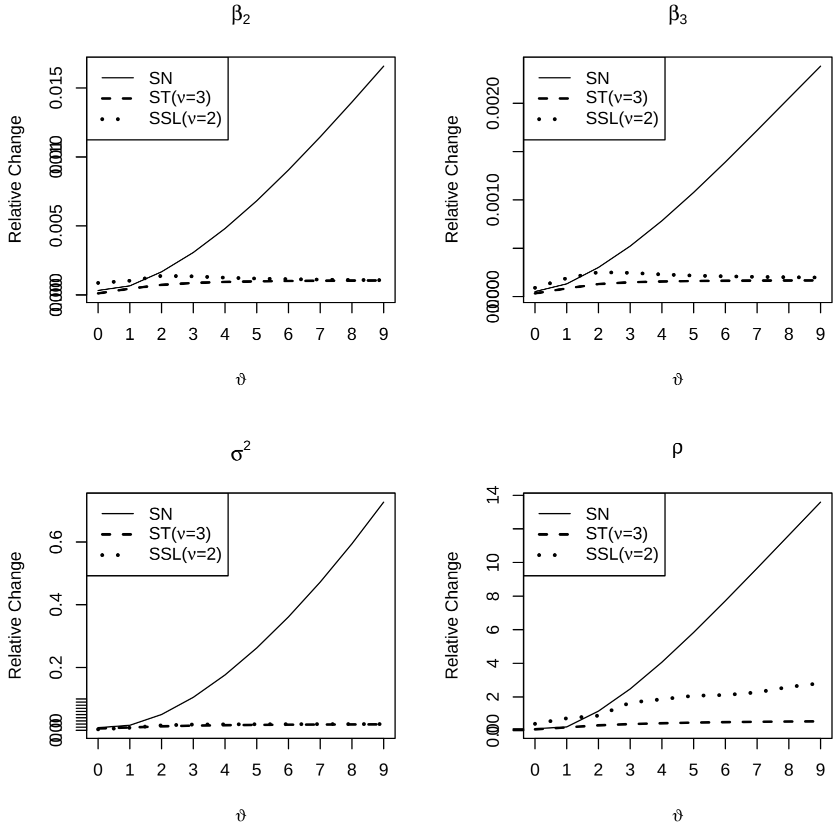

Figure 1: Simulated data. Average changes on estimates of β_2, β_3, σ^2 and ρ for the nonlinear growth-curve model (20).

deal with presence of outliers in the data. The goal of the second part is to show that the ML estimates based on the ECME algorithm do provides good asymptotic properties.

We performed a Monte Carlo simulation study with the following nonlinear growth-curve model:

$$Y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{20}$$

where $\varepsilon_i \sim SMSN(-\sqrt{\frac{2}{\pi}} k_1 \Delta m_i^{1/2}, \sigma_i^2, \lambda; H)$, with $\sigma_i^2 = \sigma^2 \exp\left(\sum_{j=1}^p \rho_j z_{ij}\right)$, for $j = 1$, are independent identically with mean zero. The variable x_i ranging from 0.1 to 20 and these values were held fixed throughout the simulations and $z_i = x_i/10$. The parameter values were set around the estimates obtained in the application $\beta_1 = 37, \beta_2 = 43, \beta_3 = 0.6, \sigma^2 = 0.5, \lambda = -3$ and $\rho = -0.2$.

4.1. Robustness of estimates

The goal of this simulation study is to compare the performance of the ML estimates in the presence of outliers. In order to do this, we generated 1,000 data set of size $n = 50$, considering $\epsilon \sim SN(-\sqrt{2/\pi}\Delta m_i, \sigma_i^2, \lambda)$ in (20). To guarantee the presence of outliers we constructed $Y_i^* = Y_i - \vartheta$, where i is a the corresponding central value of sample and $\vartheta = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. In each replication, we obtained the parameters estimates with and without outliers denoted by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{(i)}$, respectively, by fitting both datas under skew-normal (SN-HNLM), skew-t (ST-HNLM) with different values of $\nu = 3$, 6,

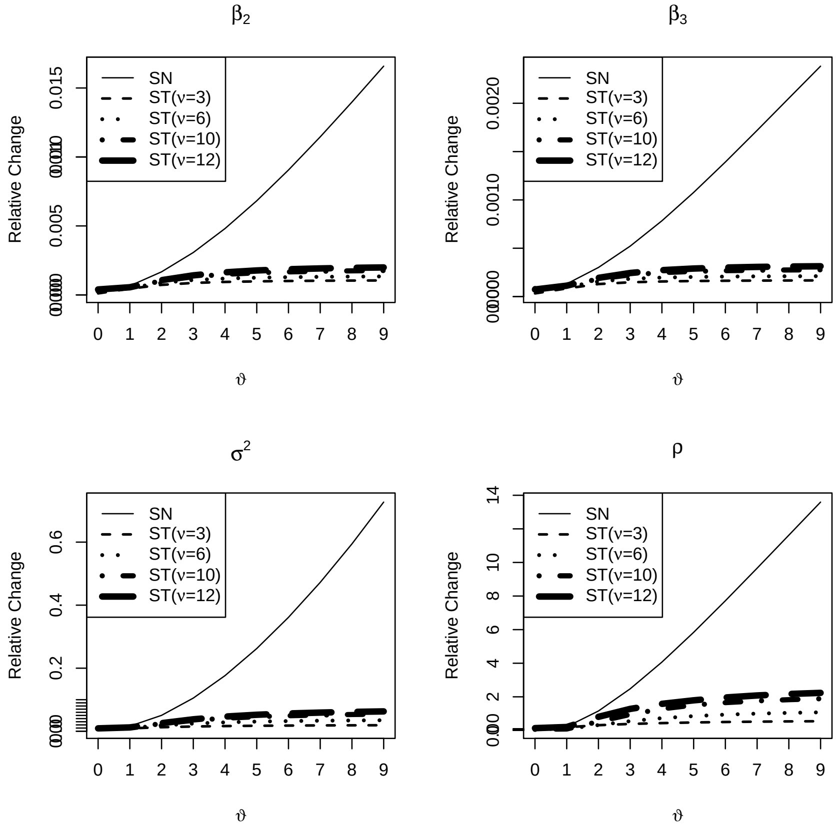

Figure 2: Simulated data. Average changes on estimates of β_2, β_3, σ^2 and ρ for the nonlinear growth-curve model (20).

10 and 12 and skew-slash (SSL-HNLM) with $\nu = 2$. Then we computed the relative changes $\left|\frac{\hat{\gamma}_{(i)}}{\hat{\gamma}} - 1\right|$, where $\boldsymbol{\gamma} = \beta_1,\, \beta_2,\, \beta_3,\, \sigma^2$ or ρ.

Figures 1 and 2, show the average values of relative changes on the estimation for the $1,000$ replications. Notice from these figures that in the SN model the influence increase when ϑ increases. However, for the SMSN models with heavy tails these measure varied little, which indicates that these models are more robust than the SN-HNLM in the presence of discrepant observations. In the ST models the influence increase when ν also increases. Note also that, unlike the other models considered, in the ST-HNLM and SSL-HNLM the influence of outliers on $\widehat{\boldsymbol{\beta}}$, $\widehat{\rho}$ and $\widehat{\sigma}^2$ is not strictly increasing when ϑ increases. These figures shows that, although exists an extreme observation in all models considered, extreme observations seem to be more influent in the skew-normal model, which reflects the ability of models with heavier tails than skew-normal to reduce and control the influence of outliers on parameter estimates. In conclusion, heavy-tailed in a asymmetric context, are less sensitive in the presence of outliers. These results agree with similar considerations, presented in Vanegas and Cysneiros (2010), in a symmetric context.

Table 1: Simulated data. Bias in the estimates of the parameters with different sizes of samples considering the SN, ST and the SSL distribution in the nonlinear growth-curve model (20).

SN-HNLM						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	0.018142	0.170583	2.13575e-05	-0.084223	-0.118008	0.796860
100	0.006792	0.100482	3.91535e-05	-0.043359	0.029257	-0.811886
200	0.002079	0.064363	1.92594e-04	-0.021745	0.013798	-0.358806
300	0.002328	0.050321	1.20312e-04	-0.007175	0.003215	-0.192273
500	0.001778	-0.033103	-9.26027e-05	0.007755	-0.006792	-0.092239
ST-HNLM						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	0.067886	0.069632	-7.45632e-04	-0.147138	0.008791	-0.074835
100	0.012163	0.200567	2.69659e-04	-0.056125	0.035765	-0.748369
200	0.007974	0.017946	-1.40273e-04	-0.014836	0.011951	-0.359338
300	0.000959	-0.045411	-2.45914e-04	-0.000850	0.005841	-0.231270
500	-0.001173	0.020128	9.32806e-05	0.003383	0.000531	-0.087790
SSL-HNLM						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	0.024367	0.373273	2.46772e-04	-0.128733	0.006428	-0.824971
100	0.002395	0.255741	6.62715e-04	-0.045246	0.025884	-0.716985
200	0.004480	0.026543	-1.24525e-04	-0.023700	0.017441	-0.348624
300	-0.000788	0.057916	1.28792e-04	-0.010133	0.011916	-0.208165
500	-0.000653	0.030382	3.68057e-05	-0.005529	0.002849	-0.062118

4.2. Asymptotic Properties

The main focus of this simulation study is the evaluation of bias and mean square error of the ECME estimates. Here the size samples were fixed as $n = 30, 50, 100, 150, 200, 250, 300, 400, 500$. For each combination of parameters and sample size, 500 samples from the MSNS-HNLM in (20) were generated under 3 different situations, i.e. under skew normal (SN-HNLM), skew-t (ST-HNLM) with $\nu = 3$, and under the skew-slash (SSL-HNLM) with $\nu = 2$. The values of ν are chosen in order to yield a highly skewed and heavy tails distribution in the random effects. Then using our proposed ECME algorithm, we compute the *Bias* and mean squared error (*MSE*) for each parameter over the 500 samples and the 3 different situations. They are defined as:

$$Bias(\gamma) = \frac{1}{500}\sum_{i=1}^{500}\left(\widehat{\gamma}^{(i)} - \gamma\right) \quad \text{and} \quad MSE(\gamma) = \frac{1}{500}\sum_{i=1}^{500}\left(\widehat{\gamma}^{(i)} - \gamma\right)^2,$$

where $\boldsymbol{\gamma} = \beta_1,\ \beta_2,\ \beta_3,\ \sigma^2$ or ρ and $\widehat{\gamma}^{(i)}$ is the ECME estimate of $\boldsymbol{\gamma}$ when the data is sample i. The result are show in Table 1 and Figure 3. We can see a patterns of convergence to zero of the Bias and MSE when n increases. The worst case scenario seems to happen while estimating the scale and skewness parameters of the random effect, perhaps due to the well known inferential problems related to the skewness parameter in skew-normal models, or maybe a sample size greater than 500 is needed to obtain a reasonably pattern of convergence. As a general rule, we can say that bias and MSE tend to approach to zero when the sample size is increasing, indicating that the approximates MLEs based on the proposed EM-type algorithm do provide good asymptotic properties.

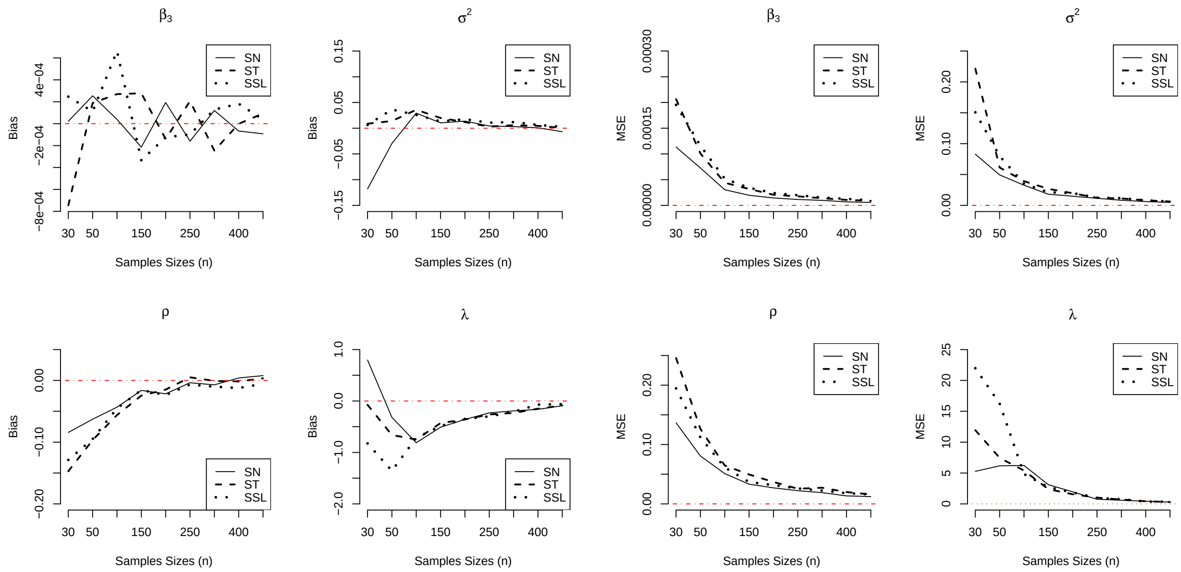

Figure 3: Simulated data. Bias and MSE in the ECME estimates of β_3, σ^2, ρ and λ for the nonlinear growth-curve model (20).

5. 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.

5.1. 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 $\boldsymbol{\theta} = (\boldsymbol{\rho}^\top, \boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$, 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 $\widehat{\boldsymbol{\theta}}_{(i)} = (\boldsymbol{\rho}_{(i)}^\top, \boldsymbol{\beta}_{(i)}^\top, \sigma_{(i)}^2, \lambda_{(i)})^\top$ 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.

$$\widehat{\boldsymbol{\theta}}_{(i)} = \widehat{\boldsymbol{\theta}} + \{\mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1}\dot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}}), \tag{21}$$

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.,

$$GD_i = (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top \mathbf{M}(\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}) \tag{22}$$

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 (21) into Equation (22), 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

$$LD_i(\boldsymbol{\theta}) = 2\{\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_{(i)})\}, \tag{23}$$

Substituting (21) into (23), we obtain the following approximation:

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

5.2. 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 $\boldsymbol{\Delta}^*_{rs} = \partial^2\ell(\boldsymbol{\theta})/\partial\theta_r\partial\psi_s$, for $r = 1, \ldots, (p+2)$ and $r = 1, \ldots, q$, both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\psi} = \boldsymbol{\psi}_o$ (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 case weight perturbation for the SMSN-HNLM given in (7).

• 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

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

where K_i is defined in equation (8) and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ 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 $\boldsymbol{\Delta}_i^* = U_i(\boldsymbol{\theta})$ is as given in (9).

5.3. 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 $\widehat{\mu}_i = \eta(\widehat{\boldsymbol{\beta}}, \mathbf{x}_i)$ and $\widehat{Var}(y_i) = \widehat{\sigma}^2 \widehat{m}_i (k_2 - \frac{2}{\pi} k_1^2 \widehat{\delta}^2)$, with $\widehat{m}_i = m(\widehat{\boldsymbol{\rho}}, \mathbf{z}_i)$. In this case, $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$ and $\widehat{\delta}$ denote the maximum likelihood estimators of $\boldsymbol{\rho}$, $\boldsymbol{\beta}$, 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.

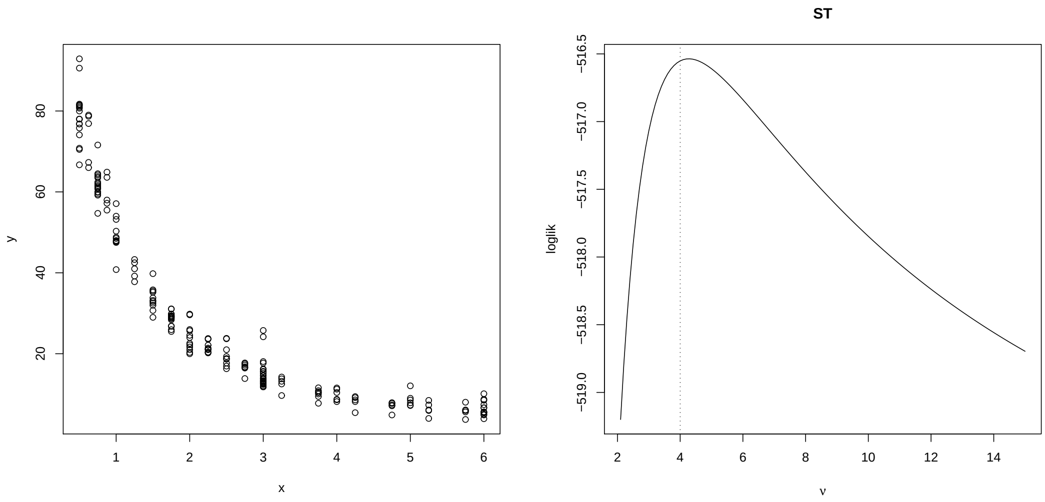

Figure 4: Ultrasonic calibration data. On the left the scatter-plot of the data set and on the right the plot of the profile log-likelihood of the parameter ν for fitting a ST-NLM.

6. Application on ultrasonic calibration data

In this section, we provide an application of the results derived in the previous sections using the ultrasonic calibration data described in Lin et al. (2009a). The required numerical evaluations were implemented using the statistical software package R (R Development Core Team, 2009). These data are the result of a NIST study involving ultrasonic calibration, where the response variable is ultrasonic response (y), and the predictor variable is metal distance (x) (see Figure 4). This data set was analyzed by Lin et al. (2009a), who considered a homoscedastic skew-t-normal nonlinear model. Following these authors, we consider the following SMSN-HNLM:

$$Y_i = \frac{\exp(-\beta_1 x_i)}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} SMSN(-\sqrt{\frac{2}{\pi}} k_1 \Delta, \sigma_i^2, \lambda; H), \tag{24}$$

where $\sigma_i^2 = \sigma^2 x_i^{\rho}$, for $i = 1, \ldots, 214$. In our analysis we will assume the SN, the ST and the SSL distributions from the SMSN class for comparative purposes.

6.1. Estimation models:

We choose the value of $\boldsymbol{\nu}$ by maximizing the the likelihood function as illustrated on the right of Figure 4; for the ST model we found $\nu = 4$ and for the SSL we found $\nu = 2$. Actually, with $\nu = 2$ the variance of the skew-slash distribution is finite. Further we fit Normal (N-HNLM) and Student-t models (T-HNLM) for comparative purposes. Table 2 contains the ML estimates of the parameters from the five models, together with their corresponding standard errors calculated via the observed information matrix. The AIC model selection criterion indicate that the heavy-tails SMSN models present the best fit, with the ST model significantly better. 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 (ρ, σ^2 and λ) are not comparable since they are on different scale.

Table 2: ML estimation results for fitting various mixture models on the Ultrasonic Calibration Data. SE are the asymptotic standard errors based on the observed information matrix.

	N-HNLM		T-HNLM		SN-HNLM		ST-HNLM		SSL-HNLM	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	0.148	0.0159	0.157	0.0135	0.154	0.0151	0.156	0.0142	0.155	0.0145
β_2	0.005	0.0004	0.005	0.0003	0.006	0.0004	0.005	0.0004	0.005	0.0004
β_3	0.012	0.0008	0.012	0.0006	0.012	0.0007	0.012	0.0007	0.012	0.0007
ρ	-0.959	0.1263	-1.035	0.0829	-0.987	0.1272	-0.959	0.1802	-1.025	0.1579
σ^2	16.20	2.1045	8.837	0.6165	33.785	5.8221	11.323	3.0454	13.86	3.2922
λ					2.248	0.4398	0.885	0.3799	1.477	0.4238
ν	-	-	4			-	4	-	2	
log-likelihood	-531.076		-519.328		-521.454		-516.5663		-517.507	
AIC	1072.152		1050.656		1054.908		1047.133		1049.014	

6.2. Model checking

In order to detect incorrect specification of the error distribution and the systematic component (24), 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 N-HNLM, SN-HNLM and T-HNLM, since there are not observations falling outside the envelope. Moreover, there is clear evidence of lack of fit for the N-HNLM and SN-HNLM. In the following we proceed our analysis using just asymmetric models.

6.3. Influence diagnostics analysis:

In this sub-section, we compute case-deletion measures and analysis of local influence for the ultrasonic calibration data by using the heteroscedastic nonlinear regression models under SMSN distributions.

- *Case deletion model:*

  Here case-deletion measures GD_i^l and LD_i^1, as presented in Subsection 5.1, are computed. The results are displayed in Figure 5. Considering the generalized Cook's distance GD_i^l, we observe that cases (♯146, ♯147 and ♯176) are identified as the most influential in the estimation of the parameters under the SN model. Meanwhile, no observation are influential under the ST-HNLM and SSL-HNLM. The same pattern can be seen when we use the likelihood distance LD_i^1, where only observations (♯146, ♯147, and ♯176) are influential under the ST-HNLM, SSL-HNLM and SN-HNLM models. For this data set the ST and SSL models accommodates better the influential observations. As expected, the influence of such observations are reduced when we consider distributions with heavier tails than the skew-normal one.

  Note from Figure 6 that when we use distributions with heavier tails 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 ($\widehat{u}_i, i = 1, \ldots, 214$) are indicated in Figure 6 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 $\boldsymbol{\theta}$. We consider the benchmark for $M(0)$ as described in Section 5.2, with $c^* = 2.0$

- *Case weights perturbation*

  Under this perturbation scheme, we obtain $C_{\mathbf{d}_{max}} = 1.93$ for the SN, $C_{\mathbf{d}_{max}} = 1.64$ for ST and $C_{\mathbf{d}_{max}} = 1.78$ for the SSL distribution, as values of maximum curvature. Clearly the smaller values correspond to the heavy-tails models. From Figure 7, it is noted that under the ST-HNLM and SSL-HNLM, the observation ♯176 is identified as influential. In addition, the observation ♯147 is also identified as influential under the SN-HNLM. Once again, for this data set, the ST and SSL models accommodate slightly better the influential observations than the SN-HNLM under case weights perturbation.

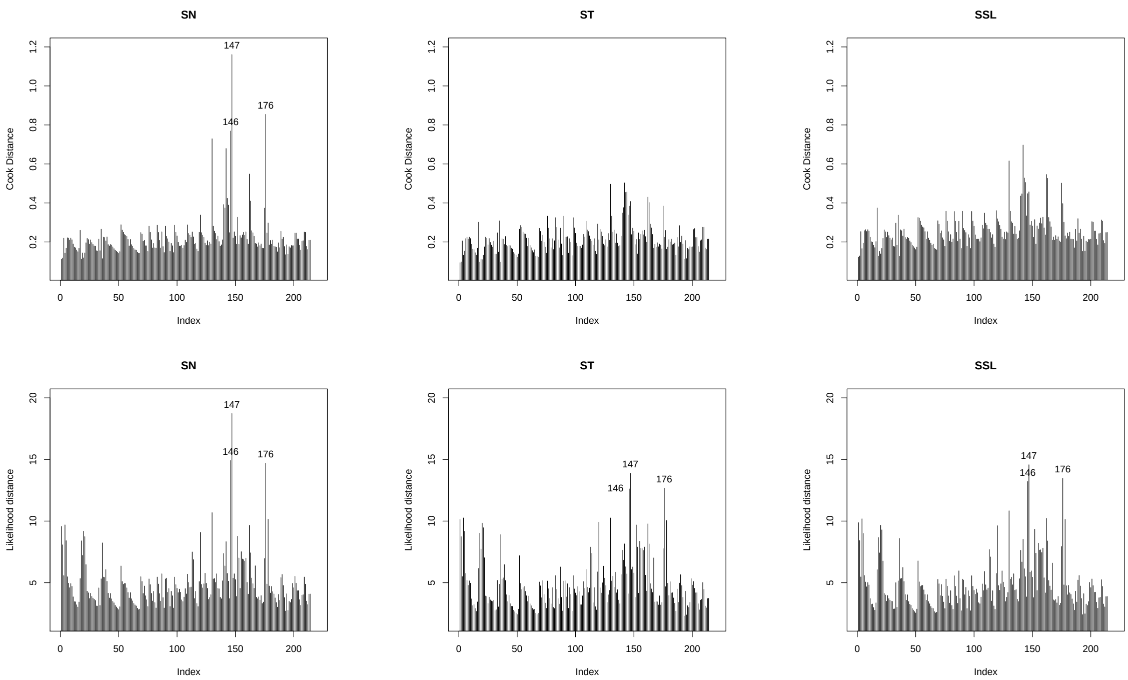

Figure 5: Ultrasonic calibration data. In the first row the index plots of the Generalized Cook's distance GD_i^l, and in the second row the index plots of the Likelihood Distance LD_i^l for SN-HNLM, ST-HNLM and SSL-HNLM.

6.4. Heteroscedasticity test

We now consider the test for heteroscedasticity for the ultrasonic calibration data based on the likelihood ratio test statistic. In the previous analysis, we saw that the ST-HNLM is the most appropriate for these data, so our analysis will be based on this distribution. From (24), it is easily seen that when $\rho = 0$, then $m_i = x_i^\rho = 1$ and thus $\sigma_i^2 = \sigma^2$ for all i. Hence, the likelihood ratio test for the homogeneity of scalar parameter becomes the test of hypothesis $H_0 : \rho = 0$, which is provided by

$$LR = 2\left(\ell(\widehat{\rho}, \widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma}^2, \widehat{\lambda} - \ell(0, \widetilde{\boldsymbol{\beta}}^\top, \widetilde{\sigma}^2, \widetilde{\lambda})\right)$$

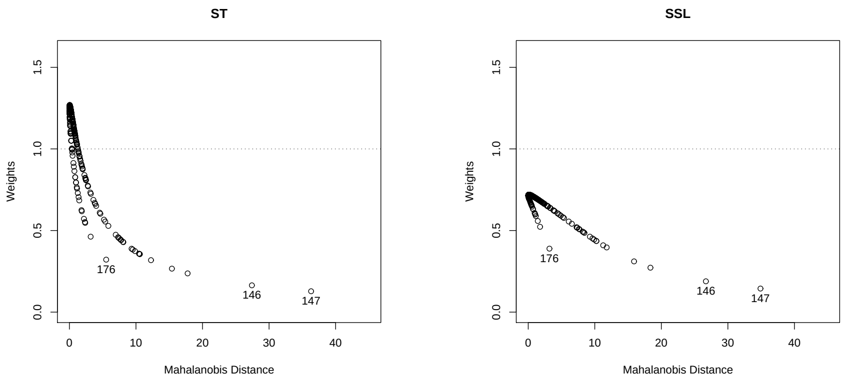

Figure 6: Estimated u_i for the ST-HNLM and the SSL-HNLM, on the ultrasonic calibration data.

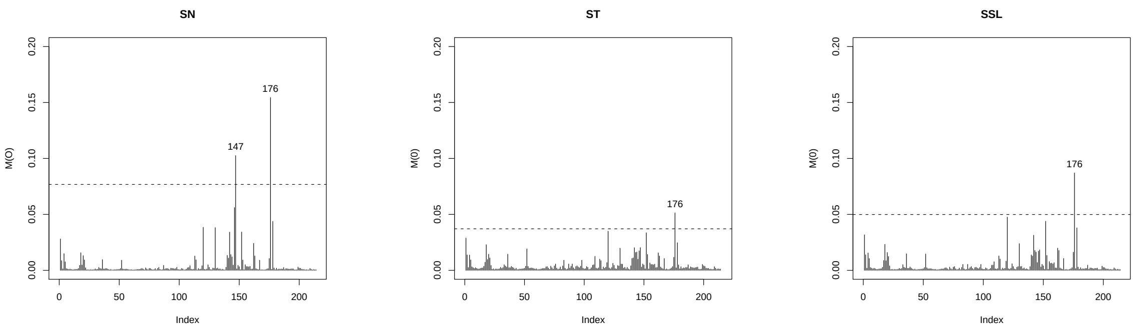

Figure 7: Ultrasonic calibration data. Index plot of M(0) in the case weights perturbation using SN-HNLM, ST-HNLM and SSL-HNLM models.

where $\widetilde{\boldsymbol{\beta}}^\top, \widetilde{\sigma}^2, \widetilde{\lambda}$ denote the ML estimates under H_0 (the homoscedastic SMSN-NLM).

Based on the statistic SR and a little computation, we get $LR_{ST} = 33.0472$ and the corresponding *p-value* ≈ 0, which rejects the hypothesis of homoscedasticity. This result is in accordance with that obtained by Lin et al. (2009b) using the score statistic test and therefore the assumption of homogeneity of variance is not suitable for the ultrasonic calibration data.

We believe that the likelihood ratio statistics is very sensitive to the presence of influential observations, $\sharp 146$, $\sharp 147$ and $\sharp 176$, (based in the diagnostic analysis) so we eliminate these observations from the full data and by similar computation we get $LR_{ST} = 31.2995$ and the corresponding *p-value* ≈ 0, which indicates that we can not reject the hypothesis H_0 and therefore the heteroscedasticity of the data is preserved, i.e. for this data the presence of possible influential observations do not affect the heteroscedasticity of the model.

Table 3 presents the results of the LR statistics for the three asymmetric models with the full data and the incomplete data (without the influential observations), we note that the hypothesis of homoscedasticity is rejected in all the cases. These results reinforces the fact that there is strong evidence of presence of heteroscedasticity and it should be considered to make the relevant inferences in the ultrasonic calibration data. Similar conclusions emerged when we chose $m_i(\boldsymbol{\rho}, \mathbf{z}_i) = \exp\left(\sum_{j=1}^{p} \rho_j z_{ij}\right)$, for $j = 1$.

Table 3: Likelihood ratio test statistic and *p-value* for various mixture models on the ultrasonic calibration data.

	All observations		All - {influential observations}	
Model	Statistics	*p-value*	Statistics	*p-value*
SN-HNLM	68.4013	1.110223e-16	67.3665	≈ 0
ST-HNLM	33.0472	8.994735e-09	31.2995	≈ 0
SSL-HNLM	44.3483	2.748557e-11	41.1843	≈ 0

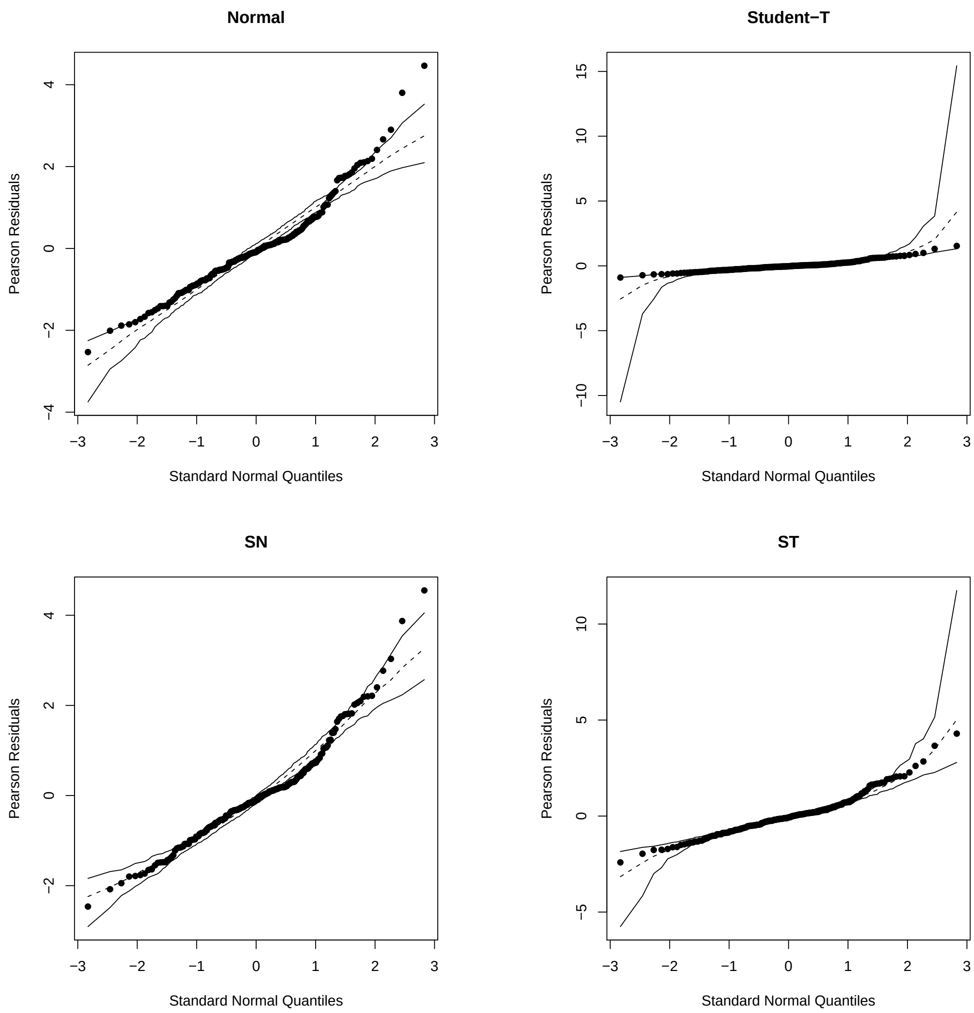

Figure 8: Ultrasonic Calibration Data. Q–Q plots and simulated envelopes for the Pearson Residuals

7. Conclusions

Modeling NLM is a research area with several challenging aspects. In this paper, we have proposed the application of a new class of asymmetric distributions, called the SMSM distribution, to the heteroscedastic NLM, where both mean and dispersion parameters vary across observations through nonlinear regression structures. An EM-type algorithm to obtain MLEs is developed by exploring the statistical properties of the SMSN class. A small simulation study is presented, showing the robust aspect of this flexible class against outlying and influential observations and that the maximum likelihood estimates based on the EM-type algorithm do provide good asymptotic properties. From our illustrated ultrasonic calibration data it is encouraging that the use of ST-HNLM offers better fitting than the skew-normal (and Student-t) counterpart. Moreover, through influence diagnostic procedures some aspects of robustness of the maximum likelihood estimators under SMSN distributions were note.

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-HNLM. 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-HNLM to multivariate settings and nonlinear mixed effects models. Finally, the proposed EM algorithm has been coded and implemented in the R package (R Development Core Team, 2009) and is available from the authors upon request.

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

Appendix : 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 $B_i = (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}) - b\sigma m_i^{1/2}\delta)/(\sigma m_i^{1/2}) = C_i - b\delta$, with $C_i = (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}))/(\sigma m_i^{1/2})$.

- 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 with respect to λ.

- 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}$$

References

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

Barroso, L., Cordeiro, G., 2005. Bartlett corrections in heteroskedastic t regression models. Statistics & Probability Letters 75 (2), 86–96.

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 doi:10.1016/j.csda.2009.09.031.

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. 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.

Cysneiros, F., Cordeiro, G., Cysneiros, A., 2009. Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. Journal of Statistical Computation and Simulation - (999), 1–11.

Cysneiros, F., Paula, G., Galea, M., 2007. Heteroscedastic symmetrical linear models. Statistics & 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.

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., XIE12, F., WEI, B., 2009a. Statistical Diagnostics for Skew-t-Normal Nonlinear Models. Communications in Statistics-Simulation and Computation 38 (10), 2096–2110.

Lin, J. G., Xie, F. C., Wei, B. C., 2009b. Statistical diagnostics for skew-t-normal nonlinear models. Communications in Statistics - Simulation and Computation 38, 2096–2110.

Liu, C., Rubin, D. B., 1994. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. Biometrika 80, 267–278.

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

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.

R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL `http://www.R-project.org`

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

Xie, F. C., , Lin, J. G., Wei, B. C., 2009a. Diagnostics for skew-normal nonlinear regression models with ar(1) errors. Computational Statistics & Data Analysis (In press).

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