Research Article
Likelihood Inference of Nonlinear Models Based on a Class of Flexible Skewed Distributions

Xuedong Chen,[1] Qianying Zeng,[2] and Qiankun Song[2]

[1] School of Science, Huzhou Teachers' College, Huzhou 313000, China
[2] School of Management, Chongqing Jiaotong University, Chongqing 400074, China

Correspondence should be addressed to Qiankun Song; qiankunsong@163.com

This paper deals with the issue of the likelihood inference for nonlinear models with a flexible skew-t-normal (FSTN) distribution, which is proposed within a general framework of flexible skew-symmetric (FSS) distributions by combining with skew-t-normal (STN) distribution. In comparison with the common skewed distributions such as skew normal (SN), and skew-t (ST) as well as scale mixtures of skew normal (SMSN), the FSTN distribution can accommodate more flexibility and robustness in the presence of skewed, heavy-tailed, especially multimodal outcomes. However, for this distribution, a usual approach of maximum likelihood estimates based on EM algorithm becomes unavailable and an alternative way is to return to the original Newton-Raphson type method. In order to improve the estimation as well as the way for confidence estimation and hypothesis test for the parameters of interest, a modified Newton-Raphson iterative algorithm is presented in this paper, based on profile likelihood for nonlinear regression models with FSTN distribution, and, then, the confidence interval and hypothesis test are also developed. Furthermore, a real example and simulation are conducted to demonstrate the usefulness and the superiority of our approach.

1. Introduction

The common assumption of distribution for random error is normal in statistical modeling. This assumption may lack the robustness against departures from normality and/or outliers and may result in misleading inferential results [1, 2]. For the past few years, there is an increasing interest in developing more flexible parametric families capable of adopting as closely as possible real data, which exhibit quite substantial nonnormal characteristics such as skewness and heavy tails. In a variety of applications, one popular option is to modify a symmetric probability density function of a variable, thereby introducing skewness. An important advantage of this sort of approach compared with other approaches to robustness is an explicit statement of the probabilistic setting, leading to a clear interpretation of the results [3]. Following this idea, the skw-normal (SN) distribution was firstly introduced by [4], and, then, the skew-t (ST) distribution was introduced by [5]; the skew-t-normal (STN) was introduced by [6]; moreover, some extensions to these multivariate cases were studied by [7, 8] and so on. Since then, several authors

have tried to extend these results to more general forms of skew-symmetric distributions, of which here we would like to mention [9], in this paper; they proposed a general framework of distributions which is called flexible skew-symmetric (FSS) distribution. As pointed out by [10] that this distribution family enjoys a sufficient flexibility in that with different choice of submodel settings, the FSS distribution includes several known distributions such as the SN and ST as its special cases.

However, in many practical applications, it is not rare at all to encounter a multimodality, sometimes with an even irregular shape, and, for this case, all the distributions mentioned above appear to be unsufficient to describe the multimodal feature of the data. A solution to this problem is to use finite mixture models. In [11], the authors worked with a mixture model with component densities belonging to the STN distribution and a computationally feasible EM-type algorithm was developed for calculating the maximum likelihood (ML) estimates of parameters. Unfortunately, although the proposed methodology is useful for analyzing multimodal asymmetric data, it suffers from the problem of

"model identification" as the number of the parameters to be estimated is usually large. As a result, in this paper, we deal with a new extension of the class of FSS distributions, which is referred to as flexible skew-t-normal (FSTN) distribution. This new distribution is proposed within the general framework of the FSS distributions in combination with the definition of STN distribution. In practical applications, it is able to regulate the density in a more flexible way to offer robustness and it can be treated as an appealing option for accommodating data with skewness and heavy tails as well as multimodality jointly.

On the other hand, nonlinear regression models are widely applied in the fields of economics, engineering, biomedical research, and so forth, where the nonlinear function of unknown parameters is used to explain or investigate the nonlinear relationship of random phenomena under study. More recently, several authors have used a class of skewed distributions in the context of nonlinear regression models, and some valuable results were obtained. For example, [12] developed the robust estimation and the local influence analysis for regression model with SMSN distribution. From Bayesian point of view, [13] considered the Bayesian estimation and the case influence diagnostics for nonlinear regression models with SMSN distributions. More related literature could be found in [14–17]. Generally speaking, for model fitting of the nonlinear regression with skewed distributions, a popular approach is to consider the hierarchical representation of variables with a specific distribution, in which the postulated distribution is expressed as several conditional distributions of simpler forms such as normal and Student's t and Gamma. Based on that, EM algorithm or Bayesian hierarchical approach then can be implemented effectively for conducting model estimation and statistical inference.

In this paper, our aim is to develop an approach to likelihood inference of nonlinear regression models with FSTN assumption. As there is no stochastic representation for FSTN distribution, all the methods cited above become unavailable for our considered problem and an alternative way is to return to the original Newton-Raphson iterative procedure for model estimation. Under the nonlinear regression paradigm, the accuracy of estimates is affected by the strength of nonlinearity and the corresponding confidence interval and hypothesis test require the assumption of normality of the estimators or distribution, which is too restrictive. Besides, considering that in many practical applications, rather than the total parameters, we are usually interested in a proper subset of them. By taking all these factors into account, in this paper we focus on the parameters of interest and propose a modified Newton-Raphson iterative algorithm for calculating the ML estimates based on profile likelihood. Furthermore, the confidence interval and hypothesis test for the parameters of interest are also considered. We conduct an application and a simulation study to compare the algorithm effectiveness and distribution robustness for nonlinear regression model in terms of fitting performance and model selection. The results from the numerical examples illustrate the usefulness and the superiority of our methodology.

The remainder of this paper is organized as follows. In Section 2, we briefly discuss the FSS distribution and FSTN distributions. In Section 3, we present the likelihood inference including the quantities of the first- and the second-order derivatives as well as the standard Newton-Raphson iterative formula. In Section 4, we give an introduction of profile inference for our proposed model, where the confidence estimation and hypothesis test are presented too. Section 5 gives numerical examples using both simulated and real data to illustrate the performance of the proposed methodology. Finally, some concluding remarks are given in Section 6.

2. Models and Notation

The class of skewed distributions such as SN, ST, and STN perform to be plausible for modeling skewness or (and) heavy tails underlying the observations. The actual situation is that it is not rare at all to encounter multimodality, sometimes with an even more irregular shape, and, for this case, the aforementioned distributions become unsufficient. In this paper, with the adoption of a sufficiently flexible class of distributions, we consider one of these extensions, referred to as the family of flexible skew-symmetric (FSS) distributions which is introduced by [9] with the following density function of type:

$$f(y; \mu, \omega, \boldsymbol{\alpha}) = 2\omega^{-1} f_0(z) G\{P_K(z)\}, \qquad (1)$$

where f_0 and G are symmetric univariate density and distribution function, respectively, that is, $f_0(x) = f_0(-x)$, $G(-x) = 1 - G(x)$, and $P_K(x) = \alpha_1 x + \alpha_3 x^3 + \cdots \alpha_{2K-1} x^{2K-1}$ is an odd polynomial of degree K (i.e., a polynomial including only terms of odd degree), $z = \omega^{-1}(x - \mu)$, $\mu \in R$ is the location parameter, $\omega > 0$ is the scale parameter, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_3, \ldots, \alpha_{2K-1})$ and $\alpha_1, \alpha_3, \ldots, \alpha_{2K-1} \in R$ are shape parameters.

In general, the density function of STN distribution can be represented as $2\omega^{-1} t(z; v)\Phi(\alpha z)$, where t and Φ, respectively, denote the univariate standard Student's t density function and the univariate standard normal distribution function and v is the degrees of freedom. The skewness is regulated by the shape parameter α and the tail thickness of the distribution is controlled by v. Commonly, in comparison with the SN distribution, the STN distribution exhibits obvious feature of heavy tails when $v \leq 10$.

In this paper, we work with one version of (1) and the specific definition can be presented as follows. Let $f_0(z) = t(z; v)$ and $G(x) = \Phi(x)$, where $\Phi(x)$ is defined as before and

$$t(z; v) = \frac{\Gamma(v+1)/2}{\omega(v\pi)^{1/2} \Gamma(v/2)} \left(1 + \frac{z^2}{v}\right)^{-((v+1)/d)}, \qquad (2)$$

that is, the density function of univariate Student's t distribution with 0 location, ω scale, and v degrees of freedom. The above extension of (1) is referred to as flexible skew-t-normal (FSTN) distribution, denoted by $FSTN(\mu, \omega, \boldsymbol{\alpha}, v)$.

It is noted that the FSTN distribution is proposed within the general framework of FSS distribution by combining with

the definition of STN distribution and, as a consequence, it shares analogous feature with these two distributions. For all of that, the FSTN distribution presents some interesting and peculiar features and is able to regulate the density in a more flexible way. To be particular, except for modeling skewness and tail thickness, the FSTN distribution allows for multimodality, depending on the specific setting of $\boldsymbol{\alpha}$. For the purpose of comparison and illustration, we assume $f_0(z) = \phi(z)$ and $G(z) = \Phi(z)$ in FSS distribution, which is denoted by FSN, and, then, we set $K = 2$ in $P_K(z)$, that is, $P_K(x) = \alpha_1 x + \alpha_3 x^3$ in STN distribution, and, for this case, different selections of α_1 and α_3 determine whether the density is unimodal or bimodal. Moreover, the same assumption for G and P in FSTN distribution is made.

Figure 1 displays the density functions of FSN and STN as well as FSTN distributions with four different situations considered, namely, $\alpha_1 = 1$, $\alpha_3 = 0$, and $\nu = 10$; $\alpha_1 = \alpha_3 = 1$ and $\nu = 6$; $\alpha_1 = 1$, $\alpha_3 = -1$, and $\nu = 4$; $\alpha_1 = -1$, $\alpha_3 = 1$, and $\nu = 4$, respectively, with $\mu = 1$, $\omega = 1.5$ for all cases. By examination of Figure 1, we can detect how these three densities change with different combinations of $\boldsymbol{\alpha}$ and ν. For instance, in Figure 1(a), FSN, STN, and FSTN appear to be very close, while STN and FSTN are heavy tailed to a little extent. In Figure 1(b), both FSN and FSTN are unimodal when α_1 and α_3 keep the same sign, and the ranking for the degree of skewness is STN, FSTN, and FSN in turn; that is, the STN and FSTN distributions have thicker tails compared to FSN distribution. With opposite sign of α_1 and α_3, both FSN and FSTN distributions are bimodal and highly skewed in Figures 1(c) and 1(d); moreover, in the same direction, FSTN has thicker tails than FSN distribution. Our proposed FSTN distribution can be treated as a proper compromise between the FSS distribution and the STN distribution. It allows for a wider range of tail behavior compared to FSS distribution whereas it is able to accommodate multimodality which cannot be described by STN distribution. From the applied viewpoint, the FSTN distribution is an appealing option which can be expected to yield robust inferential results in the presence of outlying observations.

3. Likelihood Inference

Consider n independent observations satisfying a nonlinear regression model as $[y_i] \sim \mathrm{FSTN}(\mu_i, \omega, \boldsymbol{\alpha}, \nu)$ with $\mu_i = \eta(\mathbf{x}_i, \boldsymbol{\beta})$ for $i = 1, 2, \ldots, n$. Here, $\mathbf{x}_i$ is a p-dimensional vector and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters. Also, let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ be the $n \times p$ design matrix; $\eta(\cdot, \cdot)$ is a known twice differentiable function. Then, the corresponding log-likelihood for parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \omega, \nu)$ is given by

$$l(\boldsymbol{\theta}) = \text{constant} - n \log \omega + \sum_{i=1}^{n} \log t(z_i) + \sum_{i=1}^{n} \log \Phi\{P_K(z_i)\}.$$

(3)

And we have

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -\frac{1}{\omega} \sum_{i=1}^{n} \left(\frac{t'(z_i)}{t(z_i)} + P_K'(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right) \frac{\partial \eta_i}{\partial \boldsymbol{\beta}},$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \omega} = -\frac{1}{\omega} \sum_{i=1}^{n} \left(\frac{t'(z_i)}{t(z_i)} z_i + z_i P_K'(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} + n \right),$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \nu} = \sum_{i=1}^{n} \frac{z_i^2}{t(z_i)} + \frac{n}{2} \left(DG\left(\frac{\nu+1}{2}\right) - DG\left(\frac{\nu}{2}\right) - \frac{1}{2} \right),$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_{2j-1}} = \sum_{i=1}^{n} z_i^{2j-1} \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} P_K'(z_i), \quad j = 1, 2, \ldots, K,$$

(4)

where $\eta_i = \eta(\mathbf{x}_i, \boldsymbol{\beta})$ and $DG(x) = (d[\log \Gamma(x)])/dx$. The corresponding second-order derivatives of (3) can be shown as

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$$

$$= \sum_{i=1}^{n} \left\{ \left(\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \right)^T \frac{1}{\omega^2} \left[h(z_i) + P_K'(z_i) Q\{P_K(z_i)\} \right. \right.$$
$$\left. + P_K''(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$$
$$\left. - \frac{1}{\omega} \left[\frac{t'(z_i)}{t(z_i)} + P_K'(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \frac{\partial^2 \eta_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right\},$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \omega} = \sum_{i=1}^{n} \left\{ \frac{1}{\omega^2} \left[\frac{t'(z_i)}{t(z_i)} + P_K'(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \right.$$
$$+ \frac{z_i}{\omega^2} \left[h(z_i) + P_K'(z_i) Q\{P_K(z_i)\} \right.$$
$$\left. \left. + P_K''(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \right\} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}},$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \nu} = -\frac{1}{\omega} \sum_{i=1}^{n} \left[\frac{(1 + \nu^{-1}) z_i^3 - t(z_i) z_i^2}{\nu^2 t^2(z_i)} \right] \frac{\partial \eta_i}{\partial \boldsymbol{\beta}},$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \alpha_{2j-1}}$$

$$= \sum_{i=1}^{n} \left\{ z_i^{2j-1} \left[P_K'(z_i) Q\{P_K(z_i)\} + P_K''(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \right\}$$
$$\times \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad j = 1, 2, \ldots, K,$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \omega^2} = \frac{1}{\omega^2} \sum_{i=1}^{n} \left\{ \left[\frac{t'(z_i)}{t(z_i)} + P_K'(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] 3z_i \right.$$
$$+ z_i^2 \left[h(z_i) + P_K'(z_i) Q\{P_K(z_i)\} \right.$$
$$\left. \left. + P_K''(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \right\},$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \omega \partial \nu} = -\frac{1}{\omega} \sum_{i=1}^{n} \left[\frac{(1 + \nu^{-1}) z_i^3 - t(z_i) z_i^2}{\nu^2 t^2(z_i)} \right],$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \omega \partial \alpha_{2j-1}}$$

$$= -\sum_{i=1}^{n} \left[\frac{1}{\omega} \left(P_K'(z_i) Q\{P_K(z_i)\} + P_K''(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right) z_i \right]$$
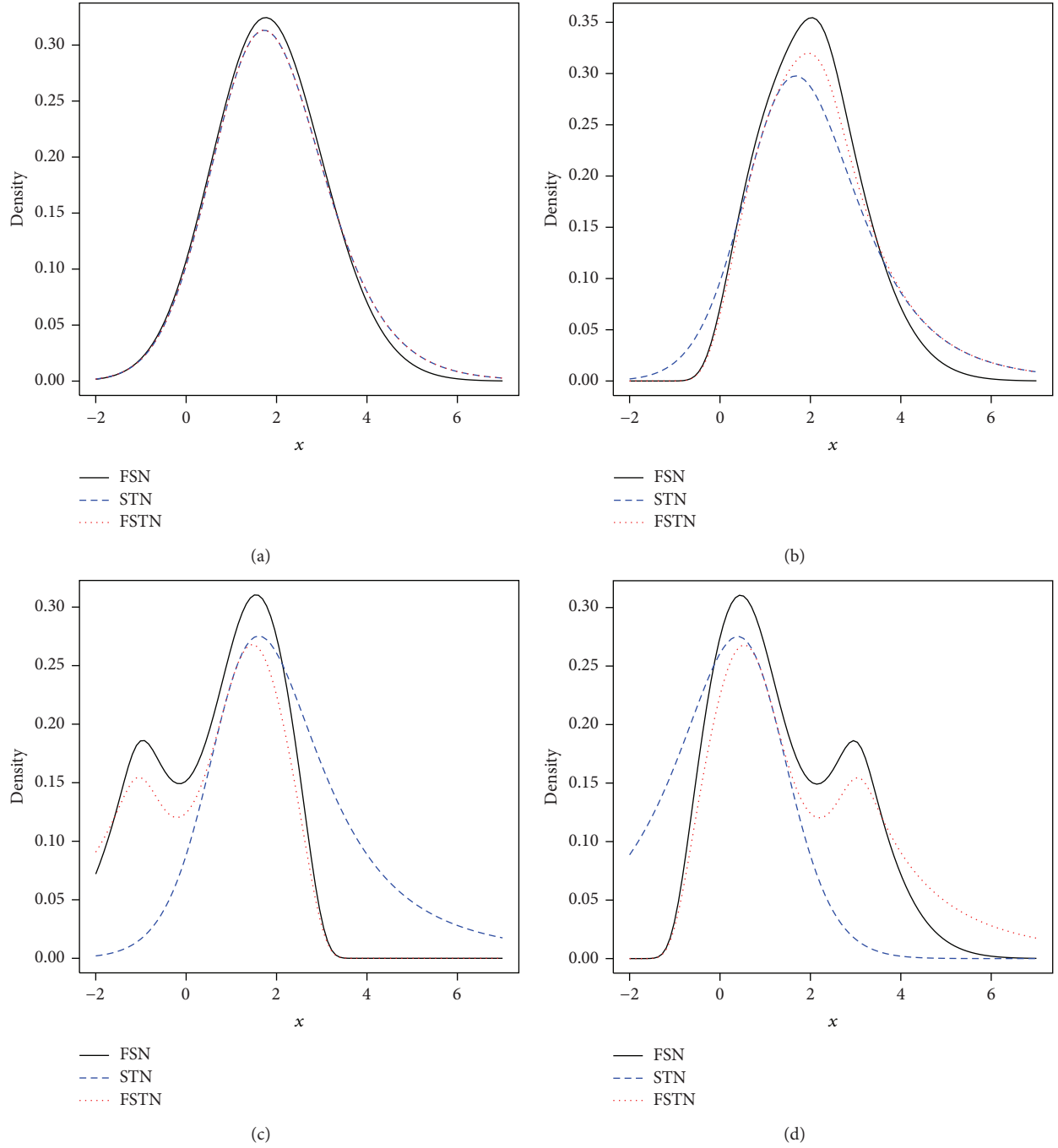$$\times z_i^{2j-1}, \quad j = 1, 2, \ldots, K,$$

FIGURE 1: ((a)–(d)) Density functions of FSN, STN, and FSTN distributions with four different situations.

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial v^2} = \frac{n}{2}\left[TG\left(\frac{v+1}{2}\right) - TG\left(\frac{v}{2}\right)\right] + \sum_{i=1}^{n}\log\left(t\left(z_i\right)\right),$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\alpha_{2j-1}\partial\alpha_{2k-1}} = \sum_{i=1}^{n}z_i^{j+k}Q\left\{P_K\left(z_i\right)\right\}, \quad j,k = 1,2,\ldots,K,$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial v\partial\alpha_{2j-1}} = 0, \quad j = 1,2,\ldots,K,$$

(5)

where $TG(x) = (d^2[\log\Gamma(x)])/dx^2$, $h(x) = (t''(x)t(x) - [t'(x)]^2)/t^2(x)$, and $Q(x) = (\phi(x)/\Phi(x))[x + (\phi(x)/\Phi(x))]$.

Assume $U(\boldsymbol{\theta}) = (U^T(\boldsymbol{\beta}), U(\omega), U(v), U^T(\boldsymbol{\alpha}))^T$ is the gradient or score vector and $H(\boldsymbol{\theta})$ is the Hessian matrix composed of the above second-order derivatives. To obtain the ML estimate of $\boldsymbol{\theta}$, the Newton-Raphson iteration algorithm is defined by

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} - [H(\boldsymbol{\theta})]_{\theta=\widehat{\theta}^{(k)}}^{-1}[U(\boldsymbol{\theta})]_{\theta=\widehat{\theta}^{(k)}}. \tag{6}$$

It is noted that the above iterative procedure is an unpartitioned algorithm; that is, all the parameters including nonlinear regression coefficients $\boldsymbol{\beta}$, scale parameter ω, and shape parameter $\boldsymbol{\alpha}$ as well as tail thickness parameter ν are estimated simultaneously. For our considered problem, there are at least two difficulties that may be encountered for (6); the first one is that once the number of the parameters to be estimated becomes large, the corresponding computational burden turns to be heavy with an unacceptable estimation error, and the second one is as follows: when the strength of nonlinearity of the link function $\eta(\cdot, \cdot)$ changes, the iterative process may become unstable or even nonconvergent, leading to the poor estimation results. Considering the needs of practical problems, rather than the total parameter set, we are usually interested in a proper subset of it. To improve the efficiency of the algorithm and to facilitate statistical inference of the nonlinear models with FSTN distribution, we put forward the following profile likelihood method based on (3) and (6).

4. Profile Likelihood Inference

4.1. Profile Estimation Algorithm. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)$ be a partition of $\boldsymbol{\theta}$, where $\boldsymbol{\beta}$ is a parameter vector of nonlinear regression, an interest parameter, and $\boldsymbol{\gamma}$ is a q-dimension ($q = K + 2$) nuisance parameter with $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \omega, \nu)^T$. Similarly, the partition of H and U is given as $H(\boldsymbol{\theta}) = \left(\begin{smallmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{smallmatrix} \right)$, where $H_{11} = (\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T)$, $H_{12} = (\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^T) = (\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\omega, \partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\nu, \partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\alpha_1, \ldots, \partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\alpha_{2k-1})$, $H_{12} = [H_{21}]^T$, and $H_{22} = (\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T)$, and the diagonal elements of H_{22} are given by $\partial^2 l(\boldsymbol{\theta})/\partial\omega^2, \partial^2 l(\boldsymbol{\theta})/\partial\nu^2, \partial^2 l(\boldsymbol{\theta})/\partial\alpha_1^2, \ldots, \partial^2 l(\boldsymbol{\theta})/\partial\alpha_{2k-1}^2$, respectively, and the off-diagonal elements in H_{22} can be obtained similarly. Let $U(\boldsymbol{\theta}) = (U_1, U_2)$, corresponding to the partition of $\boldsymbol{\theta}$.

In the subsequent context, we focus on the estimation of $\boldsymbol{\theta}$ based on profile likelihood method [18]. Firstly, suppose $\boldsymbol{\beta}$ is known and we rewrite the original likelihood function (3) as

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\beta}, \boldsymbol{\gamma}) = l_\beta(\boldsymbol{\gamma}), \tag{7}$$

where notation $l_\beta(\boldsymbol{\gamma})$ denotes that $\boldsymbol{\beta}$ is fixed but $\boldsymbol{\gamma}$ varies. For each $\boldsymbol{\beta}$, to estimate $\boldsymbol{\gamma}$ we can obtain

$$\widehat{\boldsymbol{\gamma}}_\beta = \arg\max_\gamma l_\beta(\boldsymbol{\gamma}). \tag{8}$$

Alternatively, to estimate $\boldsymbol{\beta}$, we evaluate the maximum value of $l_\beta(\boldsymbol{\gamma})$ over $\widehat{\boldsymbol{\gamma}}_\beta$ and have

$$\widehat{\boldsymbol{\beta}}_p = \arg\max_\beta l_\beta(\widehat{\boldsymbol{\gamma}}_\beta) = \arg\max_\beta l(\boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_\beta), \tag{9}$$

where $l(\boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_\beta)$ and $\widehat{\boldsymbol{\beta}}_p$ are referred to as the profile likelihood function and the profile ML estimation, respectively.

Following [19], we define the profile Newton-Raphson iteration formula as follows:

$$\widehat{\boldsymbol{\beta}}_p^{(k+1)} = \widehat{\boldsymbol{\beta}}_p^{(k)} - \left[H_{\beta|\gamma} \right]^{-1} \left(U_{\beta|\gamma} \right),$$
$$\widehat{\boldsymbol{\gamma}}_p^{(k+1)} = \widehat{\boldsymbol{\gamma}}_p^{(k)} - \left((H_{22})^{-1} U_1 + (H_{22})^{-1} H_{21} \left[H_{\beta|\gamma} \right]^{-1} \left(U_{\beta|\gamma} \right) \right), \tag{10}$$

where $U_{\beta|\gamma} = U_1 - H_{12}(H_{22})^{-1}U_2$, $H_{\beta|\gamma} = H_{11} - H_{12}(H_{22})^{-1}H_{21}$, and all the matrices and the vectors on the right hand side of (10) are evaluated at $\widehat{\boldsymbol{\beta}}_p^{(k)}$ and $\widehat{\boldsymbol{\gamma}}_p^{(k)}$.

Note that, both in (6) and in (10), the strength of nonlinearity of link function $\eta(\cdot, \cdot)$ is reflected by $\partial^2\eta_i/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$ to large extents. Therefore, by examination of the expression of $\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$, we find that when the element of

$$\left| \omega \left[\frac{t'(z_i)}{t(z_i)} + P'_K(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \frac{\partial^2\eta_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \right| \tag{11}$$

is much less than the corresponding element of

$$\left| \frac{\partial\eta_i^T}{\partial\boldsymbol{\beta}} \left[h(z_i) + P'_K(z_i) Q\{P_K(z_i)\} + P''_K(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \right.$$
$$\left. \times \frac{\partial\eta_i}{\partial\boldsymbol{\beta}} \right|, \tag{12}$$

the following approximating result can be obtained as

$$H_{11}^* = \sum_{i=1}^n \left\{ \left(\frac{\partial\eta_i}{\partial\boldsymbol{\beta}} \right)^T \frac{1}{\omega^2} \left[h(z_i) + P'_K(z_i) Q\{P_K(z_i)\} \right. \right.$$
$$\left. \left. + P''_K(z_i) \frac{\phi\{P_K(z_i)\}}{\Phi\{P_K(z_i)\}} \right] \frac{\partial\eta_i}{\partial\boldsymbol{\beta}} \right\}. \tag{13}$$

And then, the iteration formulas of $U_{\beta|\gamma}^*$, $[H_{\beta|\gamma}^*]^{-1}$, $\widehat{\boldsymbol{\beta}}_p$, and $\widehat{\boldsymbol{\beta}}_p^*$ can also be obtained just as before. The above estimation procedure is referred to as profile modified Newton-Raphson iteration algorithm.

The stopping rule for the above algorithm can be presented as that iteration proceeds until some distance involving two successive evaluations of the profile likelihood $l(\boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_\beta)$, such as $|l(\widehat{\boldsymbol{\beta}}^{(k+1)}, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}^{(k+1)}}) - l(\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}^{(k)}})|$ or $|l(\widehat{\boldsymbol{\beta}}^{(k+1)}, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}^{(k+1)}})/l(\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}^{(k)}}) - 1|$, is small enough; for example, $\varepsilon < 10^{-4}$ is adopted in this paper.

The asymptotic covariance matrix of the ML estimates for profile likelihood can be evaluated by inverting the expected information matrix; however, it does not have a closed form expression; the observed information matrix $J(\boldsymbol{\beta}) = -(\partial^2 l(\boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_\beta)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T)$ can be used as a replacement which is estimated by $H_{11} - H_{12}(H_{22})^{-1}H_{21}$, where H_{11}, H_{12}, and H_{22} can be obtained as similar as above.

The choice of the initial values plays an important role in nonlinear regression fitting; in this paper, the specific steps

for choosing the starting values are implemented as follows:

  (i) compute the initial value $\widehat{\boldsymbol{\beta}}^{(0)}$ based on the nonlinear regression model with standard normal assumption;

  (ii) with $\widehat{\boldsymbol{\beta}}^{(0)}$ fixed, compute the initial values $\omega^{(0)}$ and $\boldsymbol{\alpha}^{(0)}$ for the SN and finite mixture SN assumptions, respectively.

In order to simplify the estimation of parameter v, we have fixed integral values for v from 2 to 40 by one; choose the value of v that maximizes the profile likelihood as $v^{(0)}$, and, then, the initial values of $\widehat{\boldsymbol{\beta}}^{(0)}, \omega^{(0)}, \boldsymbol{\alpha}^{(0)}, v^{(0)}$ that are required in the estimation procedure are all obtained.

4.2. Profile Confidence Estimation and Hypothesis Test. Confidence interval and hypothesis test play an important role in statistical inference and, in the subsequent content, we will consider the profile confidence estimation and hypothesis test for the parameters of interest in nonlinear regression models. Suppose $J(\boldsymbol{\theta}) = E[-H(\boldsymbol{\theta})]$; the following regular conditions for likelihood inference are assumed:

  (R-i) $\int (\partial \log L(\mathbf{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}) \cdot L(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = 0$;

  (R-ii) $(\partial/\partial \boldsymbol{\theta}) \int L(\mathbf{x}; \boldsymbol{\theta}) = \int (\partial L(\mathbf{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}) d\mathbf{x} = 0$;

  (R-iii) $(\partial/\partial \boldsymbol{\theta}) \int g(\mathbf{x}) L(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = \int g(\mathbf{x})(\partial L(\mathbf{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}) d\mathbf{x}$, where $g(\mathbf{x})$ is arbitrary measurable function;

  (R-iv) $E(\partial \log L(\mathbf{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta})^2 > 0$.

Apart from the above assumptions, in this paper, some additional conditions are assumed to hold for that:

  (A-i) $\eta(\mathbf{x}_i, \boldsymbol{\beta})$ is twice continuously differentiable with respect to $\boldsymbol{\beta} \in R^p$;

  (A-ii) $D = (\partial \eta_1/\partial \boldsymbol{\beta}, \ldots, \partial \eta_n/\partial \boldsymbol{\beta})^T$ is full rank for all $\boldsymbol{\beta}$;

  (A-iii) assume $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$; then, there exist a neighbour region N_0 of $\boldsymbol{\theta}_0$ and a constant c such that $\lambda_{\min}(J(\boldsymbol{\theta})) > c$ for any $\boldsymbol{\theta} \in N_0$.

Under the above assumption, following [20], we have

$$2\left\{l\left(\widehat{\boldsymbol{\beta}}_P, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}_P}\right) - l\left(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}_{\beta_0}\right)\right\} \longrightarrow \chi^2(p), \qquad (14)$$

where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$ and the convergence is under the meaning of convergence in probability. Based on the profile likelihood theory, the confidence region of $\boldsymbol{\beta}$ with the confidence level $100(1 - \alpha)\%$ is given by

$$C(Y) = \left\{\boldsymbol{\beta} : 2\left[l\left(\widehat{\boldsymbol{\beta}}_P, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}_P}\right) - l\left(\boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_\beta\right)\right] \leq \chi^2(p, \alpha)\right\}, \quad (15)$$

where $\chi^2(p, \alpha)$ denotes the upper α percentile of chi-square distribution with p degrees of freedom. Furthermore, the hypothesis test

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0 \qquad (16)$$

can also be considered and the corresponding test statistics are presented by $2[l(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}_{\widehat{\beta}}) - l(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}_{\beta_0})]$. Unlike standard

likelihood method, profile likelihood confidence intervals and hypothesis test do not need assumption of normality of the estimator or distribution which is too restrictive; they are based on an asymptotic chi-square distribution of the log profile likelihood ratio test statistics and these properties bring a lot of convenience and feasibility in practical computation.

5. Simulation Study

To investigate the experimental performance of our methodology, we undertake a simulation study to compare the fitting performance of misspecified distribution as well as understanding the large sample properties of the ML estimates. To realize this purpose, we generate the artificial data from the following two nonlinear models:

Model (1): drug responsiveness model:

$$y_i = \beta_0 - \frac{\beta_0}{1 + (x_i/\beta_1)^{\beta_2}} + \varepsilon_i, \qquad (17)$$

Model (2): nonlinear growth-curve model:

$$y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x)} + \varepsilon_i, \qquad (18)$$

with two kinds of skewed distributions for random error as follows: Case (I): $\varepsilon \sim \text{STN}(\mu, \omega, \alpha, v)$ and Case (II): $\varepsilon \sim \text{FSTN}(\mu, \omega, \boldsymbol{\alpha}, v)$.

The true parameter population is chosen as follows: $\boldsymbol{\beta} = (2, 5, -0.8)$ is assumed to be the same for the above two models. In Model (1), let $\mu = \delta_1, \omega = 1.5, \alpha = 1.5$, and $v = 4$ and in Model (2) let $\mu = \delta_2, \omega = 1.5, \boldsymbol{\alpha} = (1, -0.8)$, and $v = 4$. Note that, for the values of the degrees of freedom v, a relative value ($v = 4$) can yield a heavy-tailed distribution as we need.

In this simulation, there are totally four simulated data sets corresponding to two nonlinear models, Model (1) and Model (2), along with two distributions, Case (I) and Case (II), for random error. Similar to previous analysis, each simulated data set is fitted under STN and FSN as well as FSTN scenarios using three different estimation algorithms. To shed light on the experimental performance of our methodology, an interesting comparison can be made by examining how often we can recognize the true model.

Table 1 shows the absolute value of the average bias between the true and the estimated parameters and the percentages of each model being ranked as the best model based on AIC criterion out of 500 replications are also presented in Table 1.

By examination of Table 1, we can find the following: (i) the difference for the estimation results between drug responsiveness model and growth-curve model is significant, indicating that the nonlinearity of model imposes an impact on parameter estimation; (ii) when the true distribution for random error is STN, the three fitting distributions have roughly the same behavior; it is hard to tell which approach for parameter estimation is good and which approach is bad and the similar results can be obtained for model selection based on AIC; (iii) when the true distribution for random error is FSTN, the FSN and the FSTN distributions

TABLE 1: Results of the simulation study.

Method	Model	Error	Fitting model	β_0	β_1	β_2	(%)
NR	1	Case I	STN	0.048	0.044	0.034	38.4
			FSN	0.052	0.043	0.031	30.2
			FSTN	0.051	0.048	0.035	31.4
		Case II	STN	0.941	0.445	0.374	10.5
			FSN	0.035	0.014	0.011	43.8
			FSTN	0.033	0.010	0.010	45.7
	2	Case I	STN	0.067	0.061	0.053	36.2
			FSN	0.070	0.057	0.055	30.2
			FSTN	0.073	0.072	0.048	33.6
		Case II	STN	0.963	0.469	0.391	8.5
			FSN	0.047	0.023	0.017	44.6
			FSTN	0.043	0.022	0.021	46.9
P-NR	1	Case I	STN	0.049	0.040	0.032	32.6
			FSN	0.050	0.047	0.038	40.1
			FSTN	0.052	0.049	0.036	27.3
		Case II	STN	1.029	0.420	0.248	11.4
			FSN	0.039	0.018	0.010	39.6
			FSTN	0.038	0.017	0.009	49.0
	2	Case I	STN	0.070	0.050	0.047	30.3
			FSN	0.067	0.057	0.056	42.3
			FSTN	0.073	0.071	0.050	27.4
		Case II	STN	1.052	0.446	0.265	10.2
			FSN	0.053	0.031	0.020	39.6
			FSTN	0.052	0.022	0.017	50.2
MP-NR	1	Case I	STN	0.045	0.049	0.033	32.3
			FSN	0.051	0.047	0.036	38.3
			FSTN	0.045	0.044	0.031	29.4
		Case II	STN	0.773	0.373	0.266	14.5
			FSN	0.031	0.018	0.011	38.4
			FSTN	0.028	0.010	0.008	47.1
	2	Case I	STN	0.068	0.072	0.062	28.1
			FSN	0.076	0.074	0.062	41.2
			FSTN	0.073	0.060	0.054	30.7
		Case II	STN	0.794	0.394	0.290	11.3
			FSN	0.037	0.026	0.017	39.4
			FSTN	0.036	0.021	0.013	49.3

are adopted for parameter estimation and samples of different sizes (n = 50, 100, and 200) are generated from Models (1) and (2). We compute the 95% confidence interval for the parameters of interest and the mean square error (MSE) for different model, where $\text{MSE}(\boldsymbol{\beta}) = (1/500) \sum_{i=1}^{500} \|\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}\|^2$. The length of the 95% confidence interval and the MSE results are summarized in Table 2. From Table 2 we can see that the length of the confidence interval and the MSE tend to decrease as the sample size increases as expected.

Tables 1 and 2 show that in general FSTN distribution enjoys more robustness and flexibility in modeling data with skewness and heavy tails as well as multimodality in comparison with other skewed alternatives and the implementation of MPNR method brings more accuracy and improvement for model estimation in the context of nonlinear regression with this new distribution.

6. Conclusion

We have proposed a new skewed distribution based on the general FSS distribution framework, called the FSTN distribution, which is allowed to accommodate multimodality, asymmetry, and heavy tails jointly to offer greater flexibility than SN and STN counterparts. Moreover, we have developed a modified profile of Newton-Raphson iterative algorithm for estimating the parameters of interest of nonlinear model with FSTN distribution and the interval estimation and hypothesis test in a profile likelihood paradigm are also considered.

Numerical studies reveal that, in the context of nonlinear regression analysis, if the true distribution is STN or FSN whereas the fitting distribution is FSTN, the estimation results are not influenced by this misspecification of distribution assumption. However, once the true distribution is FSTN while the fitting distribution is STN, the estimation results appear to be somewhat disappointing, which shows the robustness of FSTN distribution. In general, the combination of FSTN distribution with MPNR method brings more accuracy and improvement on the estimation of nonlinear regression.

So far the present methodology is limited to the complete data analysis; the extensions of this paper include missing data version as well as Bayesian analysis of this model, which will be reported in another paper.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

outperform the STN distribution by producing estimates with lower bias and higher AIC proportion; furthermore, the PNR and MPNR methods perform better than the traditional NR method in general.

To study the consistence properties of ML estimate, we focus on the situation that the true distribution for random error is Case (II) whereas the fitting distribution is FSTN too. For this case, two estimation algorithms, PNR and MPNR,

TABLE 2: The table shows the length of 95% confidence interval of parameters of interest and MSE of specified model.

Model	Sample size	Profile-NR				Modified profile-NR			
		β_0	β_1	β_2	MSE	β_0	β_1	β_2	MSE
1	50	1.015	0.675	0.374	0.195	0.959	0.589	0.341	0.167
	100	0.558	0.365	0.152	0.051	0.519	0.254	0.132	0.073
	200	0.172	0.134	0.076	0.004	0.120	0.141	0.081	0.005
2	50	1.258	0.828	0.449	0.240	1.152	0.692	0.390	0.154
	100	0.569	0.448	0.208	0.099	0.516	0.427	0.211	0.022
	200	0.232	0.181	0.122	0.009	0.214	0.145	0.120	0.008

References

[1] G. Verbeke and E. Lesaffre, "A linear mixed-effects model with heterogeneity in the random-effects population," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 217–221, 1996.

[2] P. Ghosh, M. D. Branco, and H. Chakraborty, "Bivariate random effect model using skew-normal distribution with application to HIV-RNA," *Statistics in Medicine*, vol. 26, no. 6, pp. 1255–1267, 2007.

[3] A. Azzalini and M. G. Genton, "Robust likelihood methods based on the skew-t and related distributions," *International Statistical Review*, vol. 76, no. 1, pp. 106–129, 2008.

[4] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, vol. 12, no. 2, pp. 171–178, 1985.

[5] A. Azzalini and A. Capitanio, "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, vol. 65, no. 2, pp. 367–389, 2003.

[6] H. W. Gómez, O. Venegas, and H. Bolfarine, "Skew-symmetric distributions generated by the distribution function of the normal distribution," *Environmetrics*, vol. 18, no. 4, pp. 395–407, 2007.

[7] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.

[8] A. Azzalini and A. Capitanio, "Statistical applications of the multivariate skew normal distribution," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 61, no. 3, pp. 579–602, 1999.

[9] Y. Ma and M. G. Genton, "Flexible class of skew-symmetric distributions," *Scandinavian Journal of Statistics*, vol. 31, no. 3, pp. 459–468, 2004.

[10] A. Azzalini, "The skew-normal distribution and related multivariate families," *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 159–200, 2005.

[11] H. J. Ho, S. Pyne, and T. I. Lin, "Maximum likelihood inference for mixtures of skew Student-*t*-normal distributions through practical EM-type algorithms," *Statistics and Computing*, vol. 22, no. 1, pp. 287–299, 2012.

[12] C. B. Zeller, V. H. Lachos, and F. E. Vilca-Labra, "Local influence analysis for regression models with scale mixtures of skew-normal distributions," *Journal of Applied Statistics*, vol. 38, no. 2, pp. 343–368, 2011.

[13] V. G. Cancho, D. K. Dey, V. H. Lachos, and M. G. Andrade, "Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: estimation and case influence diagnostics," *Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 588–602, 2011.

[14] M. D. Branco and D. K. Dey, "A general class of multivariate skew-elliptical distributions," *Journal of Multivariate Analysis*, vol. 79, no. 1, pp. 99–113, 2001.

[15] F. C. Xie, J. G. Lin, and B. C. Wei, "Diagnostics for skew-normal nonlinear regression models with AR(1) errors," *Computational Statistics and Data Analysis*, vol. 53, no. 12, pp. 4403–4416, 2009.

[16] F.-C. Xie, B.-C. Wei, and J.-G. Lin, "Homogeneity diagnostics for skew-normal nonlinear regressions models," *Statistics & Probability Letters*, vol. 79, no. 6, pp. 821–827, 2009.

[17] L. H. Vanegas and F. J. Cysneiros, "Assessment of diagnostic procedures in symmetrical nonlinear regression models," *Computational Statistics & Data Analysis*, vol. 54, no. 4, pp. 1002–1016, 2010.

[18] T. A. Severini, "An approximation to the modified profile likelihood function," *Biometrika*, vol. 85, no. 2, pp. 403–411, 1998.

[19] G. K. Smyth, "Partitioned algorithms for maximum likelihood and other non-linear estimation," *Statistics and Computing*, vol. 6, no. 3, pp. 201–216, 1996.

[20] B. Wei, *Exponential Family Nonlinear Models*, Springer, Singapore, 1998.