JIRSS (2018)
Vol. 17, No. 02, pp 13-35
DOI: 10.29252/jirss.17.2.3

Nonlinear Regression Models Based on Slash Skew-Elliptical Errors

S. Pirzadeh Nahooji[1], **Rahman Farnoosh**[2] **and Nader Nematollahi**[3]

[1] Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
[2] School of Mathematics, Iran University of Science and Technology, Narmak, Tehran, Iran.
[3] Department of Statistics, Allameh Tabataba'i University, Tehran, Iran.

Received: 01/11/2016, Revision received: 01/03/2018, Published online: 07/08/2018

Abstract. In this paper, the nonlinear regression models when the model errors follow a slash skew-elliptical distribution, are considered. In the special case of nonlinear regression models under slash skew-t distribution, we present some distributional properties, and to estimate their parameters, we use an EM-type algorithm. Also, to find the estimation errors, we derive the observed information matrix analytically. To describe the influence of the observations on the ML estimates, we use a sensitivity analysis. Finally, we conduct some simulation studies and a real data analysis to show the performance of the proposed model.

Keywords. EM algorithm, Nonlinear regression, Skew normal distribution, Skew-t distribution, Slash skew-elliptical distribution, Slash skew-t distribution.

MSC: 62J02; 62E15.

S. Pirzadeh Nahooji (si_pirzad@yahoo.com)
Corresponding Author: Rahman Farnoosh (rfarnoosh@iust.ac.ir)
Nader Nematollahi (nematollahi@atu.ac.ir).

1 Introduction

One of the most interesting problems in regression theory is the nonlinear regression models (NLM) which are used in many fields such as finance, economics, sociology, engineering and biomedical sciences. These models are used when the response variable is a nonlinear function of explanatory variables and the unknown parameters. Usually, the error term in these models is a random variable which has a normal distribution or, in general, a symmetric distribution. However, when the data under the study do not possess a symmetric property or have heavy tails, then the normal NLM is not appropriate. In these cases, the use of the asymmetric and heavy tail distributions for the error terms in NLM are more appropriate than the normal NLM.

Recently, some researchers have used flexible parametric non-normal distributions for error term variables in NLM. Some of these asymmetric and heavy tailed distributions are skew normal distribution (Azzalini, 1985, 2005), the skew-t distribution (Jones and Faddy, 2003), the skew-elliptical distribution (Branco and Dey, 2001; Sahu *et al.*, 2003) and the Slash-elliptical distribution (Alcantara and Cysneiros, 2017). In the context of non-normal NLM, Cancho *et al.* (2010) introduced the skew-normal NLM (SN-NLM) and used an efficient EM-type algorithm for estimation of its parameters. Xie *et al.* (2009a,b) developed score tests for testing homogeneity in the SN-NLM proposed by Cancho *et al.* (2010). Garay *et al.* (2011) extend the SN-NLM by assuming that the model errors follow a mean-zero, scale mixtures of skew-normal distribution. Alcantara and Cysneiros (2017) used slash-elliptical distribution for error term in NLM and estimate its parameters by the maximum likelihood method. Such classes of distributions contain skewed versions of classical distributions such as skew-normal, skew-t, skew-slash (Gomez and Venegas, 2008) and skew contaminated normal distribution (Basso *et al.*, 2010). Lachos *et al.* (2011) defined scale mixtures of skew-normal heteroscedastic NLM and estimate its parameters by applying an EM-type algorithm.

Although the above researchers used asymmetric and heavy tailed distributions for modeling error terms in NLM, there are some situations in which the data need a more flexible skewed and heavy tailed distribution for the model errors (see Section 5.2 below). This motivates us to employ a slash skew-elliptical distribution for the distribution of the model errors in NLM, which was introduced by Farnoosh *et al.* (2013). This class of distributions can be more skewed and heavy tailed than the skew-normal and the skew-t distributions, so it is appropriate for the distribution of random errors of NLM which have the skewed and heavy tailed properties.

The content of the paper is as follows. Section 2 will provide a brief overview on the slash-skew elliptical distributions and a special case of it, i.e., slash skew-t distribution.

The slash skew elliptical NLM and the special case, NLM with slash skew-t distribution, are introduced in Section 3. In Section 4, in the special case of slash skew-t NLM, an EM-type algorithm is constructed to estimate the parameters. In addition, the observed information matrix is calculated analytically and a sensitivity analysis is introduced for investigating the influence of observations on the ML estimates. Also, the uses of standardized residuals in these asymmetrical NLM are evaluated in the presence of outliers. Simulation studies and an application to a real data are illustrated in Section 5 to show the performance of the proposed model. Finally, in Section 6, concluding remarks are presented.

2 Slash Skew-elliptical Distributions

In this section, we review definitions and the properties of slash skew-elliptical distribution which is introduced by Farnoosh *et al.* (2013) . A random variable has slash skew-elliptical (SLSEL) distribution if it can be written as

$$Y = \mu + \sigma \frac{X}{U^{\frac{1}{q}}}, \tag{2.1}$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and $q > 0$ is the shape parameter that controls the tails. The random variable X here is a skew elliptical random variable with location 0, scale 1 and $-\infty < \lambda < \infty$ is the skewness parameter, denoted by $X \sim SEL(0, 1, \lambda; g)$ where $g(.)$ is the density generator function of an elliptical distribution . U is the uniform random variable which is independent of X. Farnoosh *et al.* (2013) denoted this distribution by $Y \sim SLSEL(\mu, \sigma, \lambda, q; g)$. The probability density function (p.d.f.) of Y is given by

$$f_Y(y) = \begin{cases} \frac{q\sigma^q}{|y-\mu|^{q+1}} \int_0^{(\frac{y-\mu}{\sigma})^2} u^{\frac{q-1}{2}} g(u) F_g(\lambda \sqrt{u} h(y-\mu)) du & y \neq \mu \\ \frac{q}{\sigma(q+1)} g(0) & y = \mu, \end{cases} \tag{2.2}$$

where $h(t) = \frac{t}{|t|}$. When $q \to \infty$, we have skew elliptical distribution ($SEL(\mu, \sigma, \lambda; g)$). Also, if $\lambda = 0$, slash skew-elliptical distribution reduces to slash elliptical distribution ($SLEL(\mu, \sigma, q; g)$) defined by Gomez and Venegas (2008). Now, we represent the following results of slash skew-elliptical distribution. These results can be used to simulate slash skew- elliptical random variable and to implement an EM-type algorithm. The proofs can be found in Farnoosh *et al.* (2013).

Theorem 2.1. *If $X \sim SLSEL(\mu, \sigma, \lambda, q; g)$ and $Y = aX + b$, $a, b \in \mathbb{R}$, then $Y \sim SLSEL(a\mu + b, |a|\sigma, \lambda h(a), q; g)$.*

Theorem 2.2. *If $Y|U = u \sim SEL(0, u^{-\frac{1}{q}}, \lambda; g)$ and $U \sim U(0,1)$, then $Y \sim SLSEL(0, 1, \lambda, q; g)$.*

Theorem 2.3. *If $Z \sim SLSEL(0, 1, \lambda, q; g)$ and $Y \sim SLSEL(\mu, \sigma, \lambda, q; g)$, then*

$$\mu_k = E(Z^k) = \frac{q}{q-k} a_k, \quad k = 0, 1, 2, ..., \quad (q > k),$$

where $a_k = 2\int_{\mathbb{R}} x^k g(x^2) F_g(\lambda x) dx$ is the r-th moment of $SEL(0, 1, \lambda; g)$ distribution and

$$\mu_k^{'} = E(Y^k) = \sum_{i=0}^{k} \binom{k}{i} \sigma^i \mu^{k-i} \mu_i, \quad k = 0, 1, 2, ..., \quad (q > k).$$

2.1 A Special Case

In the special case of (2.1), let X have a skew-t distribution with skewness λ and r degrees of freedom, i.e., $X \sim ST(0, 1, \lambda, r)$. Therefore, Y has the p.d.f. given by (2.2) with the generator function

$$g(t) = \frac{\Gamma(\frac{1+r}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}} (1 + \frac{t}{r})^{-\frac{1+r}{2}}, \qquad t \in \mathbb{R}. \tag{2.3}$$

Farnoosh *et al.* (2013) have called this distribution slash skew-t distribution (SLST), and denoted it by $Y \sim SLST(\mu, \sigma, \lambda, q, r)$. They illustrated that the SLST distribution can be more skewed and heavier tailed than the other skew distributions like skew-t and skew slash.

Further, if $Y \sim SLST(\mu, \sigma, \lambda, q, r)$, from Theorem 2.3 we have,

$$\begin{aligned} \mu_1^{'} &= E(Y) = \mu + c\delta\sigma, && q > 1,\ r > 1, \\ \mu_2^{'} &= E(Y^2) = \mu^2 + 2\mu c\delta\sigma + \frac{rq}{(q-2)(r-2)}\sigma^2, && q > 2,\ r > 2, \end{aligned} \tag{2.4}$$

and

$$Var(Y) = \frac{rq}{(q-2)(r-2)}\sigma^2 - (c\delta\sigma)^2, \qquad q > 2,\ r > 2, \tag{2.5}$$

where $\delta = \frac{\lambda}{\sqrt{(1+\lambda^2)}}$ and $c = \frac{q}{q-1} \frac{\Gamma(\frac{r-1}{2})}{\Gamma(\frac{r}{2})} \sqrt{\frac{r}{\pi}}$.

We remind that if $X \sim ST(0,1,\lambda,r)$, Azzalini and Capitanio (2003) showed that X can be represented by $X = \frac{Z}{\sqrt{V}}$ where $Z \sim SN(0,1,\lambda)$ and $V \sim \frac{1}{r}\chi_r^2$ (chi-square distribution with r degrees of freedom). If $Z \sim SN(0,1,\lambda)$, Henze (1986) showed that a stochastic representation of Z is given by

$$Z = \delta|T_0| + (1-\delta^2)^{\frac{1}{2}} T_1,$$

where T_0 and T_1 are independent standard normal random variables and $|.|$ denotes the absolute value.

Consequently, if $Y \sim SLST(\mu,\sigma,\lambda,q,r)$, from equation (2.1) and $X \sim ST(0,1,\lambda,r)$, we can write the stochastic representation of Y as follows

$$\begin{aligned} Y &= \mu + \sigma\delta U^{\frac{-1}{q}} V^{\frac{-1}{2}} |T_0| + \sigma(1+\delta^2)^{\frac{1}{2}} U^{\frac{-1}{q}} V^{\frac{-1}{2}} T_1 \\ &= \mu + \xi T + \Lambda^{\frac{1}{2}} U^{\frac{-1}{q}} V^{\frac{-1}{2}} T_1, \end{aligned} \tag{2.6}$$

where $\xi = \sigma\delta$, $\Lambda = \sigma^2(1-\delta^2)$ and $T = U^{\frac{-1}{q}} V^{\frac{-1}{2}} |T_0|$. This representation is useful in the simulation data from SLST distribution and to implement the EM-type algorithm.

3 The Slash Skew-elliptical Nonlinear Regression Model

We denote the nonlinear regression model based on slash skew-elliptical distribution errors by SLSEL-NLM, and this model is defined as

$$Y_i = \psi(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where Y_i is the response variable, $\psi(.)$ is an injective and twice continuously differentiable function with respect to the parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, $\mathbf{x}_i$ is a vector of explanatory variable values. The random errors $\varepsilon_i \sim SLSEL(-\sigma a_1(\frac{q}{q-1}), \sigma, \lambda, q; g)$ for $q > 1$, where $a_1 = \int_{\mathbb{R}} x g(x^2) F_g(\lambda x) dx$ and $g(.)$ is the density generator function of elliptical distribution, which corresponds to the regression model, where the error distribution has mean zero and consequently the regression parameters are all comparable. Thus, from the above discussion and Theorems 2.1 and 2.3, we can obtain the following result,

$$E(Y_i) = \psi(\boldsymbol{\beta}, \mathbf{x}_i), \quad Var(Y_i) = q\sigma^2(\frac{a_2}{q-2} - \frac{a_1 q}{(q-1)^2}), \quad q > 2,$$

where $a_2 = \int_{\mathbb{R}} x^2 g(x^2) F_g(\lambda x) dx$ and $Y_i \sim SLSEL(\psi(\boldsymbol{\beta}, \mathbf{x}_i) - \sigma a_1(\frac{q}{q-1}), \sigma, \lambda, q; g)$, $i = 1, \ldots, n$. Hereafter, we focus on a special case of slash skew-elliptical nonlinear regression model, which is slash skew-t nonlinear regression model (SLST-NLM) and compare it with other members of this class of distributions, i.e., a skew normal distribution that has been used by Garay *et al.* (2011) and slash skew normal distribution for distributions of the model errors of nonlinear model.

Remark 1. Alcantara and Cysneiros (2017) used slash-elliptical distribution as error terms in nonlinear regression model 3.1 to construct slash-elliptical nonlinear regression model (SLEL-NLM), i.e.,

$$Y_i = \mu_i(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim SEL(0, \phi, q; g), \quad i = 1, \dots, n. \tag{3.2}$$

Note that when $\lambda = 0$, the SLSEL-NLM reduces to SLEL-NLM (3.2). Therefore, the model (3.2) is a special case of the model (3.1). The model (3.2) contains only heavy tail distributions for error terms, but the model (3.1) contains heavy tail and skew distributions for error terms.

3.1 The Slash Skew-t Nonlinear Regression Model

In slash skew-t nonlinear regression model (SLST-NLM), the random errors in equation (3.1) have slash skew-t distribution ($\varepsilon_i \sim SLST(-c\sigma\delta, \sigma, \lambda, q, r)$) where $c = \frac{q}{q-1}\frac{\Gamma(\frac{r-1}{2})}{\Gamma(\frac{r}{2})}\sqrt{\frac{r}{\pi}}$. So the responses variable Y_i has slash skew-t distribution, i.e., $Y_i \sim SLST(\psi(\boldsymbol{\beta}, \mathbf{x}_i) - c\sigma\delta, \sigma, \lambda, q, r)$, $i = 1, \dots, n$. Therefore

$$E(Y_i) = \psi(\boldsymbol{\beta}, \mathbf{x}_i), \quad Var(Y_i) = \frac{q}{q-2}\frac{r}{r-2}\sigma^2 - (c\delta\sigma)^2, \quad q > 2,\ r > 2.$$

Pirzadeh *et al.* (2015) obtained the Bayes estimate of the parameters of SLST-NLM via a Metropolis-Hastings algorithm.

4 Maximum Likelihood Estimation via an EM-type Algorithm

In this section, we derive the maximum likelihood estimate of the parameters of slash skew-t nonlinear regression model. We used an EM-type algorithm (Dempster *et al.*, 1977) which is used similarly by Garay *et al.* (2011) and Lachos *et al.* (2011) for the estimation of the parameters of scale mixtures of skew-normal nonlinear regression. Let $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ be n independent random variables, where $Y_i \sim SLST(\psi(\boldsymbol{\beta}, \mathbf{x}_i) - c\sigma\delta, \sigma, \lambda, q, r)$. Since error variance must be finite and when $\varepsilon_i \sim SLST(-c\sigma\delta, \sigma, \lambda, q, r)$ variance of ε_i is finite if $q > 2,\ r > 2$, so in computation we suppose that $q > 2,\ r > 2$.

On the other hand, Lango *et al.* (1989) and Berkane *et al.* (1994), pointed out difficulties in estimating r and q because of the problems of unbounded and local maximum in the likelihood function. Thus, at first, we determine the values of the parameters r and q based on the behavior of the tail of distribution or likelihood function and then estimate other parameters of SLST-NLM model (see Sections 5.1 and 5.2 for more details).

Stochastic representation (2.6) is useful for hierarchical representation for Y_i, in the EM-algorithm. Accordingly, $(U_1, V_1), \ldots, (U_n, V_n)$ are considered as missing data in equation (2.6) and Y_i is considered as observed data. Let $(Y_i, U_i, V_i), i = 1, \ldots, n$ be the complete data and $\theta = (\boldsymbol{\beta}^T, \sigma, \lambda)^T$ where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$. From (2.6), for $i = 1, \ldots, n$, we have,

$$\begin{aligned}
&Y_i|U_i = u_i, V_i = v_i, T_i = t_i \sim N(\psi(\boldsymbol{\beta}, x_i) + \xi t_i, u_i^{\frac{-2}{q}} v_i^{-1}\Lambda),\\
&T_i|U_i = u_i, V_i = v_i \sim TN(-c, u_i^{\frac{-2}{q}} v_i^{-1})I(-c, \infty),\\
&U_i \sim U(0,1),\\
&V_i \sim \frac{1}{r}\chi_r^2 = gamma(\frac{r}{2}, \frac{r}{2}),
\end{aligned}$$

where $TN(a, b^2)I(r, s)$ denotes the truncated normal distribution $N(a, b^2)$ on (r, s). A useful result is that the conditional distribution of T_i given y_i, u_i and v_i is

$$T_i|Y_i = y_i, U_i = u_i, V_i = v_i \sim TN(\mu_{T_i} - c, u_i^{\frac{-2}{q}} v_i^{-1} M_{T_i}^2)I(-c, \infty),$$

where $\mu_{T_i} = \frac{\xi}{\Lambda+\xi^2}(y_i - \psi(\boldsymbol{\beta}, x_i) + \xi c)$ and $M_{T_i}^2 = \frac{\Lambda}{\Lambda+\xi^2}$. Since we have, $f(y_i, u_i, v_i, t_i) = f(y_i|u_i, v_i, t_i) f(t_i|u_i, v_i) f(u_i) f(v_i)$, so,

$$\begin{aligned}
f(y_i, u_i, v_i, t_i) &= \frac{u^{\frac{-1}{q}} v^{\frac{-1}{2}}}{\sqrt{2\pi\Lambda}} \exp(\frac{-u^{\frac{2}{q}} v}{2\Lambda}(y_i - \psi(\boldsymbol{\beta}, x_i) - \xi t_i)^2)\\
&\times \frac{\frac{u^{\frac{-1}{q}} v^{\frac{-1}{2}}}{\sqrt{2\pi}} \exp(\frac{-u^{\frac{2}{q}} v(t_i+c)^2}{2})}{\int_{-c}^{\infty} \frac{u^{\frac{-1}{q}} v^{\frac{-1}{2}}}{\sqrt{2\pi}} \exp(\frac{-u^{\frac{2}{q}} v(t_i+c)^2}{2}) dt_i} \times \frac{(\frac{r}{2})^{\frac{r}{2}}}{\Gamma(\frac{r}{2})} v^{\frac{r}{2}-1} e^{\frac{-rv}{2}}.
\end{aligned}$$

The log-likelihood function for the complete data is given by

$$\ell_c(\theta|\mathbf{y}, t, u, v) = k - \frac{n}{2}\log\Lambda - \frac{1}{2\Lambda}\sum_{i=1}^{n} u^{\frac{2}{q}} v(y_i - \psi(\boldsymbol{\beta}, x_i) - \xi t_i)^2,$$

where k is a constant that is independent of θ. Now, the conditional expectation of $l_c(\theta|\mathbf{y}, t, u, v)$ given observed data y_i and current estimates of the parameters $\hat{\theta}$, is given by

$$\begin{aligned} Q(\theta|\hat{\theta}) &= E(l_c(\theta)|y_i, \hat{\theta}) = k - \frac{n}{2}\log\Lambda \\ &- \frac{1}{2\Lambda}\sum_{i=1}^{n}\Bigg[E(U_i^{\frac{2}{q}}V_i|y_i,\hat{\theta})(y_i-\psi(\boldsymbol{\beta},x_i))^2 + \xi^2 E(U_i^{\frac{2}{q}}V_iT_i^2|y_i,\hat{\theta}) \\ &- 2\xi(y_i-\psi(\boldsymbol{\beta},x_i))E(U_i^{\frac{2}{q}}V_iT_i|y_i,\hat{\theta})\Bigg]. \end{aligned} \tag{4.1}$$

Let

$$\begin{aligned} &\hat{k}_i = E\left[U_i^{\frac{2}{q}}V_i|\hat{\theta}, y_i\right], \quad \hat{S}_i = E\left[U_i^{\frac{2}{q}}V_i W_\Phi(\frac{\mu_{T_i}U_i^{\frac{1}{q}}V_i^{\frac{1}{2}}}{M_{T_i}})|\hat{\theta}, y_i\right], \\ &\hat{S}_{2i} = E\left[U_i^{\frac{2}{q}}V_iT_i|\hat{\theta}, y_i\right], \qquad \hat{S}_{3i} = E\left[U_i^{\frac{2}{q}}V_iT_i^2|\hat{\theta}, y_i\right], \end{aligned} \tag{4.2}$$

where $W_\Phi(x) = \frac{\phi(x)}{\Phi(x)}$ and $\phi(.)$ and $\Phi(.)$ denote the standard normal pdf and cumulative distribution function (c.d.f.), respectively. Since $Y_i|u_i, v_i \sim SN(\psi(\boldsymbol{\beta}, x_i) - c\xi, u^{\frac{-q}{2}}u^{\frac{-1}{2}}\sigma, \lambda)$ and using known properties of conditional expectation we have,

$$\begin{aligned} \hat{k}_i &= \frac{2}{\hat{\sigma}r\sqrt{r}f(y_i)}\int_0^1\int_0^\infty u_i^{\frac{3}{q}}w_i^{\frac{3}{2}}p(w_i)\phi(u_i^{\frac{1}{q}}w_i^{\frac{1}{2}}(\frac{y-\psi(\hat{\boldsymbol{\beta}},x_i)+c\xi}{\hat{\sigma}\sqrt{r}})) \\ &\times \Phi(u_i^{\frac{1}{q}}w_i^{\frac{1}{2}}\hat{\lambda}(\frac{y-\psi(\hat{\boldsymbol{\beta}},x_i)+c\xi}{\hat{\sigma}\sqrt{r}}))dw_i du_i, \end{aligned} \tag{4.3}$$

$$\begin{aligned} \hat{S}_i &= \frac{2}{\hat{\sigma}rf(y_i)}\int_0^1\int_0^\infty u_i^{\frac{2}{q}}w_ip(w_i)\phi(u_i^{\frac{1}{q}}w_i^{\frac{1}{2}}(\frac{y-\psi(\hat{\boldsymbol{\beta}},x_i)+c\xi}{\hat{\sigma}\sqrt{r}})) \\ &\times \Phi(u_i^{\frac{1}{q}}w_i^{\frac{1}{2}}\hat{\lambda}(\frac{y-\psi(\hat{\boldsymbol{\beta}},x_i)+c\xi}{\hat{\sigma}\sqrt{r}}))dw_i du_i, \end{aligned} \tag{4.4}$$

where $f(y_i)$ is the pdf of $SLST(\psi(\hat{\boldsymbol{\beta}}, x_i) - c\xi, \hat{\sigma}, q, r)$ (given by (2.2) with generator (2.3)) and $p(w_i)$ is the pdf of χ_r^2. Likewise, from properties of conditional expectation we have,

$$\hat{S}_{2i} = E_{u_i,v_i|y_i}\left[U_i^{\frac{2}{q}} V_i E\left[T_i|\hat{\theta}, y_i, u_i, v_i\right]\right], \tag{4.5}$$

$$\hat{S}_{3i} = E_{u_i,v_i|y_i}\left[U_i^{\frac{2}{q}} V_i E\left[T_i^2|\hat{\theta}, y_i, u_i, v_i\right]\right]. \tag{4.6}$$

Since $T_i|Y_i = y_i, U_i = u_i, V_i = v_i \sim TN(\mu_{T_i} - c, u_i^{\frac{-2}{q}} v_i^{-1} M_{T_i}^2) I(-c, \infty)$, we have

$$\begin{aligned}
E\left[T_i|\hat{\theta}, y_i, u_i, v_i\right] &= \hat{\mu}_{T_i} - c + U_i^{\frac{-1}{q}} V_i^{\frac{-1}{2}} \hat{M}_{T_i} W_\Phi\left(\frac{\hat{\mu}_{T_i} U_i^{\frac{1}{q}} V_i^{\frac{1}{2}}}{\hat{M}_{T_i}}\right), \\
E\left[T_i^2|\hat{\theta}, y_i, u_i, v_i\right] &= (\hat{\mu}_{T_i} - c)^2 + U_i^{\frac{-2}{q}} V_i^{-1} \hat{M}_{T_i}^2 \\
&+ U_i^{\frac{-1}{q}} V_i^{\frac{-1}{2}} \hat{M}_{T_i}^2 (\hat{\mu}_{T_i} - 2c) W_\Phi\left(\frac{\hat{\mu}_{T_i} U_i^{\frac{1}{q}} V_i^{\frac{1}{2}}}{\hat{M}_{T_i}}\right).
\end{aligned}$$

Now, by replacing the above expressions in (4.5) and (4.6), we have

$$\hat{S}_{2i} = \hat{k}_i(\hat{\mu}_{T_i} - c) + \hat{M}_T \hat{S}_i, \tag{4.7}$$

$$\hat{S}_{3i} = \hat{k}_i(\hat{\mu}_{T_i} - c)^2 + \hat{M}_T^2 + \hat{M}_T(\hat{\mu}_{T_i} - c)\hat{S}_i. \tag{4.8}$$

The above expressions are being implemented for maximizing the expected complete data function over θ, or the Q-function which is used in the M-step of the algorithm. From (4.1), (4.2), (4.7) and (4.8), the Q-function is given by

$$\begin{aligned}
Q(\theta|\hat{\theta}^{(k)}) &= E(l_c(\theta)|y_i, \hat{\theta}^{(k)}) = k - \frac{n}{2}\log\Lambda \\
&- \frac{1}{2\Lambda}\sum_{i=1}^{n}\left[\hat{k}_i^{(k)}(y_i - \psi(\boldsymbol{\beta}, x_i))^2 + \xi^2 \hat{S}_{2i}^{(k)} - 2\xi(y_i - \psi(\boldsymbol{\beta}, x_i))\hat{S}_{3i}^{(k)}\right],
\end{aligned} \tag{4.9}$$

where $\hat{\theta}^{(k)}$ is an updated value of $\hat{\theta}$ in k-th iteration of the EM-algorithm. In the E-step of the algorithm, given the observation y_i and current estimates $\hat{\theta}$, the conditional expectations $\hat{S}_i$, $\hat{k}_i$, $\hat{S}_{2i}$ and $\hat{S}_{3i}$ must be computed. For computing $\hat{k}_i$ and $\hat{S}_i$ Monte Carlo integration can be employed, which yields the so-called MC-EM algorithm. For the M-step of the algorithm, there is a need to maximize the Q-function over θ. The M-step turns out to be analytically intractable. But it is possible to be replaced by a sequence of conditional maximization (CM) steps. The resulting method is known as ECM algorithm (Meng and Rubin, 1993). Similar to Garay *et al.* (2011), we use ECM algorithm as follows.

E-step: Given a current estimate $\hat{\theta}^{(k)} = (\hat{\boldsymbol{\beta}}^{(k)}, \hat{\sigma}^{(k)}, \hat{\lambda}^{(k)})$ and observation $\mathbf{y} = (y_1, \ldots, y_n)$, compute $\hat{k}_i^{(k)}$ and $\hat{S}_i^{(k)}$, $i = 1, \ldots, n$ from (4.3) and (4.4) by Monte Carlo integration and then compute $\hat{S}_{2i}^{(k)}$ and $\hat{S}_{3i}^{(k)}$ from (4.7) and (4.8).

CM-step: Derive $\hat{\theta}^{(k+1)}$ by maximizing $Q(\theta|\hat{\theta}^{(k)})$ over θ, which are given by the following expressions

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} (\mathbf{z}^{(k)} - \psi(\boldsymbol{\beta}, \mathbf{x}))^T \hat{\mathbf{K}}^{(k)} (\mathbf{z}^{(k)} - \psi(\boldsymbol{\beta}, \mathbf{x})),$$

$$\xi^{(k+1)} = \frac{\sum_{i=1}^{n} \hat{S}_{2i}(y_i - \psi(\hat{\boldsymbol{\beta}}^{(k+1)}, x_i))}{\sum_{i=1}^{n} \hat{S}_{3i}},$$

$$\hat{\Lambda}^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{k}_i^{(k)} (y_i - \psi(\hat{\boldsymbol{\beta}}^{(k+1)}, x_i))^2 + (\xi^2)^{(k+1)} \hat{S}_{2i}^{(k)} - 2\xi^{(k+1)} (y_i - \psi(\hat{\boldsymbol{\beta}}^{(k+1)}, x_i \hat{S}_{3i}^{(k)})) \right),$$

where $\hat{\mathbf{K}}^{(k)} = diag(\hat{k}_1^{(k)}, \ldots, \hat{k}_n^{(k)})$ and $\mathbf{z}^k$ is the corrected observed response given by $\mathbf{z}^k = \mathbf{y} - \hat{\xi}^{(k)} \hat{\mathbf{S}}^{(k)}$, with $\hat{\mathbf{S}}^{(k)} = (\hat{S}_1^{(k)}, \ldots, \hat{S}_n^{(k)})$ and $\psi(\boldsymbol{\beta}, \mathbf{x}) = (\psi(\boldsymbol{\beta}, x_1), \ldots, \psi(\boldsymbol{\beta}, x_n))^T$. Note that estimating $\boldsymbol{\beta}$ in M-step of algorithm is equivalent to estimating $\boldsymbol{\beta}$ in the weighted nonlinear least squares in the NLM, $\mathbf{Z} = \psi(\boldsymbol{\beta}, \mathbf{x}) + \boldsymbol{\varepsilon}$, where we used NLM package in R software for estimating $\boldsymbol{\beta}$.

By using the fact that $\lambda = \xi / \sqrt{\Lambda}$ and $\sigma^2 = \xi^2 + \Lambda$, the values of $\hat{\sigma}^{2(k+1)}$ and $\hat{\lambda}^{(k+1)}$ can be found as

$$\hat{\lambda}^{(k+1)} = \frac{\hat{\xi}^{(k+1)}}{\sqrt{\hat{\Lambda}^{(k+1)}}}, \qquad \hat{\sigma}^{2(k+1)} = (\hat{\xi}^{(k+1)})^2 + \hat{\Lambda}^{(k+1)}.$$

This process is iterated until a suitable convergence rule is satisfied, e.g. if $||\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}|| < 0.0001$.

4.1 The Observed Information Matrix

In this section, we obtain the observed information matrix of the nonlinear regression model based on SLST distribution, which is defined by

$$\mathbf{J}_0(\Theta|\mathbf{y}) = -\frac{\partial^2 \ell(\Theta|\mathbf{y})}{\partial\Theta\partial\Theta^T},$$

where $\ell(\Theta|\mathbf{y})$ is the incomplete likelihood function based on observation $\mathbf{y}$. Under some regularity conditions, the covariance matrix of the maximum likelihood estimates $\hat{\Theta}$ can be approximated by the inverse of $\mathbf{J}_0(\Theta|\mathbf{y})$. The observed information matrix can be evaluated as follows,

$$\mathbf{J}_0(\hat{\Theta}|\mathbf{y}) = \sum_{i=1}^{n} \hat{\mathbf{t}}_i \hat{\mathbf{t}}_i^T, \tag{4.10}$$

where

$$\hat{\mathbf{t}}_i = \frac{\partial(\log f(y_i;\theta_j))}{\partial\theta_j}, \qquad j = 1,\ldots,p,p+1,p+2,$$

and p is the number of model parameters (see, Basford *et al.* (1997) and Lin *et al.* (2007)). Now, partition $\hat{\mathbf{t}}_i$ into components corresponding to all the parameters in Θ, i.e., $\hat{\mathbf{t}}_i = (\hat{\mathbf{t}}_{i,\boldsymbol{\beta}}, \hat{t}_{i,\sigma^2}, \hat{t}_{i,\lambda})^T$ where

$$\hat{t}_{i,\theta_j} = \frac{\partial(\log f(y_i;\theta_j))}{\partial\theta_j}, \qquad j = 1,\ldots,p,p+1,p+2.$$

Now, we define

$$\begin{aligned} I_{i1}^{F_1}(v_1,v_2) &= \int_0^1 u^{\frac{v_1}{q}} \left(1 + \frac{u^{\frac{2}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi)^2}{r\sigma^2}\right)^{\frac{-(r+v_2)}{2}} \\ &\times \; F_1(\lambda \sqrt{\frac{1+r}{r}}((r\sigma^2 + u^{\frac{2}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi)^2)^{\frac{-1}{2}} u^{\frac{1}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi); r+1)du, \end{aligned}$$

$$\begin{aligned} I_{i1}^{f_1}(v_1,v_2,v_3) &= \int_0^1 u^{\frac{v_1}{q}} \left(1 + \frac{u^{\frac{2}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi)^2}{r\sigma^2}\right)^{\frac{-(r+v_2)}{2}} \\ &\times \; ((r\sigma^2 + u^{\frac{2}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi)^2)^{\frac{-v_3}{2}} \\ &\times \; f_1(\lambda \sqrt{\frac{1+r}{r}}((r\sigma^2 + u^{\frac{2}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi)^2)^{\frac{-1}{2}} u^{\frac{1}{q}}(y_i - \psi(\boldsymbol{\beta},x_i) + c\xi); r+1)du, \end{aligned}$$

where $f_1(x;r+1)$ and $F_1(x;r+1)$ are p.d.f. and c.d.f. of Student-t distribution with $r+1$ degrees of freedom.

After some algebraic calculations, we obtain

$$\begin{aligned}\frac{\partial}{\partial\beta_j}(f(y_i;\Theta)) &= \frac{\partial(\psi(\boldsymbol{\beta},x_i))}{\partial\beta_j}\times\frac{\partial(f(y_i;\Theta))}{\partial(\psi(\boldsymbol{\beta},x_i))}\\ &= \frac{\partial(\psi(\boldsymbol{\beta},x_i))}{\partial\beta_j}\times\frac{2\Gamma(\frac{r+1}{2})}{\sigma\Gamma(\frac{r}{2})\sqrt{r\pi}}\left[\frac{(r+1)(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)^2}{r\sigma^2}I_{i1}^{F_1}(3,3)\right.\\ &+ \left.\lambda\sqrt{\frac{1+r}{r}}(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)^2 I_{i1}^{f_1}(4,1,3)-I_{i1}^{f_1}(2,1,1)\right],\quad j=1,\dots,p,\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\sigma}(f(y_i;\Theta)) &= \frac{2\Gamma(\frac{r+1}{2})}{\sigma\Gamma(\frac{r}{2})\sqrt{r\pi}}\left[-\frac{1}{\sigma}I_{i1}^{F_1}(1,1)-\frac{(r+1)(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)(\psi(\boldsymbol{\beta},x_i)-y_i)}{r\sigma^3}I_{i1}^{F_1}(3,3)\right.\\ &- r\sigma(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)I_{i1}^{f_1}(2,1,3)-c\delta(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)^2 I_{i1}^{f_1}(4,1,3)\\ &+ \left.c\delta I_{i1}^{f_1}(2,1,1)\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\lambda}(f(y_i;\Theta)) &= \frac{2\Gamma(\frac{r+1}{2})}{\sigma\Gamma(\frac{r}{2})\sqrt{r\pi}}\sqrt{\frac{1+r}{r}}(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)I_{i1}^{f_1}(2,1,1)\\ &+ \frac{c\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{r\pi}}((1+\lambda^2)^{-\frac{1}{2}}-\lambda^2(1+\lambda^2)^{-\frac{3}{2}})\left[\frac{-(r+1)(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)^2}{r\sigma^2}I_{i1}^{F_1}(3,3)\right.\\ &- \left.\lambda\sqrt{\frac{1+r}{r}}(y_i-\psi(\boldsymbol{\beta},x_i)+c\xi)^2 I_{i1}^{f_1}(4,1,3)+\lambda\sqrt{\frac{1+r}{r}}I_{i1}^{f_1}(2,1,1)\right].\end{aligned}$$

The information-based approximation (4.10) is asymptotically applicable. However, it may not be reliable unless the sample size is large.

In the next section, we use the techniques that have been used to estimate the parameters.

4.2 Residuals

The aim of residual analysis is to identify atypical observations and model misspecification when 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 use the following standardized ordinary residual (Pearson residual) that has been used by Garay *et al.* (2011),

$$r=\frac{y_i-\hat{\mu}_i}{\sqrt{Var(y_i)}},\qquad i=1,\dots,n,$$

where $Var(y_i) = \frac{q}{q-2}\frac{r}{r-2}\hat{\sigma}^2 - (c\hat{\delta}\hat{\sigma})^2$. Here $\hat{\mu}_i = \psi(\hat{\boldsymbol{\beta}}, x_i)$ and $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$, $\hat{\delta}$ denote the maximum likelihood estimators of $\boldsymbol{\beta}$, σ^2 and δ, respectively. As the distribution of the standardized residual is not known, we follow the suggestion given by Atkinson (1981) to construct the simulated envelope. The simulated envelope can be used as a helpful diagnostic tool to detect incorrect specification of the error distribution and the systematic component $\psi(\boldsymbol{\beta}, x_i)$, as well as the presence of outlying observations. An oft-voiced complaint of this method is that it may be very slow in some situations, once we need to generate and fit the model for a number $k \geq 100$ of simulated samples (see, Garay *et al.*, 2011).

4.3 Sensitivity Analysis

To detect the influence of observations on the ML estimators, we perform sensitivity analysis with scale-deletion method in this section and Section 5.2.1 to recognize observations that under small perturbation of the model exert great influence on the ML estimates. This method has been used in some papers, such as Cook (1977), Lin *et al.* (2009) and Rahnamaei *et al.* (2012). We use the case-deletion approach to detect the influence of removing the case from the analysis by evaluating the metrics such as the likelihood distance and Cook's distance (see, Cook, 1977).

Let $\hat{\Theta}_i$ be the ML estimate of Θ without the i-th observation in the sample. To assess the influence of the i-th case on the ML estimate $\hat{\Theta}$, the basic idea is to compare the difference between $\hat{\Theta}_i$ and $\hat{\Theta}$. If deletion of a case seriously influences the estimates, more attention should be paid to that case. Therefore, if $\hat{\Theta}_i$ is far from $\hat{\Theta}$, then the i-th case is considered as an influence observation. Hence, generalized Cook distance is employed to measure the change between $\hat{\Theta}_i$ and $\hat{\Theta}$, which can be expressed as

$$GD_i(\Theta) = (\hat{\Theta}_i - \hat{\Theta})^T[\mathbf{J}_0(\hat{\Theta}|\mathbf{y})](\hat{\Theta}_i - \hat{\Theta}),$$

where $\mathbf{J}_0(\hat{\Theta}|\mathbf{y})$ is the observed information matrix (in $\Theta = \hat{\Theta}$ point) discussed in Section 4.1.

Another measure of the difference between $\hat{\Theta}$ and $\hat{\Theta}_i$ is the likelihood distance

$$LD_i(\Theta) = 2\left[L(\hat{\Theta}) - L(\hat{\Theta}_i)\right].$$

In the next section, we employ sensitivity analysis to illustrate the advantage of the proposed methodology.

5 Applications

In this section, we present two applications of the SLST-NLM. The first one is based on two simulation studies and the other is a statistical analysis of the real data sets.

5.1 Simulation Studies

In this section, to show the performance of the proposed model and the given algorithm, we present two simulation studies. Simulation 1 is conducted to show the need of using heavy-tailed asymmetric models to deal with the presence of outliers in the data. Simulation 2 is conducted to show the large sample properties of the ECM-type algorithms. We generate our data from the following nonlinear growth-curve model,

$$Y_i = \frac{\beta_1}{1+\beta_2 exp(-\beta_3 x_i)} + \varepsilon_i, \quad i = 1,\dots,50, \tag{5.1}$$

where $\varepsilon_i \sim SLST(-c\delta\sigma,\sigma,\lambda,q,r)$ are independent and identically distributed variables with zero mean. The variable x_i ranges from 4 to 53 and the values were held fixed throughout the simulations. The parameter values were set around the estimates as obtained in Cancho *et al.* (2010), say, $\beta_1 = 37$, $\beta_2 = 42$, $\beta_3 = 0.73$, $\sigma^2 = 2.95$, $\lambda = -2$. To have a heavy-tailed distribution, we choose the values of r and q such that they maximize the value of tail measure $P(Y > \mu + 3\sigma)$ between values of $3 \le r \le 10$ and $3 \le q \le 10$. Thus, we obtain $r = 3$ and $q = 10$ for this simulation study.

5.1.1 Simulation 1: Robustness of estimates

In this simulation study, we want to compare the performance of the ML estimates in the presence of outliers. To do this, we generated 100 data sets of size $N = 50$, considering $\varepsilon_i \sim SN(-\sqrt{\frac{2}{\pi}}\delta\sigma,\sigma,\lambda)$ in (5.1), i.e., the data are generated from the SN-NLM model. Following Vanegas and Cysneiros (2010), to guarantee the presence of one outlier, we constructed $Y_i^* = Y_i - v$, where i is the corresponding central value of the sample and $v = 1,2,3,4,5,6,7,8,9$. In each replication, we obtained the parameter estimates with and without outliers denoted by $\hat{\theta}$ and $\hat{\theta}_{(i)}$, respectively, under the skew normal (SN-NLM), the slash skew normal (SLSN-NLM) and the slash skew-T (SLST-NLM). Then, we computed the relative changes $\left|\frac{\hat{\theta}_{(i)}}{\hat{\theta}} - 1\right|$ on estimates of β_1, β_2, β_3 and σ^2 when the outlier is removed from the data set.

Figure 1 shows the average values of relative changes on the estimates for 100 replications. We can observe that for all three models, the influence of the outliers

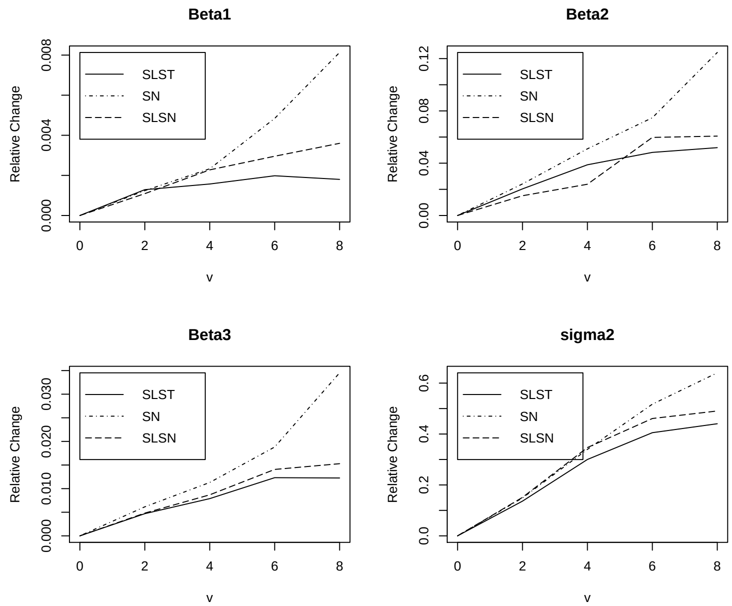

Figure 1: Simulated data: Average changes on estimates of β_1, β_2, β_3 and σ^2 for the nonlinear growth-curve model.

in the estimates increases as v increases. Note that for the SLST model, this measure increases less than the SN and the SLSN models when v increases. So we can conclude that SLST model is more robust than the SN and the SLSN models in the presence of discrepant observations. These figures reveal the models with heavier tails than the SN and the SLSN have a better capability of reducing and controlling the influence of outliers on parameter estimates.

5.1.2 Simulation 2: Consistency properties

In this simulation study, we want to evaluate the bias and mean squared error of the ECM estimates to study consistency properties. In this simulation, the sample sizes were fixed as $n = 50, 100, 200, 300, 500$. For each combination of parameters and sample size, 100 samples from (5.1) were generated under slash skew-T (SLST-NLM). Using our proposed ECME algorithm, we compute the Relative Bias and Mean Squared Error

(MSE) for each parameter over the 100 samples under the SLST-NLM . They are defined as:

$$RelativeBias(\theta) = \frac{1}{100}\sum_{i=1}^{100}\left(\frac{\hat{\theta}^{(i)} - \theta}{\theta}\right), \quad MSE(\theta) = \frac{1}{100}\sum_{i=1}^{100}\left(\hat{\theta}^{(i)} - \theta\right)^2,$$

where $\boldsymbol{\theta} = (\beta_1, \beta_2, \beta_3, \lambda, \sigma^2)$ and $\hat{\theta}^{(i)}$ is the ECME estimate of θ for the i-th sample. The results are shown in Table 1 and Figures 4 and 5. We can see that the Bias and MSE converge to zero as n increases, which imply that the approximate MLEs derived from the proposed EM-type algorithm possess good consistency properties.

Table 1: Simulated data: Relative Bias of parameter estimates with different sample sizes for SLST distribution in the nonlinear growth-curve.

n	β_1	β_2	β_3	λ	σ^2
SLST-NLM					
50	-0.0005898	0.9190525	0.0531100	1.5628815	0.1656811
100	0.0007456	0.6346358	-0.0035434	0.5519101	0.0916926
200	0.0004373	0.0625517	0.0068832	0.1481456	0.0387276
300	0.0003897	0.0454593	0.0024557	0.0374658	-0.0167413
500	0.0002683	0.0235612	0.0030808	0.0871107	0.0087041

5.2 Real Data

We use the ultrasonic calibration data described in Lin *et al.* (2009) and Lachos *et al.* (2011) to investigate our method. These data are generated from the National institutes of Standards and Technology (NIST) study by Dan Chwirut involving ultrasonic calibration (Chwirut, 1979), where the response variable and the predictor variable is metal distance. The response variable Y has mean 30.26, variance 560.7326, minimum 3.75, maximum 92.90, skewness 0.9131793 and kurtosis 2.556237. The data consist of observations, and are available freely in the R package *Nlsmsn*. The left plot in Figure 2 represents the scatter plot of response and predictor variables and shows that they have a nonlinear relationship. We consider the following nonlinear model

$$Y_i = \frac{\exp(-\beta_1 x_i)}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad \varepsilon_i \sim SLST(-c\sigma\delta, \sigma, \lambda, q, r), \quad i = 1, \dots, n. \tag{5.2}$$

We will use the SN, the SLSN and the SLST distributions for comparing the models. To implement the EM algorithm, the initial values of the parameters (σ^2, λ) are derived by the method of moments, and the model parameters $(\beta_1, \beta_2, \beta_3)$ estimates are derived by

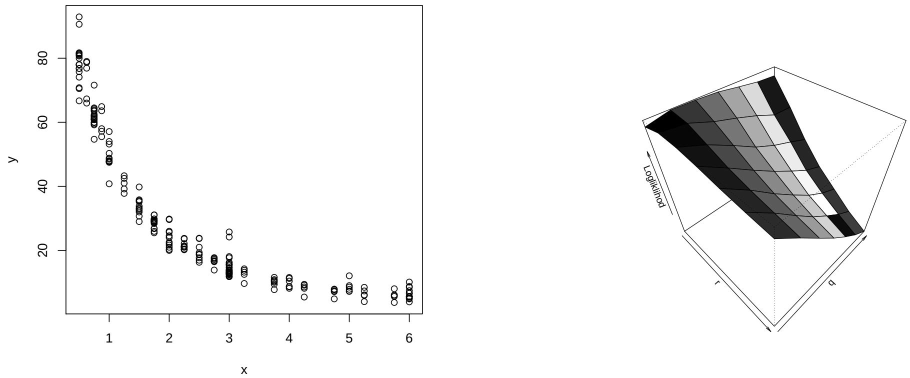

Figure 2: Ultrasonic calibration data. The left plot is the scatter plot of the data set and the right plot is the profile log-likelihood of both parameters r and q for fitting a SLST-NLM.

the least square method. These initial values are given by $\beta_1 = 0.148$, $\beta_2 = 0.005$, $\beta_3 = 0.12$, $\sigma^2 = 3.59$ and $\lambda = 0.86$. Then, we determine the values of r and q by maximizing the likelihood function which is plotted on the right side of Figure 2. For the SLST model, we found $r = 3$ and $q = 4$ and for the SLSN model, we found $q = 3$. Table 2 contains the ML estimates of the parameters of the three models, together with their corresponding standard errors calculated via the observed information matrix. For comparing the models, we also computed the Akaike Information Criterion (Akaike, 1974), ($AIC = 2p - 2\ell(\hat{\Theta})$), the Bayesian Information Criterion (Schwarz, 1978), ($BIC = p\ln(n) - 2\ell(\hat{\Theta})$) and the Efficient Determination Criterion (Bai *et al.*, 1989), ($EDC = 0.2p\sqrt{n} - 2\ell(\hat{\Theta})$), where p is the number of free parameters, n is the number of observations and $\ell(\hat{\Theta})$ is the log likelihood function.

These model selection criterions indicate that the SLST-NLM has a better fit than the SN-NLM and the SLSN-NLM for this data set. The standard errors of regression model parameters in the SLST-NLM are smaller than those in the SN-NLM and the SLSN-NLM. This suggests that the SLST-NLM seems to produce more accurate estimates than the SN-NLM and the SLSN-NLM . The estimates for the variance components (σ^2 and λ) are not comparable since they are on different scales (see Table 2).

In order to detect incorrect specification of the error distribution and systematic component in (5.2), we present Q-Q plots and simulated envelopes for the Pearson

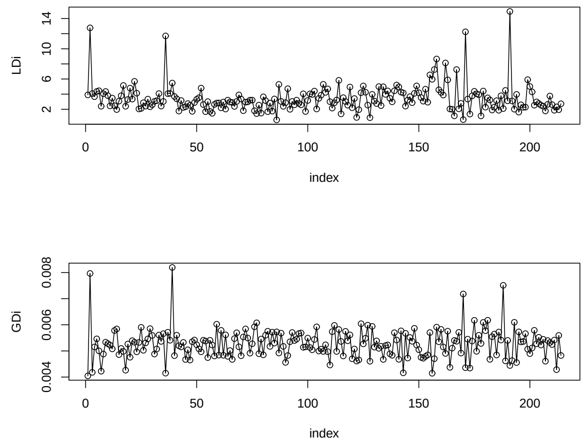

Figure 3: Plot of LD_i and GD_i for case weights perturbation for ultrasonic calibration data.

residuals (Garay *et al.*, 2011) in Figure 6. This figure clearly indicates that the SLST-NLM is more suitable for modeling this data than the SN-NLM and SLSN-NLM, since there are no observations falling outside the envelope. Moreover, there is a clear evidence of lack of fit for the SN-NLM and SLSN-NLM. Therefore, we proceed with our analysis using asymmetric models.

5.2.1 Sensitivity Analysis

In this section, we use the real data set to find the observations which are influential in parameter estimation. Let $\hat{\Theta}$ be the ML estimate of Θ in ultrasonic calibration data and $\hat{\Theta}_{(i)}$ be the ML estimate of Θ without the i-th observation. We compute the LD_i and GD_i as diagnostics for global influence. The measures LD_i and GD_i are computed and presented in Figure 3, respectively. From these figures, we observe that the cases 2, 36, 167 and 187 are influential.

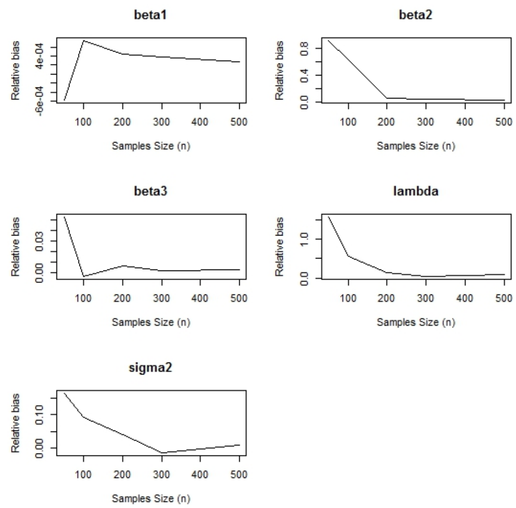

Figure 4: Relative bias of ECME estimates of $\beta_1, \beta_2, \beta_3, \lambda$ and σ^2 for SLST-NLM following the nonlinear growth curve model in (5.1).

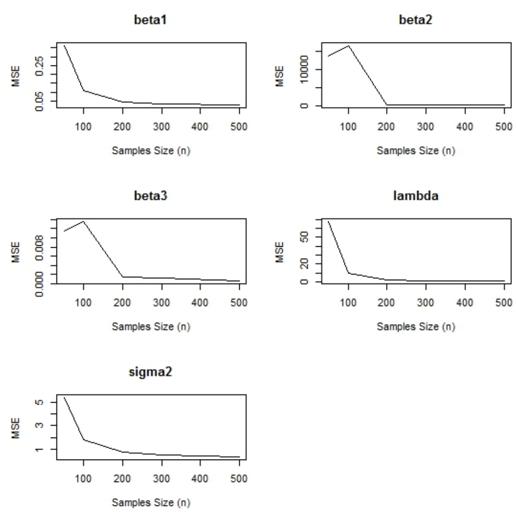

Figure 5: MSE of ECME estimates of β_1, β_2, β_3, λ and σ^2 for SLST-NLM following the nonlinear regression model.

Table 2: ML estimation results for fitting the SN-NLM, the SLSN-NLM and the SLST-NLM models to the ultrasonic calibration data.

parameter	SN-NLM		SLSN-NLM		SLST-NLM	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
β_1	0.177485	$9.327354e^{-4}$	0.186675	$1.93499e^{-5}$	0.192803	$8.690266e^{-7}$
β_2	0.006322	$1.087805e^{-5}$	0.006343	$2.715e^{-6}$	0.006357	$3.121795e^{-5}$
β_3	0.010614	$9.327354e^{-4}$	0.010377	$2.40445e^{-5}$	0.010220	$6.737256e^{-5}$
σ^2	20.86891	$1.974071e^{-3}$	13.386	$2.191338e^{-3}$	2.893346	$2.689821e^{-2}$
λ	1.780925	$3.066238e^{-5}$	1.103117	$5.33945e^{-4}$	0.651245	$6.666642e^{-3}$
r	-	-	4	-	4	-
q	-	-	-	-	3	-
AIC	1121.3098	-	1094.3464	-	1079.7192	-
BIC	1138.3968	-	1117.9082	-	1103.2810	-
EDC	1125.9385	-	1100.8266	-	1086.1994	-

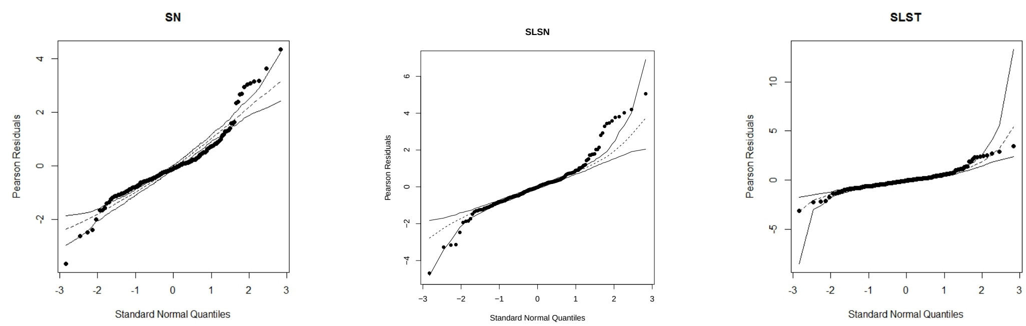

Figure 6: Q-Q plots and simulated envelopes for the Pearson residuals in the ultrasonic calibration data.

6 Conclusions

In this paper, we have proposed the application of a new asymmetric distribution, called slash skew elliptical distributions (especially slash skew-t distribution) to nonlinear regression models. We used the EM-type algorithm to obtain the maximum likelihood estimates, and applied the asymptotic method to compute the observation information matrix.

Simulation studies indicate that the method based on the SLST-NLM is more robust against outliers than the SN-NLM and SLSN-NLM. Furthermore, influential observations and the ML estimates based on the EM-type algorithm have consistency properties. The application on a data set showed that the SLST-NLM fit on real data significantly better than the SN-NLM and SLSN-NLM models.

References

Akaike, H. (1974), A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, **19** (16), 716-723.

Alcantara, I. C. and Cysneiros, F. J. A. (2017), Slash-elliptical nonlinear regression model. *Brazilian Journal of Probability and Statistics*, **31** (1), 87-110.

Atkinson, A. C. (1981), Two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13-20.

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

Azzalini, A. and Capitanio, A. (2003), Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B*, **65** (2), 367-389.

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

Bai, Z. D., Krishnaiah, P. R. and Zaho, L. C. (1989), On rates of convergence of efficient detection criteria in signal processing with white noise. *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, **35** (2), 380-388.

Basford, K. E., Greenway, D. R., McLachlan, G. J. and Peel, D. (1997), Standard errors of fitted component means of normal mixtures. *Computational Statistics*, **12**, 1-17.

Basso, R. M., Lachos, V. H., Cabral, C. R. B. and Ghosh, P. (2010), Robust mixtures of skew -normal distributions. *Computational Statistics and Data Analysis*, **54**, 2926-2941.

Berkane, M., Kano, Y. and Bentler, P. M. (1994), Pseudo maximum likelihood estimation in elliptical theory: effects of misspecification. *Computational Statistics and Data Analysis*, **18**, 255-267.

Branco, M. D. and Dey, D. K. (2001), A general case of multivariate skew - elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99-113.

Cancho, V. G., Lachos, V. H. and Ortega, E. M. M. (2010), A nonlinear regression model with skew-normal errors. *Statistical Papers*, **51**, 547-558.

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

Chwirut, D. (1979), Ultrasonic Reference Block Study. NIST. 1979 [Internet] Available from://www.itl.nist.gov/div898/strd/nls/data/chwirut2.shtml.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B,* **39**, 1-38.

Farnoosh, R., Nematollahi, N., Rahnamaei, Z. and Hajrajabi, A. (2013), A family of skew-slash distributions and estimation of its parameters via an EM Algorithm. *Journal of the Iranian Statistical Society,* **12** (2), 271-291.

Garay, A. M., Lachos, V. H. and Abanto-Valle C. A. (2011), Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society*, **40**, 115-124.

Gomez, H. W. and Venegas, O. (2008), Erratum to: A new family of slash-distribution with elliptical contours. *Statistics and Probability Letters,* **78**, 2273-2274.

Henze, N. (1986), A probabilistic representation of the skew-normal distributio. *Scandinavian Journal of Statistics,* **13**, 271-275.

Jones M. C. and Faddy M. J. (2003), A skew extension of the t distribution with applications. *Journal of the Royal Statistical Society. Series B,* **65**, 159-174.

Lachos, V. H., Bandyopadhyay, D. and Garay, A. M. (2011), Heteroscedastic nonlinear regression model based on scale mixtures of skew-normal distributions. *Statistics and Probability Letters,* **81**, 1208-1217.

Lango K. L., Little, R. and Taylor, J. (1989), Heteroscedastic nonlinear regression model based on scale mixtures of skew-normal distributions. *Journal of the American Statistical Association*, **84**, 881-896.

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

Lin, T. I., Lee, J. C. and Yen, S. Y. (2007), Finite mixture modeling using the skew normal distribution. *Statistica Sinica,* **17** (3), 909-927.

Meng, X. L. and Rubin, D. B. (1993), Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika,* **80**, 267-278.

Pirzadeh Nahooji, S., Farnoosh, R. and Nematollahi, N. (2015), Bayesian Nonlinear Regression Models based on Slash Skew-t Distribution. *European Online Journal of Natural and Social Sciences,* **4** (1), 230-240.

Rahnamaei, Z., Nematollahi, N. and Farnoosh, R. (2012), The loction-scale mixture of exponential power distribution: A Bayesian and Maximum Likelihood Approach. *Journal of Applied Mathematics,* Article ID 978282.

Sahu, S. K., Dey, D. K. and Branco, M. (2003), A new class of multivariate distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics,* **29**, 129-150.

Schwarz, G. (1978), Estimating the dimension of a model. *The Annals of Statistics,* **6**, 461-464.

Vanegas, L. H. and Cysneiros, F. J. A. (2010), Assesment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics and Data Analysis,* **54**, 1002-1016.

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

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