Published in final edited form as:

Stat Probab Lett. 2011 August 1; 81(8): 1208–1217. doi:10.1016/j.spl.2011.03.019.

Heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions

Victor H. Lachos^a, Dipankar Bandyopadhyay^{b,*}, and Aldo M. Garay^a

Victor H. Lachos: hlachos@ime.unicamp.br; Dipankar Bandyopadhyay: bandyopd@musc.edu; Aldo M. Garay: amedina@ime.unicamp.br

^a Departamento de Estatística, Universidade Estadual de Campinas, Brazil

^b Division of Biostatistics and Epidemiology, Medical University of South Carolina, USA

Abstract

An extension of some standard likelihood based procedures to heteroscedastic nonlinear regression models under scale mixtures of skew-normal (SMSN) distributions is developed. We derive a simple EM-type algorithm for iteratively computing maximum likelihood (ML) estimates and the observed information matrix is derived analytically. Simulation studies demonstrate the robustness of this flexible class against outlying and influential observations, as well as nice asymptotic properties of the proposed EM-type ML estimates. Finally, the methodology is illustrated using an ultrasonic calibration data.

Keywords

EM algorithm; Homogeneity; Nonlinear regression models; Scale mixtures; Skew-normal

1. Introduction

Normal nonlinear regression models (N-NLM) are usually applied to model symmetrical data for which nonlinear functions of unknown parameters are used, in order to explain or describe the phenomena under study. However, it is well-known 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, the normality assumption is replaced with more flexible classes of distributions. For instance, Cysneiros and Vanegas (2008) studied the symmetrical nonlinear regression model and performed an analytical and empirical study to describe the behavior of the standardized residuals. Cancho et al. (2010) introduced the skew-normal nonlinear regression model (SN-NLM) and uses an efficient EM algorithm for ML estimation. Xie et al. (2009a,b) developed score tests for testing homogeneity in the SN-NLM proposed by Cancho et al. (2010). Recently, Cysneiros et al. (2010) introduced the class of heteroscedastic symmetric nonlinear regression models where a joint iterative process for estimating the mean and dispersion parameters are proposed. They also derive the n^{-1} biases of the MLEs of the mean and dispersion parameters. A

© 2011 Elsevier B.V. All rights reserved.

*Corresponding author: Division of Biostatistics and Epidemiology, Department of Medicine, Medical University of South Carolina, 135 Cannon Street, Suite 303, Charleston, SC 29425. bandyopd@musc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

common feature of these classes of NLMs is that the Gaussian NLM is a particular member of the class.

Although these models are attractive, there is a need to check the distributional assumptions of the model errors because these can present skewness and heavy-tails simultaneously. Applications of asymmetric classes of distributions with heavy tails are quite common in the literature, see for example, the book edited by Genton (2004). Branco and Dey (2001) proposed to use scale mixtures of skew-normal (SMSN) distributions in order to deal with the problem of atypical data in an asymmetrical context. Interestingly, this rich class 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) distributions. In this article, we extend the heteroscedastic symmetric nonlinear regression models (Cysneiros et al., 2010), 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-heterogeneous nonlinear model (SMSN-HNLM) is defined.

The rest of the paper is organized as follows. Section 2 provides a brief introduction to the class of univariate scale-mixture of skew-normal (SMSN) distributions. The heteroscedastic asymmetric nonlinear model (SMSN-HNLM) is introduced in Section 3 and an EM-type algorithm is proposed for ML estimation. In Section 4, we present results of simulation studies in which the robustness characteristics (in presence of outliers) as well as the consistency properties of the EM-estimates are evaluated. The methodology proposed is applied to a real data set in Section 5 and some concluding remarks are presented in Section 6.

2. Scale-mixtures of Skew-normal (SMSN) distributions

We start with the definition of SMSN distributions. A random variable Y belongs to the SMSN family if it can be written as

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

where μ is a location parameter, Z is skew-normal random variable with location 0, scale σ^2 , skewness λ (i.e. $Z \sim SN(0, \sigma^2, \lambda)$), U is a positive random variable with cumulative distribution function (cdf) $H(\cdot; \nu)$ and ν is a scalar or vector parameter indexing the distribution of U . Given $U = u$, we have $Y|U = u \sim SN(\mu, u^{-1}\sigma^2, \lambda)$. Thus, the probability density function (pdf) of Y is given by

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

where $\varphi(\cdot; \mu, \sigma^2)$ 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 $f_0(y) = \int_0^\infty \varphi(y; \mu, u^{-1}\sigma^2) dH(u; \nu)$.

For a random variable Y with the pdf (2), a convenient stochastic representation (Basso et al., 2010) is given by

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

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 an absolute value. We thus have $E[Y] = \mu + \sqrt{\frac{2}{\pi}} k_1 \Delta$ and $V ar[Y] = \sigma^2(k_2 - b^2\delta^2)$, where $b = -\sqrt{\frac{2}{\pi}} k_1$ and $k_m = E[U^{-m/2}]$. We refer to Basso et al. (2010) for details and additional properties related to this class of distributions which includes various sub-classes, some of which are the *skew-normal distribution* denoted by $SN(\mu, \sigma^2, \lambda)$ when $U = 1$, the *skew-t distribution* denoted by $ST(\mu, \sigma^2, \lambda; \nu)$ when $U \sim \text{Gamma}(\nu/2, \nu/2)$, $\nu > 0$ and the *skew-slash distribution* denoted by $SSL(\mu, \sigma^2, \lambda; \nu)$ when $U \sim \text{Beta}(\nu, 1)$, $\nu > 0$.

3. The SMSN heteroscedastic nonlinear regression model

The heteroscedastic nonlinear regression model based on SMSN distributions (SMSN-HNLM) is defined as

$$Y_i = \eta(\beta, \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4)$$

where the Y_i are the responses, $\eta(\cdot)$ is an injective and twice continuously differentiable function with respect to the parameter vector $\beta = (\beta_1, \dots, \beta_p)^\top$, $\mathbf{x}_i$ is a vector of explanatory variable values, the random errors $\varepsilon_i \sim \text{SMSN}(-\sqrt{\frac{2}{\pi}} k_1 \Delta m_i^{1/2}, m_i \sigma^2, \lambda; H)$ where $m_i = m(\rho, \mathbf{z}_i)$ is a known positive continuously differentiable function, $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$ contains values of explanatory variables and $\rho = (\rho_1, \dots, \rho_q)^\top$. Following the properties from SMSN

distribution, we have $Y_i \sim \text{SMSN}(\eta(\beta, \mathbf{x}_i) + b_1 \Delta w_i^{1/2}, \sigma^2 w_i, \lambda; H)$, for $i = 1, \dots, n$. The NLM defined in (4) corresponds to the regression model with mean zero error distribution and hence the regression parameters are all comparable. We have $E[Y_i] = \eta(\beta, \mathbf{x}_i)$ and $V ar[Y_i] = \sigma^2 m_i(k_2 - b^2 d^2)$. We consider ν to be known. Thus, the log-likelihood function for the parameter vector $\theta = (\rho^\top, \beta^{\text{top}}, \sigma^2, \lambda)^\top$ given the observed sample $\mathbf{y} = (y_1, \dots, y_n)^\top$ is

$$\ell(\theta) = \sum_{i=1}^n \ell_i(\theta), \quad \text{where}$$

$$\ell_i(\theta) = \log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log m_i + \log K_i, \quad (5)$$

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

$U(\theta) = \sum_{i=1}^n U_i(\theta)$ where $U_i(\theta) = \frac{\partial \ell_i(\theta)}{\partial \theta}$, and the observed information matrix $\mathbf{J}(\theta) = - \sum_{i=1}^n \frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta^\top}$ can be easily obtained, the detailed expressions of these partial derivatives are presented in the online Supplementary Material associated with this paper. Note that, since we have a closed-form expression for the observed information matrix for θ , the Newton-Raphson (NR) method can be easily applied to obtain ML estimates. An oft-voiced complain of the

NR algorithm is that it may not converge unless good starting values are used. In the next sub-section, we discuss a technique more elaborate to find the ML estimates of the parameters vector θ , based on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977).

3.1. Parameter estimation via the EM-algorithm

To develop an efficient EM-type algorithm for ML estimation of the parameters of SMSN-HNLM, 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

$$Y_i|T_i=t_i \sim N_1(\eta(\beta, \mathbf{x}_i) + \Delta m_i^{1/2} t_i, U_i^{-1} m_i \Gamma), \quad (6)$$

$$T_i|U_i \sim TN_1(b, u_i^{-1}; (b, \infty)), \quad (7)$$

$$U_i \sim H(, \nu), \quad (8)$$

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 $TN_1(\mu_{T_i} + b, u_i^{-1} M_T^2; (b, \infty))$, with

$$M_T^2 = \frac{\Gamma}{\Delta^2 + \Gamma}, \mu_{T_i} = \frac{\Delta}{m_i^{1/2}(\Delta^2 + \Gamma)} (y_i - \eta(\beta, \mathbf{x}_i) - \Delta b m_i^{1/2}).$$

Now we proceed for the E-step of the algorithm. To represent the estimator of the parameter $\xi = g(\theta)$, we will use the general notation $\hat{\xi} = g(\hat{\theta})$, where $g(\cdot)$ is a generic function of θ . Letting $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{t} = (t_1, \dots, t_n)^\top$ and $\mathbf{u} = (u_1, \dots, u_n)^\top$, the complete-data log-likelihood function $\ell_c(\theta|\mathbf{y}, \mathbf{t}, \mathbf{u})$ associated with $(\mathbf{y}, \mathbf{t}, \mathbf{u})$ is given by

$$\ell_c(\theta|\mathbf{y}, \mathbf{t}, \mathbf{u}) = c - \frac{n}{2} \log \Gamma - \frac{1}{2} \sum_{i=1}^n \log m_i - \frac{1}{2\Gamma} \sum_{i=1}^n \frac{u_i}{m_i} (y_i - \eta(\beta, \mathbf{x}_i) - \Delta m_i^{1/2} t_i)^2, \quad (9)$$

where c is a constant that is independent of θ . Letting $\hat{u}_i = E[U_i|\theta = \hat{\theta}, y_i]$,

$\widehat{ut}_i = E[U_i t_i | \theta = \hat{\theta}, y_i]$, $\widehat{ut}_i^2 = E[U_i^2 t_i^2 | \theta = \hat{\theta}, y_i]$ and using known properties of conditional expectations, we obtain

$$\widehat{ut}_i = u_i(\widehat{\mu}_{T_i} + b) + \widehat{M}_T \widehat{\tau}_{1_i}, \quad \widehat{ut}_i^2 = \widehat{u}_i(\widehat{\mu}_{T_i} + b)^2 + \widehat{M}_T^2 + \widehat{M}_T(\widehat{\mu}_{T_i} + 2b)\widehat{\tau}_{1_i}, \quad (10)$$

where $\widehat{\tau}_{1_i} = E\left[U_i^{1/2} W_\Phi\left(\frac{U_i^{1/2} \widehat{\mu}_{T_i}}{\widehat{M}_T}\right) | \theta = \hat{\theta}, y_i\right]$. In each step, the conditional expectations $\hat{u}_i = \hat{u}_{1_i}$ and can $\widehat{\tau}_{1_i}$ be easily derived from the results of Subsection 2.1 given in Basso et al. (2010). These computationally attractive expressions are quite useful in implementing the M-step,

which consists of maximizing the expected complete data function or the Q -function over θ , given by

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

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

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 the *ECM* algorithm (Meng and Rubin, 1993). The *ECME* algorithm (Liu and Rubin, 1994) which is 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 ML estimation of the parameters of the SMSN distributions.

E-step—Given a current estimate $\hat{\theta}^{(k)}$, compute $\hat{u}_i^{(k)}, \hat{u}_i^{(k)}, \hat{u}_i^{(k)}$, for $i = 1, \dots, n$.

CM-step—Update $\hat{\theta}^{(k)}$ by maximizing $Q(\theta|\hat{\theta}^{(k)})$ over θ , which leads to the following expressions

$$\hat{\beta}^{(k+1)} = \arg\min_{\beta} (\mathbf{z}^{(k)} - \eta(\beta, \mathbf{x}))^\top \hat{\mathbf{U}}^{(k)} (\mathbf{z}^{(k)} - \eta(\beta, \mathbf{x})), \quad (11)$$

$$\hat{\Delta}^{(k+1)} = \frac{\sum_{i=1}^n \hat{u}_i^{*(k)} (y_i - \eta(\beta^{(k+1)}, \mathbf{x}_i))}{\sum_{i=1}^n \hat{u}_i^{(k)}}, \quad (12)$$

$$\hat{\Gamma}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left((y_i - \eta(\beta^{(k+1)}, \mathbf{x}_i))^2 \hat{u}_i^{*(k)} - 2\Delta^{(k+1)} (y_i - \eta(\beta^{(k+1)}, \mathbf{x}_i)) \hat{u}_i^{*(k)} + (\Delta^2)^{(k+1)} \hat{u}_i^{(k)} \right) \quad (13)$$

$$\hat{\rho}^{(k+1)} = \arg\max_{\rho} \sum_{i=1}^n \log(f(y_i|\beta^{(k+1)}, \sigma^{2(k+1)}, \lambda^{(k+1)}, \nu)). \quad (14)$$

where $\hat{u}_i^{*(k)} = \hat{u}_i^{(k)} / \hat{m}_i^{(k)}$, $\hat{u}_i^{*(k)} = \hat{u}_i^{(k)} / \hat{m}_i^{1/2(k)}$, $\hat{\mathbf{U}}^{(k)} = \text{Diag}(\hat{u}_1^{*(k)}, \dots, \hat{u}_n^{*(k)})$, $\mathbf{z}^{(k)}$ is the corrected observed response given by $\mathbf{z}^{(k)} = \mathbf{y} - \hat{\Delta}^{(k)} \hat{\tau}^{(k)}$ with $\hat{\tau}^{(k)} = (\hat{\tau}_1^{(k)}, \dots, \hat{\tau}_n^{(k)})^\top$, $\hat{\tau}_i^{(k)} = \hat{u}_i^{*(k)} / \hat{u}_n^{*(k)}$ and $\eta(\beta, \mathbf{x}) = (\eta(\beta, \mathbf{x}_1), \dots, \eta(\beta, \mathbf{x}_n))^\top$. This process is iterated until some distance involving two successive evaluations of the actual log-likelihood $\ell(\theta)$, like $|\ell(\hat{\theta}^{(k+1)}) - \ell(\hat{\theta}^{(k)})|$ or $|\ell(\hat{\theta}^{(k+1)}) / \ell(\hat{\theta}^{(k)}) - 1|$, (where $|\cdot|$ denotes absolute value) is small enough. An interesting observation is that the M-step used to estimate β is equivalent to the weighted nonlinear least squares in the NLM, $\mathbf{z} = \eta(\beta, \mathbf{x}) + \varepsilon$, for which reliable and efficient algorithms are available in standard softwares as SAS, R, Ox and Matlab. Note that $\hat{\sigma}^{2(k+1)}$ and $\hat{\lambda}^{(k+1)}$ can

be recovered using the fact that $\lambda = \Delta / \sqrt{\Gamma}$ and $\sigma^2 = \Delta^2 + \Gamma$. The approximate variance-covariance matrix of the ML estimates can be evaluated by inverting the expected information. In SMSN-HNLM, the expected information matrix does not have a closed-form

expression. Instead, the observed information matrix, $\mathbf{J}(\theta) = - \sum_{i=1}^n \frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta^T}$, can be obtained by applying standard matrix differentiations and the results are provided in the Appendix. The asymptotic covariance matrix of $\hat{\theta}$ can be approximated by $\mathbf{J}^{-1}(\hat{\theta})$.

Remark—Although several authors have addressed parameter estimation of mixing variables in nonlinear regression (see, for example, Lange and Sinsheimer, 1993; Jamshidian, 1999) based upon the EM algorithm and its variants, Fernández and Steel (1999) discusses potential problems that may arise in the estimation of degrees of freedom, in particular for the Student-*t* distribution. This is due to the apparent unboundedness of the likelihood function near the boundary of the parameter space, and hence the ML scheme as developed in Lange and Sinsheimer (1993) are questionable because they do not provide sufficient information on whether these estimates correspond to local, or global maximas. Interestingly, Lucas (1997) notes that only under fixed degrees of freedom, the parameter estimates behave robust against extreme observations. A plausible (and simple) alternative is to assume that the parameter ν associated with the mixture variables U is known, which has been adopted in this work. Recent works in the context of elliptical distributions have considered the parameter ν to be known. See for instance Vanegas and Cysneiros (2010a), and Meza et al. (2011).

3.2. Notes on implementation

It is well known that maximum likelihood estimation in nonlinear models may face some computational hurdles, in the sense that the method may not give maximum global solutions if the starting values are far from the real parameter values. Thus, the choice of starting values for the EM algorithm in the non-linear context plays a big role in parameter estimation. In our example, we consider the following procedure to calculate the EM estimates under the skew-normal (SN-HNLM):

- Compute the initial values $\hat{\beta}^{(0)}$, $\hat{\rho}^{(0)}$ and $(\sigma^2)^{(0)}$. For this step, there are many optimization procedures available in standard softwares such as *nlreg()* or *ellipticalhet()* in R, considering normality.
- Compute the initial value $\hat{\lambda}^{(0)}$ by considering twice the signal of the skewness coefficient of the residuals in the step above.
- We use the EM estimates of the SN-HNLM as initial values for the corresponding skew-t (ST-HNLM), skew-slash (SSL-HNLM) and skew-contaminated normal (SCN-HNLM) parameters.

In order to estimate ν in the ST-HNLM and SSL-HNLM we have fixed integer values for ν from 3 to 30 and 2 to 30 by 1, respectively, choosing the value of ν that maximizes the likelihood function. A similar procedure has been adopted for the SCN-HNLM. See Figure 4 for more details.

4. Simulation study

In order to study the performance of our proposed model and algorithm, we present two simulation studies. Simulation 1 shows the necessity of heavy-tailed asymmetric models to deal with presence of outliers in the data whereas Simulation 2 shows that the ECME-based ML estimates have nice large sample properties. To conduct our simulation studies, we

generate our data from the following nonlinear growth-curve model (see, Cancho et al., 2010):

$$Y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \varepsilon_i, \quad i=1, \dots, n, \quad (15)$$

where $\varepsilon_i \sim SMSN(-\sqrt{\frac{2}{\pi}} k_1 \Delta m_i^{1/2}, \sigma_i^2, \lambda; H)$, with $\sigma_i^2 = \sigma^2 \exp(\rho z_i)$, are independent and identically distributed with zero mean. The variable x_i ranges from 0.1 to 20 and these values were fixed throughout the simulations with $z_i = x_i/10$. The parameter values were set around the estimates as obtained in Cancho et al. (2010), say, $\beta_1 = 37$, $\beta_2 = 43$, $\beta_3 = 0.6$, $\sigma^2 = 0.5$, $\lambda = -3$ and $\rho = -0.2$. The initial values were calculated as indicated in Section 3.2 and the maximum number of iteration considered was 100.

4.1. Simulation 1: Robustness of estimates

The goal of this study is to compare the performance of the ML estimates in the presence of outliers. In order to do so, we generated 1000 data sets of size $n = 50$, considering

$\varepsilon \sim SN(-\sqrt{2/\pi} \Delta m_i, \sigma_i^2, \lambda)$ in (15). To guarantee the presence of outliers, we constructed $Y_i^* = Y_i - \vartheta$, where i is the corresponding central value of sample and $\vartheta = 0, 1, 2, \dots, 9$. In each replication, we obtain parameters estimates with and without outliers denoted by $\hat{\theta}$ and $\hat{\theta}_{(i)}$, respectively, by fitting both Y and Y^* under skew-normal (SN-HNLM), skew-t (ST-HNLM) with $\nu = 3, 6, 10$ and 12 and skew-slash (SSL-HNLM) with $\nu = 2$. Then we computed the relative changes $|\frac{\hat{\gamma}_{(i)}}{\gamma} - 1|$, where $\gamma = \beta_1, \beta_2, \beta_3, \sigma^2$ or ρ . Figures 1 and 2 display the average relative changes on the parameter estimations for 1000 replications. Specifically, Figure 1 compares the SN case with ST ($\nu = 3$) and SSL ($\nu = 2$), whereas Figure 2 compares the SN with ST for various values of ν as discussed above. Notice from these figures that in the SN model, a considerable increase in relative change is noticed with increasing ϑ . However for the SMSN models with heavy tails, these measure varied little which indicates that they are more robust than the SN-HNLM in presence of discrepant observations/outliers. Note that unlike the other models considered, the influence of outliers on $\hat{\beta}$, $\hat{\rho}$ and $\hat{\sigma}^2$ for ST-HNLM and SSL-HNLM is not strictly increasing when ϑ increases. These figures reveal that models with heavier tails than the SN has better capability of reducing and controlling the influence of outliers on parameter estimates. These results agree with similar considerations, presented in Vanegas and Cysneiros (2010b) in a symmetric context.

4.2. Simulation 2: Consistency properties

The main focus of this simulation study is the evaluation of bias and mean square error of the ECME estimates to study consistency properties. Here the sample sizes were fixed as $n = 30, 100, 200, 300, 500$. For each combination of parameters and sample size, 500 samples from (15) 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 tailed distribution. Using our proposed ECME algorithm, we now compute the 'Relative Bias' and 'Mean Squared Error' (MSE) for each parameter over the 500 samples under the 3 different situations.

They are defined as:

$$\text{Relative Bias}(\gamma) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\gamma}^{(i)} / \gamma - 1) \quad \text{and} \quad \text{MSE}(\gamma) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\gamma}^{(i)} - \gamma)^2,$$

where $\gamma = (\beta_1, \beta_2, \beta_3, \sigma^2, \rho, \lambda)$ and $\hat{\gamma}^{(i)}$ is the ECME estimate of γ for the i th sample. The results are shown in Table 1 and Figure 3. To conserve space, we presented plots of Relative Bias and MSE of β_3, σ^2, ρ and λ in Figure 3. We can see a pattern of convergence to zero of the Bias and MSE as n increases. To conclude, we can comment that (relative) bias and MSE tend to approach zero with increasing sample size, indicating that the approximate MLEs derived from the proposed EM-type algorithm do possess good consistency properties.

5. Application

In this section, we use the ultrasonic calibration data described in Lin et al. (2009) to illustrate our methodology. These data are generated from a National Institutes of Standards and Technology (NIST) study by Dan Chwirut involving ultrasonic calibration (Chwirut, 1979), where the response variable is ultrasonic response y and the predictor variable is metal distance x (see left panel of Figure 4). The data consist of 214 observations, and is available freely in the R package *NISTnls*. This data set was analyzed by Lin et al. (2009) who considered a homoscedastic skew-t-normal nonlinear model. Following them, we consider the following SMSN-HNLM:

$$Y_i = \frac{\exp(-\beta_1 x_i)}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} SMSN\left(-\sqrt{\frac{2}{\pi}} k_1 \Delta, \sigma_i^2, \lambda; H\right), \quad (16)$$

where $\sigma_i^2 = \sigma^2 x_i^\rho$ for $i = 1, \dots, 214$. In our analysis, we will use the SN, the ST and the SSL distributions from the SMSN class for purpose of comparison. We choose ν by maximizing the likelihood function as illustrated in the right panel of Figure 4. For the ST model, $\nu = 4$ and for the SSL model, $\nu = 2$. 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 comparison. Table 2 shows the ML estimates of the parameters from fitting the five models and their corresponding standard errors calculated via the observed information matrix. The AIC and BIC model selection criterion indicate that the heavy-tailed SMSN models present the best fit, with the ST model providing significantly better fit. Although the estimates of regression parameters are very similar in all the three SMSN 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 estimates. The estimates for the variance components (σ^2 and λ) are not comparable since they are on different scale.

In order to detect incorrect specification of the error distribution and the systematic component in (16), we present QQ-plots and simulated envelopes for the Pearson residuals (Garay et al., 2011) in Figure 5. The figures clearly indicate that the ST-HNLM is more suitable for modeling this data than the N-HNLM, SN-HNLM and T-HNLM, since there are no observations falling outside the envelope. Moreover, there is clear evidence of lack of fit for the N-HNLM and SN-HNLM. Henceforth, we proceed with our analysis using asymmetric models. We now consider the test of heteroscedasticity for the ultrasonic calibration data using the likelihood ratio (LR) test statistic. Because ST-HNLM is the most appropriate model for these data, our analysis will be based on this distribution. From (16), it is easily seen that when $\rho = 0$, $m_i = x_i^\rho = 1$ and thus $\sigma_i^2 = \sigma^2$ for all i . Hence, the LR test is

equivalent to testing $H_0: \rho = 0$, which is given by $LR = 2 \left(\ell(\hat{\rho}, \hat{\beta}^\top, \hat{\sigma}^2, \hat{\lambda}) - \ell(0, \hat{\beta}^\top, \hat{\sigma}^2, \hat{\lambda}) \right)$ where $\hat{\beta}^\top, \hat{\sigma}^2, \hat{\lambda}$, denote the ML estimates under H_0 (the homoscedastic ST-NLM). After some computations, we obtain $LR_{ST} = 33.04722$ and the corresponding p -value ≈ 0 , which

rejects the hypothesis of homoscedasticity. This result is in accordance with that obtained by Lin et al. (2009) using the score test statistic and therefore the assumption of homogeneity of variance is not suitable for the ultrasonic calibration data.

The robustness of the ST-HNLM and SSL-HNLM can also be studied through the influence of a single outlying observation on the ML estimates of β . In particular, we can assess how much the ML estimates of θ is influenced by a change of ∇ units in a single observation Y_k . We replace a single y_k with $y_k(\nabla) = y_k - \nabla$ and record the relative change in the estimates $((\hat{\theta}(\nabla) - \hat{\theta})/\hat{\theta})$, where $\hat{\theta}$ is the original estimate and $\hat{\theta}(\nabla)$ is the estimate for the contaminated data. In this example, we contaminated the observation on subject 130 and varied ∇ between 0 and 8 with increments of 1. The results under SN-HNLM, ST-HNLM and SSL-HNLM for β_2 and β_3 are presented in Figure 6. As expected, the ST and SSL models are less adversely affected by variations of ∇ than the SN model.

6. Conclusions

Nonlinear regression modeling is an exciting research area that comes with interesting challenges. In this paper, we proposed application of a new class of asymmetric distributions (SMSM) to the heteroscedastic NLM, where both mean and dispersion parameters vary across observations through nonlinear regression structures. For parameter estimation, an EM-type algorithm is developed. Simulation studies reveal that our proposed class is quite robust against outlying and influential observations and that the ML estimates based on the EM-type algorithm possess nice consistency properties. The data application showed a better fit of the ST-HNLM over the usual normal (and symmetric) counterparts. The proposed EM-type algorithm has been coded and implemented in the R package (R Development Core Team, 2009) and is available from the first author on request.

Due to recent advances in computational technology, Bayesian treatments via Markov chain Monte Carlo (MCMC) techniques are also feasible in the context of SMSN-HNLM. Other extensions of the current work include but not limited to using finite mixture of skew-normal (Lin et al., 2007b) and skew-t (Lin et al., 2007a) distributions, which will be considered in a separate paper. Multivariate versions of our proposition for skew-normal (Lin, 2009) and skew-t (Lin, 2010) models are also worthy of investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research of V. H. Lachos was supported in part by grants from FAPESP-Brazil (2010/012465) and CNPq-Brazil (201384/2008-6). D. Bandyopadhyay acknowledges support from NIH/NCRR grant P20 RR017696-06. The authors thank an anonymous referee whose constructive comments led to a substantially improved version of this manuscript.

References

- Basso RM, Lachos VH, Cabral CR, Ghosh P. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*. 2010; 54:2926–2941.
- Branco MD, Dey DK. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*. 2001; 79:99–113.
- Cancho VC, Lachos VH, Ortega EMM. A nonlinear regression model with skew-normal errors. *Statistical Papers*. 2010; 51:547–558.
- Chwirut, D. Ultrasonic Reference Block Study. NIST; 1979. Unpublished

- Cysneiros F, Cordeiro G, Cysneiros A. Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*. 2010; 80 (4):451–461.
- Cysneiros FJA, Vanegas LH. Residuals and their statistical properties in symmetrical nonlinear models. *Statistics & Probability Letters*. 2008; 78:3269–3273.
- Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1977; 39:1–38.
- Fernández C, Steel M. Multivariate Student-t regression models: Pitfalls and inference. *Biometrika*. 1999; 86 (1):153–167.
- Garay A, Lachos V, Abanto-Valle C. Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society*. 2011; 40 (1):115–124.
- Genton, M. Skew-elliptical distributions and their applications: A journey beyond normality. Chapman & Hall/CRC; 2004.
- Jamshidian M. Adaptive robust regression by using a nonlinear regression program. *Journal of Statistical Software*. 1999; 4 (6):1–25.
- Lange KL, Sinsheimer JS. Normal/independent distributions and their applications in robust regression. *J Comput Graph Stat*. 1993; 2:175–198.
- Lin JG, Xie FC, Wei BC. Statistical diagnostics for skew-t-normal nonlinear models. *Communications in Statistics - Simulation and Computation*. 2009; 38 (10):2096–2110.
- Lin T. Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*. 2009; 100 (2):257–265.
- Lin T. Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing*. 2010; 20 (3):343–356.
- Lin T, Lee J, Hsieh W. Robust mixture modeling using the skew t distribution. *Statistics and Computing*. 2007a; 17 (2):81–92.
- Lin T, Lee J, Yen S. Finite mixture modelling using the skew normal distribution. *Statistica Sinica*. 2007b; 17 (3):909–927.
- Liu C, Rubin DB. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*. 1994; 80:267–278.
- Lucas A. Robustness of the student t based M-estimator. *Communications in Statistics-Theory and Methods*. 1997; 26 (5):1165–1182.
- Meng X, Rubin DB. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*. 1993; 81:633–648.
- Meza C, Osorio F, De la Cruz R. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*. 2011; 10.1007/s11222-010-9212-1
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2009. URL <http://www.R-project.org>
- Vanegas L, Cysneiros F. Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*. 2010a; 54 (4):1002–1016.
- Vanegas LH, Cysneiros FJA. Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*. 2010b; 54:1002–1016.
- Xie FC, Lin JG, Wei BC. Diagnostics for skew-normal nonlinear regression models with ar(1) errors. *Computational Statistics & Data Analysis*. 2009a; 53:4403–4416.
- Xie FC, Wei BC, Lin JG. Homogeneity diagnostics for skew-normal nonlinear regression models. *Statistics & Probability Letters*. 2009b; 79:821–827.

Figure 1.

Average relative changes on estimates of β_2 , β_3 , σ^2 and ρ for SN-HNLM, ST-HNLM and SSL-HNLM following the non-linear growth curve model in (15).

Figure 2.

Average relative changes on estimates of β_2 , β_3 , σ^2 and ρ for SN-HNLM and various cases of ST-HNLM following the non-linear growth curve model in (15).

Figure 3. Relative Bias and MSE of the ECME estimates of β_3 , σ^2 , ρ and λ for SN-HNLM, ST-HNLM and SSL-HNLM following the non-linear growth curve model in (15).

Figure 4. Ultrasonic calibration data. Left panel indicates the scatter-plot of the raw data and right panel shows the plot of the profile log-likelihood of v after fitting a ST-NLM.

Figure 5.
Q-Q plots and simulated envelopes for the Pearson Residuals in the Ultrasonic Calibration Data

Figure 6. Relative changes on the ML estimates of β_2 and β_3 when fitting a SN-HNLM, ST-HNLM and SSL-HNLM for different contaminations of ∇ for the observation 130 in the Ultrasonic Calibration Data.

Table 1

Simulated data. Relative Bias in parameter estimates with different sample sizes for the SN, ST and the SSL distribution in the nonlinear growth-curve model.

SN-HNLM						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	4.90351e-04	0.003967	3.55959e-05	0.421119	-0.236017	-0.265620
100	1.83577e-04	0.002336	6.52559e-05	0.216799	0.058515	0.270628
200	5.61969e-05	0.001496	3.20991e-04	0.108726	0.027597	0.119602
300	6.29276e-05	0.001170	2.00521e-04	0.035878	0.006430	0.064091
500	4.80572e-05	-0.000769	-1.54338e-04	-0.038779	-0.013584	0.030746
ST-HNLM						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	1.83478e-03	0.001619	-0.001242	0.735691	0.017582	0.024945
100	3.28755e-04	0.004664	0.000449	0.280626	0.071530	0.249456
200	2.15526e-04	0.000417	-0.000233	0.074182	0.023902	0.119779
300	2.59449e-05	-0.001056	-0.000409	0.004251	0.011682	0.077090
500	-3.17269e-05	0.000468	0.000155	-0.016919	0.001062	0.029263
SSL-HNLM						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	6.58573e-04	0.008680	4.11286e-04	0.643667	0.012857	0.274990
100	6.47356e-05	0.005947	1.10452e-03	0.226230	0.051769	0.238995
200	1.21090e-04	0.000617	-2.07541e-04	0.118503	0.034882	0.116208
300	-2.13150e-05	0.001346	2.14653e-04	0.050668	0.023833	0.069388
500	-1.76627e-05	0.000706	6.13428e-05	0.027645	0.005698	0.020706

Table 2

ML estimation results for fitting various HNLM models to the Ultrasonic Calibration Data. SE are the asymptotic standard errors based on the observed information matrix.

Parameter	N-HNLM		T-HNLM		SN-HNLM		ST-HNLM		SSL-HNLM	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	0.148	0.01595	0.157	0.01354	0.154	0.01512	0.156	0.01415	0.155	0.01449
β_2	0.005	0.00041	0.005	0.00033	0.006	0.00038	0.005	0.00036	0.005	0.00037
β_3	0.012	0.00076	0.012	0.00064	0.012	0.00071	0.012	0.00066	0.012	0.00068
ρ	-0.959	0.12627	-1.035	0.08299	-0.987	0.12717	-0.959	0.18024	-1.025	0.15795
σ^2	16.20	2.10449	8.837	0.61650	33.785	5.82211	11.323	3.04540	13.86	3.29217
λ					2.248	0.43980	0.885	0.37991	1.477	0.42380
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	
BIC	1086.566		1063.07		1067.322		1057.547		1059.428	