

## Bayesian Nonlinear Regression Models based on Slash Skew-t Distribution

Siavash Pirzadeh Nahooji<sup>1</sup>, Rahman Farnoosh<sup>2\*</sup>, Nader Nematollahi<sup>3</sup>

<sup>1</sup>Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran;

<sup>2</sup>School of Mathematics, Iran University of Science and Technology, Narmak, Tehran, Iran;

<sup>3</sup>Department of Statistics, Allameh Tabataba'i University, Tehran, Iran.

\*E-mail: [rfarnoosh@iust.ac.ir](mailto:rfarnoosh@iust.ac.ir).

### Abstract

This paper considers the Bayesian analysis for estimating the parameters of nonlinear regression model when the error term has a slash skew-t distribution. This model is an asymmetric nonlinear regression model which is suitable for fitting the data sets with heavy tail and skewness. The properties of this model are derived and a hierarchical representation of this model based on the stochastic representation of slash skew-t distribution is given. This representation allows us to use Markov Chain Monte Carlo method to estimate the parameters of model. To compare this model with other asymmetric nonlinear regression models, we use conditional predictive ordinate statistic and deviance information, expected Akaike information and expected Bayesian information criteria, and show the performance of the proposed model by a simulation study. Also an application of the new model to fitting a real data set is discussed.

**Keywords:** Bayesian analysis, Nonlinear regression models, Slash skew-t distribution, Skew slash distribution, Skew-t distribution.

### Introduction

Nonlinear regression model usually applied when response variable is a nonlinear function of explanatory variable. One of the fundamental assumptions of these models is the normality of the error terms. But this assumption does not hold when the data under study have skewness and/or heavy tail. For solving this problem some authors use heavy tail distributions for error terms such as Student-t, logistic and exponential power family which are symmetric and have heavy tail than the normal distribution. See for example Cysneiros and Vanegas (2008) and Cysneiros et al. (2010) in this regard. Some authors used asymmetrical distribution such as skew-normal distribution for error terms to overcome the skewness property of the data sets, for example see Cancho et al. (2010) and Xie et al. (2009 a, b). Montenegro et al. (2009) have shown that the estimation of the parameters of a nonlinear regression model with skew-normal random error can be sensitive to outliers. To solve the effect of outliers, Garay et al. (2011) and Lachos et al. (2010) considered the scale mixture of skew-normal distribution for random errors. They found the ML estimates of the model parameters and showed that these estimates did not sensitive to outlier observations when the distributions of random errors are skew-t or skew-slash distribution. Also, Cancho et al. (2011) developed a Bayesian estimation of the parameters of a nonlinear regression model with scale mixture of skew-normal distribution for its error term.

Recently Farnoosh et al. (2013) introduced a new slash skew-elliptical distribution which contains many of skew and heavy tail distributions such as slash skew-t distribution. They showed that slash skew-t distribution is more skewed and heavy tail than skew-normal and skew-t distribution. In this paper we use slash skew-t distribution as the distribution of the random error of a nonlinear regression model. We use the stochastic representation of slash skew-t to find a hierarchical representation of the proposed model and find the Bayes estimate of the model

parameter by employing a Markov Chain Monte Carlo (MCMC) method. Also by using a simulation study, we show the performance of the proposed model to other existing models by applying Conditional Predictive Ordinate (CPO) statistics, Deviance Information Criterion (DIC), Expected Akaike Information Criterion (EAIC) and Expected Bayesian Information Criterion (EBIC).

This paper has been arranged as follows. In Section 2, we will briefly introduce slash skew-t distribution and state its important properties. In Section 3 we give a rough idea of nonlinear regression model with random errors that have slash skew-t distribution. In Section 4, we will describe Bayesian method to estimate model parameters and introduce some measures of model selection which we will use to compare the models. In Section 5, for comparing the proposed model to other existing models, we conduct a simulation study and apply a real example. A discussion is given in Section 6.

### Slash skew-t distribution

In this section we discuss some properties of slash skew-t distribution which has been introduced by Farnoosh et al. (2013). A random variable  $Y$  has slash skew-t (SLST) distribution if it can be written as

$$Y = \mu + \sigma \frac{X}{U^{\frac{1}{q}}}, \quad (1)$$

where  $\mu$  is the location parameter,  $\sigma$  is a scale parameter,  $q > 0$  is the tail parameter and  $X$  is a skew-t random variable with location 0, scale 1, skewness  $\lambda$ , degree of freedom  $r$

( $X \sim ST(0,1,\lambda,r)$ ) and  $U$  is a uniform random variable ( $U \sim U(0,1)$ ),  $X$  and  $U$  are independent. Thus, the probability density function (pdf) of  $Y$  is given by

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

where  $h(t) = \frac{t}{|t|}$  and  $g(t) = \frac{\Gamma(\frac{(1+r)}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}} (1 + \frac{t^2}{r})^{-\frac{(1+r)}{2}}$  is the density generator function of

Student-t distribution and  $F_g(\cdot)$  is cumulative distribution function (cdf) of Student-t distribution.

We denote it by  $Y \sim SLST(\mu, \sigma^2, \lambda, r, q)$ .

Now, we demonstrate the subsequent results of Slash skew-t distribution. First, we present the stochastic representation of SLST. This representation will be employed to simulate random variable  $Y$ . The indicated representation will also be employed to transact Gibbs sample algorithm for hierarchical representation of random variable  $Y$ . For a proof of the following theorems see Farnoosh et al. (2013).

**Theorem 2.1** Let  $Y \sim SLST(\mu, \sigma, \lambda, q, r)$ , then  $Y$  has the following stochastic representation

$$\begin{aligned} Y &= \mu + \sigma \delta U^{\frac{-1}{q}} V^{\frac{-1}{2}} |T_0| + \sigma (1 - \delta^2)^{\frac{1}{2}} U^{\frac{-1}{q}} V^{\frac{-1}{2}} T_1 \\ &= \mu + \xi T + \Lambda^{\frac{1}{2}} U^{\frac{-1}{q}} V^{\frac{-1}{2}} T_1, \end{aligned} \quad (3)$$

where  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ ,  $\xi = \sigma\delta$ ,  $\Lambda = \sigma^2(1-\delta^2)^2$ ,  $V \sim \frac{1}{r}\chi_r^2$ ,  $U \sim U(0,1)$ ,  $T = U^{\frac{-1}{q}}V^{\frac{-1}{2}}|T_0|$  and  $T_0$  and  $T_1$  are independent standard normal random variables.

**Theorem 2.2** Let  $Y \sim SLST(\mu, \sigma^2, \lambda, r, q)$

- 1) If  $q > 1, r > 1$ , then  $E(Y) = \mu + c\delta\sigma$ ;
- 2) If  $q > 2, r > 2$ , then  $Var(Y) = \frac{q}{q-2} \frac{r}{r-2} \sigma^2 - (c\delta\sigma)^2$ ;

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

**The slash skew-t nonlinear regression model**

In this section, we assume that the random error of a nonlinear regression model is distributed as SLST distribution. So, our nonlinear regression is given by

$$Y_i = \psi(\boldsymbol{\beta}, x_i) + \varepsilon_i, \quad i = 1, \dots, n \tag{4}$$

where  $Y_i$  are responses,  $\psi(\cdot)$  is an injective and twice continuously differentiable function with respect to the parameter vector  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ ,  $\mathbf{x}_i$  is a vector of explanatory variable values and

the random errors  $\varepsilon_i \sim SLST_1(-c\delta\sigma, \sigma^2, \lambda, r, q)$  for  $q > 2, r > 2$  and  $c = \frac{q}{q-1} \frac{\Gamma(\frac{r-1}{2})}{\Gamma(\frac{r}{2})} \sqrt{\frac{r}{\pi}}$ , which

corresponds to regression model where the error distribution has mean zero. From Theorem 2.2, the following result can be deduced

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

So, we have

$$Y_i \sim SLST_1(\psi(\boldsymbol{\beta}, x_i) - c\delta\sigma, \sigma, \lambda, q, r) \text{ for } i = 1, \dots, n.$$

From (2) and (4) the log-likelihood function for  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \lambda, r, q)^T$ , given the observed sample  $\mathbf{y} = (y_1, \dots, y_n)^T$  is given by

$$\ell(\boldsymbol{\theta}) = n \log q - nq \log \sigma + n \log \Gamma\left(\frac{1+r}{2}\right) - n \log \Gamma\left(\frac{1+r}{2}\right) - \frac{n}{2} \log(r\pi) - \sum_{i=1}^n \log(S_i) \tag{5}$$

$$\text{where } S_i = \frac{1}{|y_i - \mu|^{q+1}} \int_0^{\frac{(y-\mu)^2}{\sigma^2}} u^{\frac{q-1}{2}} \left(1 + \frac{u}{r}\right)^{-\left(\frac{1+r}{2}\right)} F_g(\lambda\sqrt{uh}(y-\mu)) du.$$

In this paper we will use the Bayesian method to estimate model parameters. This method has been built based on MCMC algorithms that would achieve posterior inference for the parameters.

### Bayesian model parameter estimation

In this section we use Bayesian analysis to estimate the parameters of the nonlinear regression model with SLST random error (SLST-NLM). We use MCMC techniques. From theorem 2.1 we can obtain the following hierarchical representation which helps us to write BUGS cods simply.

$$\begin{aligned} Y_i | U_i = u_i, V_i = v_i, T_i = t_i &\sim N(\psi(\boldsymbol{\beta}, x_i) + \xi t_i, u_i^q v_i^{-1} \Lambda), \\ T_i | U_i = u_i, V_i = v_i &\sim TN(-c, u_i^q v_i^{-1}) I(-c, \infty), \\ U_i &\sim U(0, 1), \\ V_i &\sim \frac{1}{r} \chi_r^2 = \text{gamma}\left(\frac{r}{2}, \frac{r}{2}\right), \end{aligned} \quad (6)$$

where  $TN(a, b^2) I(r, s)$  denote the truncated normal distribution  $N(a, b^2)$  on  $(r, s)$ .

Let  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $\mathbf{x} = (x_1, \dots, x_n)^T$ ,  $\mathbf{t} = (t_1, \dots, t_n)^T$ ,  $\mathbf{u} = (u_1, \dots, u_n)^T$  and  $\mathbf{v} = (v_1, \dots, v_n)^T$ , so the complete likelihood function of  $\boldsymbol{\theta}$  based on  $(\mathbf{y}, \mathbf{x}, \mathbf{t}, \mathbf{u})$  is given by

$$\begin{aligned} L(\boldsymbol{\theta} | y, x, t, u, v) &\propto \prod_{i=1}^n [\phi_1(y_i; \psi(\boldsymbol{\beta}, x_i) + \xi t_i, u_i^q v_i^{-1} \Lambda) \\ &\quad \times \phi_1(t_i; -c, u_i^q v_i^{-1}) I(-c, \infty) U(u_i; 0, 1) \frac{\chi^2(v_i, r)}{r}] \end{aligned}$$

To conduct a Bayesian analysis, we need to determine the prior distribution for all unknown parameters  $\boldsymbol{\beta}$ ,  $\lambda$ ,  $\sigma^2$ ,  $r$  and  $q$ . Because there is no previous information about unknown parameters, we consider conjugate prior distributions to these parameters. For the sake of avoiding improper posterior, we select proper priors with known hyper-parameters. Therefore, we choose a normal prior distribution for the elements of  $\boldsymbol{\beta}$  with pdf given by  $\pi(\beta_j)$ . For parameter  $\xi$ , normal prior distribution with mean  $\mu_\xi$  and variance  $\sigma_\xi^2$  has been selected. The inverse Gamma prior distribution is considered for  $\Lambda$ , i.e.,

$$\Lambda \sim \text{IGamma}\left(\frac{\rho}{2}, \frac{\delta}{2}\right).$$

For degrees of freedom parameter  $r$  and tail parameter  $q$  we consider truncated exponential prior distributions on the interval  $(2, \infty)$  which are given by

$$r \sim \exp\left(\frac{\zeta}{2}\right) I_{(2, \infty)}, \quad q \sim \exp\left(\frac{\zeta}{2}\right) I(2, \infty).$$

These truncation points were chosen to persuade that the variance of  $y_i$  be finite.

We assume that prior distributions are independent. Therefore, the complete prior can be written as follows

$$\pi(\boldsymbol{\theta}) \propto \left( \prod_{j=1}^p \pi(\beta_j; \mu_{\beta_j}, \sigma_{\beta_j}^2) \right) \pi(\xi; \mu_\xi, \sigma_\xi^2) \pi(\Lambda | \rho, \delta) \pi(q) \pi(r). \quad (7)$$

By mixing the likelihood function and the prior distribution (7), the joint posterior can be achieved as follows

$$\pi(\boldsymbol{\theta}, t, u, v | y, x) \propto \prod_{i=1}^n [\phi_1(y_i; \boldsymbol{\psi}(\boldsymbol{\beta}, x_i) + \xi t_i, u_i^{2/q} v_i^{-1} \Lambda) \phi_1(t_i; -c, u_i^{2/q} v_i^{-1})] \\ \times I(-c, \infty) U(u_i; 0, 1) \frac{\chi^2(v_i, r)}{r} \pi(\boldsymbol{\theta}).$$

For inferring and estimating the parameter of the above distribution, MCMC method such as Gibbs sampler can be used. The full conditional distributions which need the Gibbs sampler can be written as follows

$$T_i | \boldsymbol{\beta}, \xi, \Lambda, r, q, \mathbf{y}, \mathbf{u}, \mathbf{v} \sim TN_1(\mu_{T_i} - c, u_i^{2/q} v_i^{-1} M_T^2) I(-c, \infty), \quad i=1, \dots, n;$$

$$\text{where } \mu_{T_i} = \frac{\xi}{\Lambda + \xi^2} (y_i - \boldsymbol{\psi}(\boldsymbol{\beta}, x_i) + c\xi), \quad M_T^2 = \frac{\Lambda}{\Lambda + \xi^2},$$

$$\xi | \boldsymbol{\beta}, \Lambda, r, q, \mathbf{y}, \mathbf{u}, \mathbf{v}, t \sim N_1\left(\frac{K\sigma_\xi^2 + \Lambda\mu_\xi}{H\sigma_\xi^2 + \Lambda}, \frac{H\sigma_\xi^2 + \Lambda}{\Lambda\sigma_\xi^2}\right),$$

$$\text{where } K = \sum_{i=1}^n u_i^{2/q} v_i t_i (y_i - \boldsymbol{\psi}(\boldsymbol{\beta}, x_i)), \quad H = \sum_{i=1}^n u_i^{2/q} v_i t_i,$$

$$\Lambda | \boldsymbol{\beta}, \xi, r, q, \mathbf{y}, \mathbf{u}, \mathbf{v}, t \sim IGamma\left(\frac{1}{2}(\rho + n), \frac{1}{2}(\delta + \sum_{i=1}^n u_i^{2/q} v_i (y_i - \boldsymbol{\psi}(\boldsymbol{\beta}, x_i) + c\xi)^2 n)\right)$$

$$\boldsymbol{\beta} | \Lambda, \xi, r, q, \mathbf{y}, \mathbf{u}, \mathbf{v}, t \propto \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}_\beta, \mathbf{D}) \prod_{i=1}^n \phi_1(\boldsymbol{\psi}(\boldsymbol{\beta}, x_i); (y_i - \xi t_i), u_i^{2/q} v_i^{-1} \Lambda)$$

$$\text{where } \boldsymbol{\mu}_\beta = (\mu_{\beta_1}, \dots, \mu_{\beta_p})^T, \quad \mathbf{D} = \text{diagonal}(\sigma_{\beta_1}^2, \dots, \sigma_{\beta_p}^2).$$

For generating samples from  $\boldsymbol{\beta}$  we use Metropolis-Hasting algorithm as follows

- a) We choose initial value for  $\boldsymbol{\beta}_0$  at the first step
- b) At the  $(r+1)$ th step  $\boldsymbol{\beta}^*$  and  $U$  are generated from  $N_p(\boldsymbol{\beta}^{(r)}, \mathbf{S})$  and  $U(0,1)$  distributions, respectively. Then  $\boldsymbol{\beta}^{(r+1)}$  is given by

$$\boldsymbol{\beta}^{(r+1)} = \begin{cases} \boldsymbol{\beta}^* & \text{if } U \leq \alpha(\boldsymbol{\beta}^*, \boldsymbol{\beta}^{(r)}) \\ \boldsymbol{\beta}^{(r)} & \text{otherwise} \end{cases}$$

where

$$\alpha(\boldsymbol{\beta}^*, \boldsymbol{\beta}^{(r)}) = \frac{\phi_p(\boldsymbol{\beta}^*; \boldsymbol{\mu}_\beta, \mathbf{D}) \prod_{i=1}^n \phi_1(\boldsymbol{\psi}(\boldsymbol{\beta}^*, x_i); (y_i - \xi^{(r)} t_i^{(r)}), u_i^{(r)2/q} v_i^{(r)-1} \Lambda^{(r)})}{\phi_p(\boldsymbol{\beta}^{(r)}; \boldsymbol{\mu}_\beta, \mathbf{D}) \prod_{i=1}^n \phi_1(\boldsymbol{\psi}(\boldsymbol{\beta}^{(r)}, x_i); (y_i - \xi^{(r)} t_i^{(r)}), u_i^{(r)2/q} v_i^{(r)-1} \Lambda^{(r)})}.$$

The matrix  $\mathbf{S}$  is selected after generating and initial MCMC chain which would help to achieve a good convergence. Similar technique used to update  $\xi$  and  $\Lambda$ .

For completing Gibbs sampler scheme, we need the full conditional posterior distributions of  $\mathbf{u}, \mathbf{v}, r$  and  $q$ . For each element of  $\mathbf{u}$ , the pdf is:

$$\pi(u_i | \boldsymbol{\beta}, \xi, \Lambda, r, q, \mathbf{y}, \mathbf{v}, \mathbf{t}) \propto u_i^{\frac{2}{q}} \exp \left( -\frac{\frac{2}{q} u_i v_i}{2} \left( \frac{(y_i - \psi(\boldsymbol{\beta}, x_i) - \xi t_i)^2}{\Lambda} + (t_i + c)^2 \right) \right) \mathbf{I}(0, 1) \quad (8)$$

for  $i = 1, \dots, n$ .

For each element of  $\mathbf{v}$ , the pdf is:

$$\pi(v_i | \boldsymbol{\beta}, \xi, \Lambda, r, q, \mathbf{y}, \mathbf{u}, \mathbf{t}) \propto v_i^{\frac{r}{2}} \exp \left[ -\frac{v_i}{2} \left( u_i^{\frac{2}{q}} \left( \frac{(y_i - \psi(\boldsymbol{\beta}, x_i) - \xi t_i)^2}{\Lambda} + (t_i + c)^2 \right) + r \right) \right] \quad (9)$$

for  $i = 1, \dots, n$ . So we can suppose that  $v_i | \boldsymbol{\beta}, \xi, \Lambda, r, q, \mathbf{y}, \mathbf{u}, \mathbf{t}$  has  $\text{Gamma}(\frac{r}{2} + 1, w_i)$ , where

$$w_i = \frac{v_i}{2} \left( u_i^{\frac{2}{q}} \left( \frac{(y_i - \psi(\boldsymbol{\beta}, x_i) - \xi t_i)^2}{\Lambda} + (t_i + c)^2 \right) + r \right)$$

for  $i = 1, \dots, n$ .

The conditional posterior density of  $r$  is:

$$\pi(r | \boldsymbol{\beta}, \xi, \Lambda, q, \mathbf{y}, \mathbf{u}, \mathbf{v}, \mathbf{t}) \propto (n/2)^{nr/2} (\Gamma(r/2))^{-n} \times \exp \left( -\frac{r}{2} \left[ \sum_{i=1}^n (v_i - \ln v_i) + \zeta \right] \right) \mathbf{I}(2, \infty). \quad (10)$$

The conditional posterior density of  $q$  is:

$$\pi(q | \boldsymbol{\beta}, \xi, \Lambda, r, \mathbf{y}, \mathbf{u}, \mathbf{v}, \mathbf{t}) \propto u_i^{2n/q} \exp \left( -\frac{1}{2} \left[ \sum_{i=1}^n u_i^{2/q} v_i z_i + \zeta q \right] \right) \mathbf{I}(2, \infty), \quad (11)$$

where  $z_i = \left( \frac{(y_i - \psi(\boldsymbol{\beta}, x_i) - \xi t_i)^2}{\Lambda} + (t_i + c)^2 \right)$ .

Note that (8), (9), (10), (11) does not have a closed form, but a Metropolis-Hasting algorithm can be embedded in the MCMC scheme to obtain draws of  $u_i$ ,  $v_i$ ,  $r$  and  $q$ .

Finally, since  $(\xi, \Lambda)$  has a one to one correspondence to  $(\sigma^2, \lambda)$ , then  $\sigma^2$  and  $\lambda$  can be calculate  $\sigma^2 = \xi^2 + \Lambda$  and  $\lambda = \xi / \sqrt{\Lambda}$ .

### Model selection method

For comparing several models and finding the best model that fits to the given data, there are several criteria for selecting the best model. In this section we introduce some of these criteria to compare SLST-NLM to other existing models.

The first criterion is based on Conditional Predictive Ordinate (CPO) statistic which is one of the applicable model comparison criteria and recommended by Gelfand et al. (1992). Let  $\Delta$  be the full data and  $\Delta^{(-i)}$  denote the data without  $i$  observation.

In SLST-NLM model for an observed data, we have

$$g(y_i | \theta) = \frac{q\sigma^q}{|y_i - \mu_i|^{q+1}} \int_0^{\frac{(y_i - \mu_i)^2}{\sigma^2}} u^{\frac{q-1}{2}} g(u) F_g(\lambda \sqrt{u} h(y_i - \mu_i)) du,$$

where  $g(t) = \frac{\Gamma(\frac{1+r}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}}(1 + \frac{t}{r})^{-\frac{1+r}{2}}$  is the density generator function of Student-t distribution,  $F_g(\cdot)$  is cdf of Student-t distribution and  $\mu_i = \psi(\beta, x_i) - c\delta\sigma$ . The posterior density of  $\theta$  given  $\Delta^{(-i)}$  is denoted by  $\pi(\theta | \Delta^{(-i)})$ . For the  $i$  observation, the  $CPO_i$  is given by

$$CPO_i = \int_{\Theta} g(y_i | \theta)\pi(\theta | \Delta^{(-i)})d\theta = \left\{ \int_{\Theta} \frac{\pi(\theta | \Delta)}{g(y_i | \theta)} d\theta \right\}^{-1}, \quad i = 1, \dots, n.$$

In SLST model we could not find a closed form for  $CPO_i$  statistics. Therefore, a Monte Carlo estimate of  $CPO_i$  can be obtained by using a MCMC sample from the posterior distribution  $\pi(\theta | \Delta)$ . Let  $\theta_1, \dots, \theta_Q$  be a sample of size  $Q$  of  $\pi(\theta | \Delta)$  after the burn in. A Monte Carlo estimate of  $CPO_i$  (Chen et al., 2000) is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{t=1}^Q \frac{1}{g(y_i | \theta_t)} \right\}^{-1}.$$

Now a statistic based on  $CPO_i$ 's is  $B = \sum_{i=1}^n \log(CPO_i)$ . Since large value of  $CPO_i$  show a better fit of the model, then large value of  $B$  imply a better fit of the model too.

We use other criteria, including the Deviance Information Criterion (DIC) introduced by Spiegelhart et al. (2002), the Expected Akaike Information Criterion (EAIC) (Brooks, 2002) and Expected Bayesian Information Criterion (EBIC) (Carlin and Louis, 2001). These criteria can be approximated by the MCMC output as follows: estimate of DIC is  $\widehat{DIC} = 2\bar{D} - \widehat{D}$ , where  $D(\theta) = -2 \sum_{i=1}^n \log[g(y_i | \theta)]$ ,  $\bar{D} = \sum_{t=1}^Q \frac{D(\theta_t)}{Q}$  and  $\widehat{D}$  can be estimated by

$$\widehat{D} = D\left(\frac{1}{Q} \sum_{t=1}^Q \beta_t, \frac{1}{Q} \sum_{t=1}^Q \sigma_t^2, \frac{1}{Q} \sum_{t=1}^Q \lambda_t, \frac{1}{Q} \sum_{t=1}^Q r_t, \frac{1}{Q} \sum_{t=1}^Q a_t\right).$$

The EAIC and EBIC can be estimated by means of  $\widehat{EAIC} = \bar{D} + 2\#(l)$  and  $\widehat{EBIC} = \bar{D} + 2\#(l)\log(n)$ , where  $\#(l)$  is the number of model parameter.

### Simulation and real data analyses

In this section, we investigate properties of SLST-NLM model by a simulation study and the performance of this model with other models such as nonlinear regression model based on skew-t random error (ST-NLM) and nonlinear regression model based on skew-slash random error (SSL-NLM) introduced by Cancho et al. (2011). Also an application to a real data set is discussed.

#### Frequentist properties

At first by using simulated SLST-NLM data we study the frequentist properties of the parameter estimates. In this simulation we study the behavior of Bayesian estimates, based on frequentist Mean Squared Error (MSE) and the frequentist Mean. We implement the simulation study by the following logistic model:

$$Y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \varepsilon_i, \quad i = 1, \dots, 50, \tag{13}$$

where the variable  $x_i$  has a range between 1 to 50. For comparing our proposed model with other models, we use model parameters that have been applied by Cancho et al. (2011), thus we consider the values,  $\beta_1 = 30, \beta_2 = 5, \beta_3 = 0.7, \sigma^2 = 2, \lambda = -4, q = 3$  and  $r = 4$ . The error terms are



independent with  $\varepsilon_i \sim SLST(-c\xi, \sigma^2, \lambda, r, q)$ . In this simulation we consider the number of iterations equal to 100 and the fitted nonlinear regression model with three different distribution for random errors  $\varepsilon_i$ . These random errors are skew-t, skew slash and slash skew-t.

For implement the Gibbs sampler the following independent prior has been considered. For  $i = 1, 2, 3$ ,  $\beta_i \sim N_1(0, 1000)$ ,  $\xi \sim N(0, 1000)$ ,  $\Lambda \sim IG(0.01, 0.01)$ ,  $r \sim \exp(0.1)I(2, \infty)$  and  $q \sim \exp(0.1)I(2, \infty)$ .

For each generated data sets we simulate two chain of size 60000 for each parameter, for burning in, we exclude the first 10000 iterations. Also for decreasing autocorrelation, we consider thin size 10, hence the effective sample size are 5000. For each sample the posterior mean of the parameter and the DIC, EAIC, BIC and B are calculated.

The result of simulation study for ST, SSL and SLST random error distributions with estimated MSE and Mean are given in Table 1. In this table the true values have been put in to the parenthesis, MC Mean has been calculated according to  $\sum_{j=1}^{100} \hat{\theta}_{kj} / 100$  and MC MSE which has been determined by  $\sum_{j=1}^{100} (\hat{\theta}_{kj} - \theta_k)^2 / 100$  is the empirical mean squared error. In the indicated formula  $\hat{\theta}_k$  is the estimated parameter of true parameter  $\theta_k$ .

The values in Table 1 indicated that the MC MSE for model parameters are smaller in SLST model than ST and SSL models. Thus SLST-NLM would be better fitted rather than ST-NLM and SSL-NLM for nonsymmetrical data.

This statement has been approved by the values in Table 2. In this table the arithmetic average of the measurements for comparing the models (DIC, EAIC, EBIC, B) are given. These measurements would be smaller in SLST-NLM than the mentioned models. Thus, this particular model has better fitted than the competitor models.

**Table 1 Monte Carlo results based on 100 simulated samples**

Parameter	ST-NLM		SSL-NLM		SLST-NLM	
	MC Mean	MC MSE	MC Mean	MC MSE	MC Mean	MC MSE
$\beta_1(30)$	32.431	0.470	31.006	0.328	30.342	0.237
$\beta_2(5)$	3.865	0.934	4.456	0.786	4.782	0.563
$\beta_3(0.7)$	0.611	0.069	0.658	0.064	0.669	0.062
$\sigma^2(2)$	3.034	1.072	3.179	1.142	2.431	0.826
$r(4)$	4.877	0.9282	-	-	4.224	0.653
$q(3)$	-	-	2.549	0.337	2.708	0.264
$\lambda(-4)$	-5.082	1.697	-4.676	0.784	-4.351	0.743

**Table 2 Comparison between ST-NLM, SSL-NLM and SLST-NLM fitting by using different Bayesian criteria**

Model	Criterion			
	MC DIC	MC EAIC	MC EBIC	MC B
ST	138.485	88.977	87.173	-376.067
SSL	135.955	82.918	81.114	-373.054
SLST	134.954	80.891	78.783	-369.953



**Real Data**

We use a Bayesian analysis of the ultrasonic calibration data, which has been described in Lin et al. (2007) and Lachos et. al. (2011), to investigate our method. These data are generated from National institutes of Standards and Technology (NIST) study by Chwirut (1979) involving ultrasonic calibration, where the response variable  $y$  and the predictor variable is metal distance  $x$ . The data consists of 214 observations and is available freely in the R package Nlsmn. We consider

$$Y_i = \frac{\exp(-\beta_1 x_i)}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad \varepsilon_i \stackrel{i.i.d}{\sim} SLST_1(-c\delta\sigma, \sigma, \lambda, q, r). \tag{14}$$

We will use the ST, SSL and SLST distributions to make a comparison.

We consider the following independent priors to apply the Gibbs sampler algorithm: For  $i = 1, 2, 3, \beta_i \sim N_1(0, 1000), \xi \sim N(0, 1000), \Lambda \sim IG(0.01, 0.01), r \sim \exp(0.1)I(2, \infty)$  for skew-t model and  $q \sim \exp(0.1)I(2, \infty)$  for SSL model and SLST model. By considering the above priors, we generate two parallel independent runs of the Gibbs sampler chain with size 100,000 for each parameter, for burning in, we exclude the first 10000 iterations. Also to avoid correlation problem we consider thin size 10, hence, the effective sample size are 9000. For each sample the posterior mean of the parameter and the DIC, EAIC, BIC and B are calculated.

Posterior mean and standard deviation of parameter under the distributions ST, SSL and SLST have been demonstrated in Table 3. Also, in Table 4 the values of model selection criteria are given. From this table we see that the SLST-NLM has a better fit to this data than ST-NLM and SSL-NLM.

**Table 3 Ultrasonic calibration data set. Summary result from the posterior distribution, Mean and Standard Deviation (SD) for parameter under ST, SSL and SLST distributions**

parameter	ST-NLM		SSL-NLM		SLST-NLM	
	Mean	SD	Mean	SD	Mean	SD
$\beta_1$	0.1694	0.1165	0.3735	0.0491	0.3899	0.0244
$\beta_2$	0.0059	4.134e-04	0.0091	3.794e-04	0.0073	2.707e-04
$\beta_3$	0.0112	0.00061	0.0082	0.00023	0.0076	0.00012
$\sigma^2$	4.4581	1.4371	4.6132	1.4526	2.9144	0.856
$\lambda$	0.5360	0.8524	1.2443	0.3304	1.4185	0.3243
$r$	2.7652	0.5422	-	-	2.9288	0.4962
$q$	-	-	2.1199	0.1861	2.9324	0.1215

**Table 4 Ultrasonic calibration data set. Comparison between ST-NLM, SSL-NLM and SLST-NLM by using different Bayesian criteria.**

Criterion	ST-NLM	SSL-NLM	SLST-NLM
B	-588.004	-581.554	-574.391
DIC	1319.090	1270.850	1235.340
EAIC	1256.841	1171.952	1168.962
EBIC	1258.823	1173.931	1170.273

### Conclusion

In this paper, we suggest a new nonlinear regression model based on an asymmetric distribution of random error of the model. We employ a MCMC method to obtain the Bayes estimate of the model parameters.

To show the performance of the new model to the existing models, we perform a simulation study. By using B, DIC, EAIC and EBIC criteria, we see that SLST-NLM has better fit than ST-NLM and SSL-NLM for simulated data and a real data set.

### References

- Brooks, S.P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *Journal of the Royal Statistical Society. Series B.* 64 (3):616-618.
- Cancho, V.C., Lachos, V.H., & Ortega, E.M.M. (2010). A nonlinear regression model with skew-normal errors. *Statistical Papers.* 51:547-558.
- Cancho, V. C., Dey, D. K., Lachos, V. H., & Andrade, M. G. (2011). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics and Data Analysis.* 55:588-602.
- Carlin, B. P., & Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. New York, Chapman & Hall.
- Chen M. H., Shao Q. M., & Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. New York, Springer-Verlag.
- Chwirut D. (1979). Ultrasonic Reference Block Study. Available from: <http://www.itl.nist.gov/div898/strd/nls/data/chwirut2.shtml>.
- Cysneiros, F. J. A., & Vanegas, L. H. (2008). Residuals and their statistical properties in symmetrical nonlinear models. *Statistics and Probability Letters.* 78:3269-3273.
- Cysneiros, F., Cordeiro, G. M., & Cysneiros, A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation.* 80:451-461.
- Farnoosh, R., Nematollahi, N., Rahnamaei, Z., & 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., & 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.
- Gelfand, A.E, Dey, D. K., & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods In: *Bayesian Statistics*. New York, Oxford Univ Press.
- Lachos, V.H., Ghosh, P., & Arellano-Valle, R.B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica.* 20:303-322.
- Lachos, V.H., Bandyopadhyay, D., & Garay, A.M. (2011) Heteroscedastic nonlinear regression model based on scale mixtures of skew-normal distributions. *Statistics and Probability Letters.* 81:1208-1217.
- Lin, T. I., Lee, J. C., & Hsieh, W. J. (2007). Robust mixture modeling using the skew-t distribution. *Statistics and Computing.* 17:81-92.
- Montenegro, L. C., Lachos, V., & Bolfarine, H. (2009). Local influence analysis of skew-normal linear mixed models. *Communications in Statistics. Theory and Methods.* 38:484-496.

- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B.* 64(4):583-639.
- Xie, F., Lin, J., & Wei, B. (2009a). Diagnostics for skew-normal nonlinear regression models with AR.1/ errors. *Computational Statistics and Data Analysis.* 53(12):4403-4416.
- Xie, F. C, Wei, B. C., & Lin, J. G., (2009b) Homogeneity diagnostics for skew-normal nonlinear regression models. *Statistics and Probability Letters.* 79:821-827.