Sinh-skew-normal/Independent Regression Models

Rocío Maehara [1], **Heleno Bolfarine** [2], **Filidor Vilca** [*3], **N. Balakrishnan** [4]

Abstract

Skew-normal/independent (SNI) distributions form an attractive class of asymmetric heavy-tailed distributions that also accommodate skewness. We use this class of distributions here to derive a generalization of sinh-normal distributions (Rieck, 1989), called the sinh-skew-normal/independent (sinh-SNI) distribution. Based on this distribution, we then propose a general class of nonlinear regression models, generalizing the regression models of Rieck and Nedelman (1991) that have been used extensively in Birnbaum-Saunders regression models. The proposed regression models have a nice hierarchical representation that facilitates easy implementation of an EM-algorithm for the maximum likelihood estimation of model parameters and provide a robust alternative to estimation of parameters. Simulation studies as well as applications to a real dataset are presented to illustrate the usefulness of the proposed model as well as all the inferential methods developed here.

Keywords: Nonlinear regression; Birnbaum-Saunders distribution; EM-algorithm; Robust estimation; Skew-normal/Independent distribution; Sinh-normal distribution.

1 Introduction

The normal distribution has played a dominant role in both theoretical and applied statistics, but is known not to always provide an adequate representation for many datasets in practice. According to Johnson (1949), a way to overcome this problem is to build non-normal distributions by transforming a normal random variable suitably. Some examples are the Birnbaum-Saunders (BS) (Birnbaum and Saunders, 1969), sinh-normal (Rieck, 1989), log-normal (Johnson et al., 1994) and Johnson distributions, among several others. From a lifetime data viewpoint, the relationship between sinh-normal and Birnbaum-Saunders distributions have been discussed by Rieck (1989) and Leiva et al. (2010). In addition, Rieck and Nedelman (1991) considered that relationship in regression models, especially when the lifetime data are modeled by the distribution of Birnbaum and Saunders (1969), in which the scale parameter depends on covariates. The BS distribution has been widely used including in engineering, industry, business and medical sciences. In a business context, Paula et al. (2011) considered a type of BS model based on the Student-t distribution and applied it to insurance modeling. Lemonte and Cordeiro (2009) proposed a Birnbaum-Saunders nonlinear regression model (BS-NLM) based on the BS distribution, or simply sinh-normal nonregression models (sinh-normal-NLM), which

[1]Departamento Académico de Ingeniería, Universidad del Pacífico, Perú. e-mail: *rp.maeharaa@up.edu.pe*

[2]Departamento de Estatística, Universidade Estadual de São Paulo, Brazil. e-mail: *hbolfar@ime.usp.br*

[3]Departamento de Estatística, Universidade Estadual de Campinas, São Paulo, Brazil e-mail: *fily@ime.unicamp.br*

[4]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada. e-mail: *bala@mcmaster.ca*

represent a generalization of the linear regression model of Rieck and Nedelman (1991).

As the sinh-normal distribution is symmetric, it can be used in the same way as symmetric or elliptical distributions are used in building regression models. But, the sinh-normal distribution may not be suitable to represent data containing outlying observations due to its close relationship with the normal distribution. To overcome this problem, the relationship between sinh-normal and normal distributions is utilized to obtain a general class of sinh-normal distributions based on the family of scale mixtures of normal (SMN) (Andrews and Mallows, 1974) distributions, or simply normal /Independent (NI) (Lange and Sinsheimer, 1993) distribution, which contains many important unimodal and symmetric distributions such as the contaminated normal, slash, Student-t and Laplace distributions, as special cases. Generalization of the sinh-normal distribution based on the NI distributions can be found in Paula et al. (2011), Vilca et al. (2015) and Vilca et al. (2017).

When the data display skewness, the skew-normal (SN) distribution of Azzalini (1985) could be used in place of the normal distribution to accommodate the tails (left or right) in a better way. This idea has been considered in the Birnbaum-Saunders distribution setting by Vilca et al. (2011) in the prediction of extreme percentiles and Santana et al. (2011) subsequently proposed a Birnbaum-Saunders linear regression model in which the normal distribution has been replaced by the skew-normal distribution. A generalization of the skew-normal distribution, following the lines of Andrews and Mallows (1974) and Lange and Sinsheimer (1993), has been proposed by Branco and Dey (2001), called scale mixtures of skew-normal (SMSN) distributions, or simply skew-normal /Independent (SNI) distributions.

The aim of this paper is to provide a generalization of the sinh-Normal (sinh-N) distribution based on the SNI, called *sinh-skew-normal/independent (sinh-SNI)* distributions. This new family of distributions is attractive as it simultaneously models skewness and heavy tails. This class contains generalizations based on the skew-normal (SN), the skew-t (ST), the skew-slash (SSL), the skew-contaminated normal (SCN) distributions, and those based on symmetric class of NI distributions, called *sinh-NI* distribution; see Vilca et al. (2015). This generalization also results in making the inference robust against departures (for example, by the presence of few outlying observations) from the sinh-normal and sinh-SN distributions.

This family of sinh-SNI distributions is quite flexible and convenient for use in practice, and also can be used in regression setup. Inspired by Lemonte and Cordeiro (2009) and Vilca et al. (2015), who developed the Birnbaum-Saunders nonlinear regression model under sinh-N and sinh-NI distributions, respectively, we consider here an extension of that model in which the symmetric NI distribution is replaced by the asymmetric SNI distribution. The resulting model is referred to as *sinh-SNI nonlinear regression model* (sinh-SNI-NLM). We investigate some inferential aspects of this model and the robust estimation of parameters in a manner similar to that of the SNI model. A key feature of this regression model is that it can be formulated in a flexible hierarchical representation that is useful for some theoretical derivations such as the computation of the maximum likelihood (ML) estimates of the model parameters efficiently through an EM-algorithm (Dempster et al., 1977).

The rest of this paper is organized as follows. In Section 2, we briefly review sinh-normal and SNI distributions and then introduce the sinh-SNI distribution, with some of its properties that are needed in the estimation of parameters. In Section 3, we present the sinh-SNI nonlinear regression models. An EM-algorithm for determining the ML estimates is then presented and the observed information matrix is derived. In Sections 4 and 5, some numerical examples using both simulated and real data are given to illustrate the usefulness of the proposed models as well as the inferential methods developed here. Finally, some concluding remarks are made in Section 6.

2 The Sinh-SNI distribution

First, we briefly describe the sinh-Normal distribution of Rieck (1989). A random variable (r.v.) Y is said to have a sinh-Normal distribution if it is related to the standard normal distribution through the stochastic representation $Y = \mu + \sigma \operatorname{arcsinh}(\alpha Z_0/2)$, where $Z_0 \sim \mathrm{N}(0, 1)$. Following Balakrishnan et al. (2009) and Vilca et al. (2017), we extend this sinh-normal distribution by replacing the standard normal distribution by a standard SNI distribution. Recall that a random variable Z is said to have a standard SNI distribution if it is related to the skew-normal distribution through the stochastic representation

$$Z = U^{-1/2} Z_0, \tag{1}$$

where Z_0 follows a standard skew-normal distribution, denoted by $Z_0 \sim \mathrm{S}N(0, 1, \lambda)$, and U is a positive random variable (r.v.) independent of Z_0. It is well-known that Z_0 can be represented stochastically as $Z_0 = \delta H + X_1\sqrt{1-\delta^2}$, where $H = |X_0|$ and X_1 are independent, and they have half-normal and normal distributions, respectively. So, a random variable Y stochastically represented in terms of $Z = \delta U^{-1/2} H + \sqrt{1-\delta^2}\, U^{-1/2} X_1$ by

$$Y = \mu + \sigma \operatorname{arcsinh}\Big(\frac{\alpha Z}{2}\Big) = \mu + \sigma \operatorname{arcsinh}\Big(\frac{\alpha}{2}\big[\delta U^{-1/2} H + \sqrt{1-\delta^2}\, U^{-1/2} X_1\big]\Big) \tag{2}$$

is said to have a sinh-SNI distribution with shape parameter $\alpha > 0$, location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and skewness parameter $\lambda \in \mathbb{R}$. The notation $Y \sim \text{Sinh-SNI}(\alpha, \mu, \sigma, \lambda; G)$ is used for this distribution, where $G(.)$ is the cumulative distribution function (cdf) of U (indexed by a scalar or vector parameter $\boldsymbol{\nu}$). Some special cases of this distribution are as follows. When $\lambda = 0$, the distribution reduces to the distribution discussed by Balakrishnan et al. (2009) and Vilca et al. (2017) which extends the sinh-normal distribution upon replacing the standard normal distribution by a SNI distribution. When $U = 1$, the SNI distribution reduces to the skewed sinh-normal distribution (Leiva et al., 2010). Moreover, when $\sigma = 2$, the r.v. $T = \exp(Y)$ follows a scale mixture of skew-normal Birnbaum-Saunders (SNI-BS) distribution recently discussed by Maehara et al. (2017).

The stochastic representation in (2) is useful for generating random numbers from the sinh-SNI distribution and also in deriving their structural properties and in implementing the EM-algorithm for calculating the maximum likelihood estimates of the model parameters. Moreover, from the stochastic representation, we observe that the conditional distribution of Y, given $U = u$, follows the sinh-skew-normal (sinh-SN) distribution, i.e., $Y|(U = u) \sim \text{Sinh-SN}\big(\alpha_u, \mu, \sigma, \lambda\big)$, where $\alpha_u = u^{-1/2}\alpha$. From this fact, the probability density function (pdf) of Y can be expressed as

$$f_Y(y) = \frac{1}{\sigma}\,\phi_{\mathrm{SNI}}(\xi_{2y})\,\xi_{1y} = \frac{1}{\sigma}\int_0^\infty \phi\big(\xi_{2y}; 0, u^{-1}\big)\,\Phi\big(\lambda\, u^{1/2}\,\xi_{2y}\big)\,dG(u)\;\xi_{1y},\; y \in \mathbb{R}, \tag{3}$$

where $\phi_{\rm SNI}(\cdot)$ is the pdf of Z in (1), $\xi_{1y} = \xi_1(y;\alpha,\mu,\sigma) = \frac{2}{\alpha}\cosh\left(\frac{y-\mu}{\sigma}\right)$, $\xi_{2y} = \xi_2(y;\alpha,\mu,\sigma) = \frac{2}{\alpha}\sinh\left(\frac{y-\mu}{\sigma}\right)$, $\phi(\psi,\eta^2)$ denotes the $N(\psi,\eta^2)$ density function, and $\Phi(\cdot)$ is the cdf of the standard normal distribution. It can be shown easily that if $Y \sim$sinh-SNI$(\alpha,\mu,\sigma,\lambda;G)$, then $W = a + bY \sim$Sinh-SNI$(\alpha, a+b\mu, |b|\sigma,\lambda;G)$, where $a\in\mathbb{R}$ and $b\neq 0$. This result implies that the sinh-SNI family is closed under linear transformation. The moment generating function (mgf) of Y can be expressed as

$$M_Y(s) = \exp(\gamma s)\int_0^\infty \left[\frac{K_{(2s+1)/2}(\alpha_u^{-2};\lambda) + K_{(2s-1)/2}(\alpha_u^{-2};\lambda)}{K_{1/2}(\alpha_u^{-2})}\right] dG(u),\ \mathrm{s}\in\mathbb{R}, \tag{4}$$

where $\alpha_u = u^{-1/2}\,\alpha$ and

$$K_\nu(w;\lambda) = \frac{1}{2}\int_{-\infty}^{\infty} \exp\big(-w\cosh(u) - \nu\,u\big)\Phi\left(\frac{2\lambda}{\alpha}\sinh\left(\frac{u}{2}\right)\right) \mathrm{d}u,$$

which satisfies $K_\nu(w;\lambda=0) = K_\nu(w)/2$, with $K_\nu(\cdot)$ denoting the modified Bessel function of the third kind defined by

$$K_\nu(w) = \frac{1}{2}\int_{-\infty}^{\infty}\exp\big(-w\cosh(u) - \nu\,u\big)\,\mathrm{d}u.$$

Thus, the moments of Y may be approximated for some special cases.

Theorem 1. *Let $Y \sim$ Sinh-SNI$(\alpha,\mu,\sigma,\lambda;G)$ and $H|(U=u) \sim$ HN$(0,u^{-1})$. Then, the conditional distribution of Y, given $H=h$ and $U=u$, has its pdf as*

$$f_{Y|H,U}(y|h,u) = \phi\big(\lambda_{h,u} + \xi_2(y;\alpha_{\delta,u},\mu,\sigma)\big)\xi_1(y;\alpha_{\delta,u},\mu,\sigma), \quad y\in\mathbb{R},$$

where $\alpha_{\delta,u} = u^{1/2}\,\alpha\sqrt{1-\delta^2}$, $\lambda_{h,u} = -u^{1/2}\,\delta\,h/\sqrt{1-\delta^2}$, $\xi_2(y;\alpha_{\delta,u},\mu,\sigma)$ and $\xi_1(y;\alpha_{\delta,u},\mu,\sigma)$ are as given in (3) with $\alpha_{\delta,u}$ instead of α, and $\delta = \lambda/\sqrt{1+\lambda^2}$.

The conditional distribution of Y, given $H=h$ and $U=u$, is the four-parameter SHN distribution proposed by Leiva et al. (2010), denoted by $Y|(H=h,U=u) \sim$ SHN$(\alpha_{\delta,u},\mu,\sigma=2,\lambda_{h,u})$. This will be used later in the development of EM-algorithm for determining the maximum likelihood estimates of the model parameters.

Theorem 2. *Let $Y \sim$ Sinh-SNI$(\alpha,\mu,\sigma,\lambda;H)$. Then, $H|(Y=y,U=u)$ has a standard HN distribution with pdf*

$$f_{H|Y,U}(h|y,u) = \frac{\phi\left(h;\delta\,\xi_2(y;\alpha,\mu,\sigma),\ u^{-1}\,(1-\delta^2)\right)}{\Phi(\lambda\,\xi_2(y;\alpha_u,\mu,\sigma))}, \quad h>0,$$

where $\alpha_u = \alpha\,u^{-1/2}$. Moreover,

$$\begin{aligned}\mathrm{E}[UH|Y=y] &= \delta\,\xi_2(y;\alpha,\mu,\sigma)\,\kappa_y + \tau_y\sqrt{1-\delta^2},\\ \mathrm{E}[UH^2|Y=y] &= \delta^2\,\xi_2^2(y;\alpha,\mu,\sigma)\,\kappa_y + [1-\delta^2] + \delta\,\xi_2(y;\alpha,\mu,\sigma)\,\tau_y\sqrt{1-\delta^2},\end{aligned}$$

where $\kappa_y = \mathrm{E}[U|Y=y]$ and $\tau_y = \mathrm{E}\left[U^{1/2}W_\Phi\Big(\frac{\delta\,\xi_2(y;\alpha_U,\mu,\sigma)}{\sqrt{1-\delta^2}}\Big)|Y=y\right]$, with $W_\Phi(u) = \phi(u)/\Phi(u)$ for $u\in\mathbb{R}$.

Proof. Following the idea of Lachos et al. (2010), by using conditional expectation properties, we have $\mathrm{E}\left[UH|Y=y\right] = \mathrm{E}\left[U\mathrm{E}\left(H|Y,U\right)|Y=y\right]$ and $\mathrm{E}\left[UH^2|Y=y\right] = \mathrm{E}\left[U\mathrm{E}\left(H^2|Y,U\right)|Y=y\right]$. These conditional expectations depend on the conditional distribution of H, given Y and U. Thus, we have

$$\begin{aligned}
\mathrm{E}\left[UH|Y=y\right] &= \delta\xi_2(y;\alpha,\mu,\sigma)\mathrm{E}\left[U|Y=y\right] + \sqrt{1-\delta^2}\mathrm{E}[g(U,Y)|Y=y], \\
\mathrm{E}\left[UH^2|Y=y\right] &= \left(\delta\xi_2(y;\alpha,\mu,\sigma)\right)^2\mathrm{E}[U|Y=y] + 1 - \delta^2 \\
&\quad + \delta\xi_2(y;\alpha,\mu,\sigma)\sqrt{1-\delta^2}\mathrm{E}\left[g(U,Y)|Y=y\right],
\end{aligned}$$

where $g(u,y) = u^{1/2}\, W_\Phi\left(\frac{\delta\,\xi_2(y;\alpha_u,\mu,\sigma)}{\sqrt{1-\delta^2}}\right)$. □

Remark 1. *Let* $Y \sim \text{Sinh-SNI}(\alpha,\mu,\sigma,\lambda;G)$*. Then:*

i) If $\Phi_{\text{SNI}}(.;\lambda)$ *is the cdf of a standard SNI distribution, then the cdf of* Y *is given by* $F_Y(y;\alpha,\mu,\sigma,\lambda) = \Phi_{\text{SNI}}(\xi_2(y;\alpha,\mu,\sigma);\lambda)$*. Moreover,*

(a) $F_Y(y;\alpha,\mu,\sigma,-\lambda) = 2\Phi_{\text{NI}}(\xi_2(y;\alpha,\mu,\sigma)) - \Phi_{\text{SNI}}(\xi_2(y;\alpha,\mu,\sigma);\lambda)$*, where* $\Phi_{\text{NI}}(.)$ *is the cdf of a standard NI distribution;*

(b) $F_Y(y;\alpha,\mu,\sigma,\lambda=1) = \Phi^2_{\text{NI}}\big(\xi_2(y;\alpha,\mu,\sigma)\big)$*;*

(c) Let $S \sim \text{Sinh-SNI}(\alpha,\mu,\sigma,\lambda;G)$ *be independent of* Y*. Then,* $V = \max(Y,S)$ *has its density as*

$$f_V(v) = \frac{2}{\sigma}\phi_{\text{SNI}}\big(\xi_2(v;\alpha,\mu,\sigma)\big)\Phi_{\text{SNI}}\big(\xi_2(v;\alpha,\mu,\sigma)\big)\xi_1(v;\alpha,\mu,\sigma);$$

(ii) If there is $\boldsymbol{\nu}_\infty$ *such that* $\lim_{\boldsymbol{\nu}\to\boldsymbol{\nu}_\infty}\phi_{\text{SNI}}(z;\lambda) = f_0(z;\lambda)$ *for all* z*, where* $f_0(.;\lambda)$ *is the pdf of the* $\text{SN}(0,1,\lambda)$ *distribution, then* $\lim_{\boldsymbol{\nu}\to\boldsymbol{\nu}_\infty} f_Y(y) = f_{SN}(y;\lambda)$ *for all* $y \in \mathbb{R}$*, where* $f_{SN}(.;\lambda)$ *is the pdf of a* $\text{Sinh-SN}(\alpha,\mu,\sigma,\lambda)$ *distribution;*

(iii) The distribution of $d(Y) = \xi_2^2(Y;\alpha,\mu,\sigma)$ *does not depend on* λ*. The random variable* $d(Y)$ *is quite useful for testing the goodness of fit of the model as well as for detecting outliers in data; see Lange and Sinsheimer (1993).*

2.1 Special cases of the Sinh-SNI family

Some special cases of the sinh-SNI family are based on the skew-normal (SN), the skew-contaminated normal (SCN), skew-slash (SSL) and skew-Student-t (ST) models. The resulting distributions are referred to as

(i) the sinh-skew normal (sinh-SN),

(ii) sinh-skew-contaminated normal (sinh-SCN),

(iii) sinh-skew-slash (sinh-SSL),

(iv) sinh-skew-Student-t (sinh-ST)

distributions, respectively. Inspired by the works of Balakrishnan et al. (2009) and Vilca et al. (2017), the special cases are reported in Appendix A. Plots of the pdf are given in Figures 1 and 2 for values of the shape parameters $\alpha = 1,5$ and the values of the parameters $\lambda = 0, \pm1, \pm3$, with $\beta = 1$ in all cases. These plots show the great flexibility

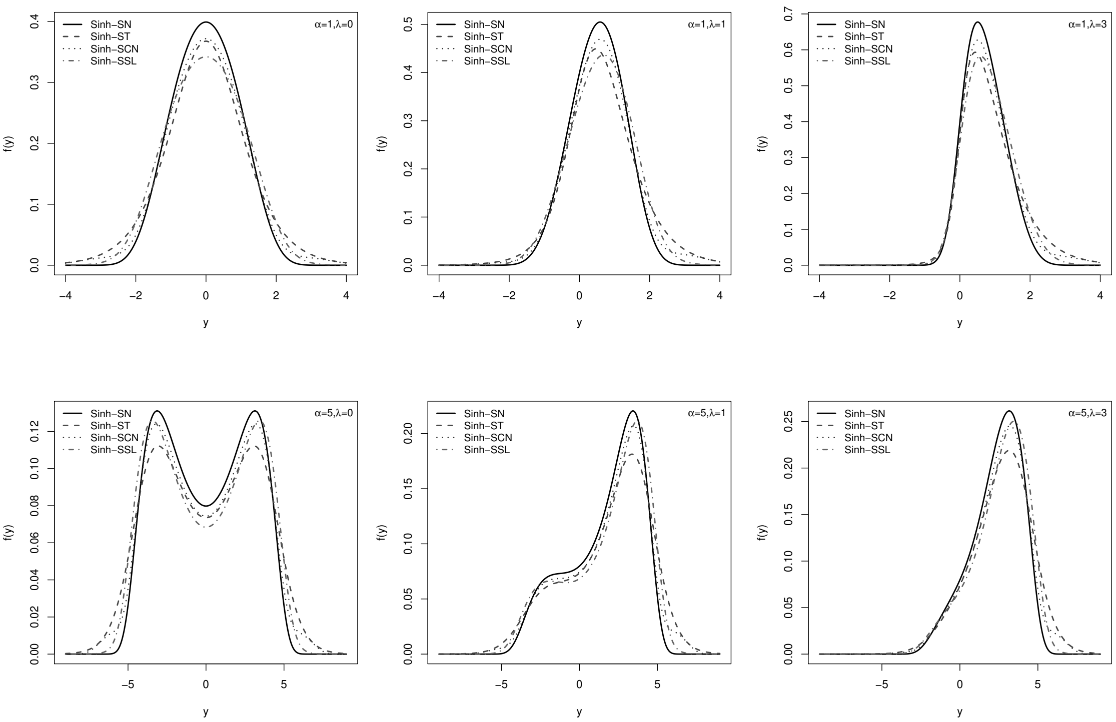


Figure 1: Density plots of the sinh-SNI distributions based on SN SCN, SSL and ST distributions for $\alpha = 1$(left) and $\alpha = 5$ (right), respectively, when $\lambda = 0, 1, 3$.

that the new distribution possesses, and hence it can be very useful in modeling data revealing different shapes. For $\lambda = 0$, we get symmetric distributions, and for $\lambda \neq 0$ asymmetric distributions; it can also exhibit bi-modality when $\alpha > 2$ and $|\lambda|$ is close to one, but when λ increases the density becomes unimodal. As in Rieck and Nedelman (1991), we will next use the class of sinh-SNI distributions to develop a flexible regression model.

3 The proposed regression model

In this section, we introduce the sinh-SNI regression model (sinh-SNI-RM) following the ideas of Lemonte and Cordeiro (2009) and Vilca et al. (2017). Consider the regression model

$$Y_i = f(\boldsymbol{\beta}; \mathbf{x}_i) + \varepsilon_i, \; i = 1, \ldots, n, \tag{5}$$

where Y_i is the response variable, $\mathbf{x}_i$ is an $m \times 1$ vector of known explanatory variables associated with the ith observable response Y_i, and $\mu_i(\boldsymbol{\beta}) = f(\boldsymbol{\beta}; \mathbf{x}_i)$ is an injective and twice differentiable function with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ being a vector of unknown nonlinear parameters. We assume that the errors

$$\varepsilon_i \sim \text{Sinh-SNI}(\alpha, 0, \sigma = 2, \lambda; G), \; i = 1, \ldots, n, \tag{6}$$

are independent. So, $Y_i \sim \text{Sinh-SNI}(\alpha, f(\boldsymbol{\beta}; \mathbf{x}_i), \sigma = 2, \lambda; G)$ with pdf

$$f_{Y_i}(y_i) \;\; = \;\; \frac{1}{2}\, \phi_{\text{SNI}}(\xi_{2i})\, \xi_{1i}, \tag{7}$$

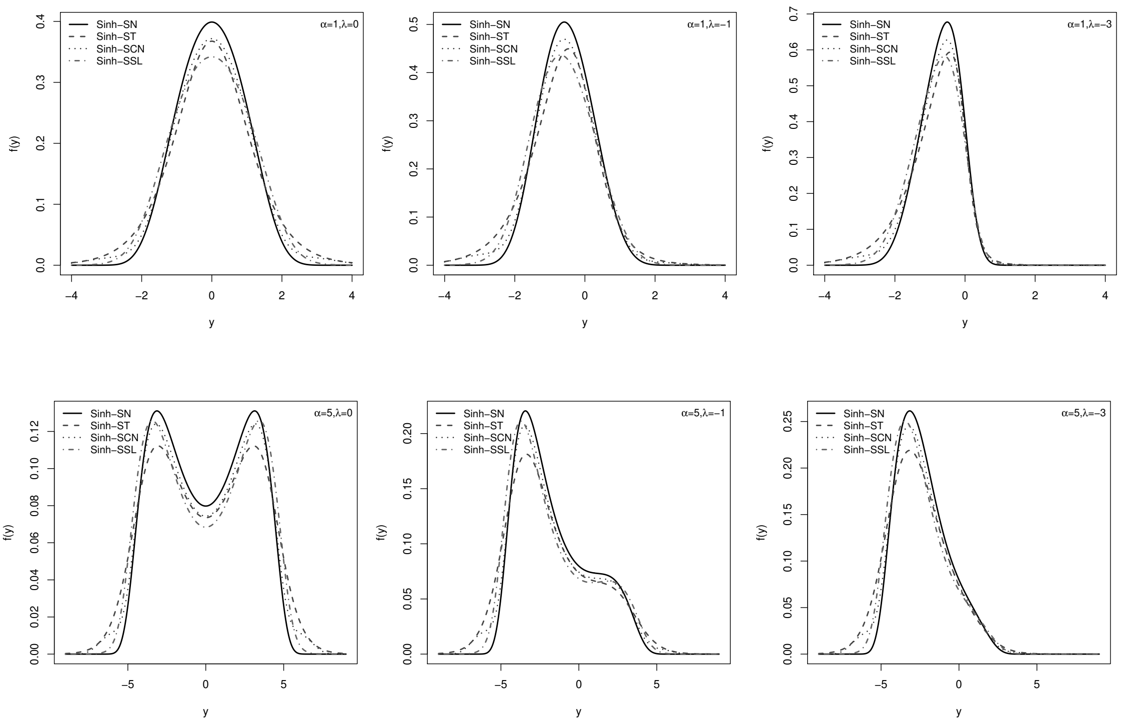


Figure 2: Density plots of the sinh-SNI distributions based on SN, SCN, SSL and ST distributions for $\alpha = 1$ (first line) and $\alpha = 5$ (second line), respectively, when $\lambda = 0, -1, -3$.

where $\xi_{1i} = \xi_1(y_i; \alpha, \boldsymbol{\beta})$ and $\xi_{2i} = \xi_2(y_i; \alpha, \boldsymbol{\beta})$ are as in (3), with $\sigma = 2$ and $\mu_i = f(\boldsymbol{\beta}; \mathbf{x}_i)$ (here, we use ξ_{1i} and ξ_{2i} instead of ξ_{1y_i} and ξ_{2y_i}, respectively). Some special cases of the proposed regression model are as follows:

(i) for $U = 1$ and $\lambda = 0$, we obtain the nonlinear regression model of Lemonte and Cordeiro (2009);
(ii) for $U = 1$ and $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, we obtain the skew-BS log-linear regression model proposed by Santana et al. (2011);
(iii) for $\lambda = 0$, we obtain the nonlinear regression model proposed by Vilca et al. (2015) and Vilca et al. (2017);
(iv) for $\lambda = 0$, $U \sim Gamma(\nu/2, \nu/2)$ and $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, we obtain the linear regression model of Paula et al. (2011).

3.1 ML estimation using an EM-algorithm

The well-known EM algorithm is a useful tool for finding maximum likelihood (ML) estimates in incomplete-data problems. Each iteration of the EM algorithm has two steps: the expectation step (E-step), wherein we calculate the conditional expectation of the complete-data log-likelihood given the observed data (Q-function) and the (current) parameter estimate, and the maximization step (M-step), wherein the (next) parameter estimate is found by maximization of the Q-function. One may refer to Dempster et al. (1977), for example. This maximization method is considered in the sinh-SNI-RM for finding the ML estimates of model parameters. Thus, by using the results in Section

2, Theorem 1 and the representation in (2), the regression model can be expressed in a hierarchical form as

$$
\begin{aligned}
Y_i|(H_i = h_i, U_i = u_i) &\overset{\text{ind}}{\sim} \sim \text{SHN}(\alpha_{\delta,u_i}, f(\boldsymbol{\beta};\mathbf{x}_i), \sigma = 2, \lambda_{h_i,u_i}), && (8)\\
H_i|U_i = u_i &\overset{\text{ind}}{\sim} \sim HN(0, u_i^{-1}), && (9)\\
U_i &\overset{\text{ind}}{\sim} h_U(u_i), \quad i = 1, \ldots, n, && (10)
\end{aligned}
$$

where $\alpha_{\delta,u_i} = u_i^{1/2}\,\alpha\,\sqrt{1-\delta^2}$ and $\lambda_{h_i,u_i} = -u_i^{1/2}\,\delta\,h_i/(\sqrt{1-\delta^2})$. We assume that the parameter vector $\boldsymbol{\nu}$ that indexes the pdf $h_U(\cdot)$ is known. An optimal value of $\boldsymbol{\nu}$ can then be chosen by using the Schwarz information criterion; see Spiegelhalter et al. (2002) and Lange et al. (1989).

Let $\mathbf{u} = (u_1, \ldots, u_n)^\top$ be the unobserved data. Then, the complete data is $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ which correspond to the original data $\mathbf{y} = (y_1, \ldots, y_n)^\top$ augmented with $\mathbf{u}$. Thus, under the hierarchical representation given in equations (8)-(10), it follows that the complete-data log-likelihood function for $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \lambda)^\top$, given $\mathbf{y}_c$ (without the additive constant), can be expressed as

$$
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = -\frac{n}{2}\log(1-\delta^2) + \sum_{i=1}^{n}\log\left(\xi_{1i}\right) - \frac{1}{2(1-\delta^2)}\sum_{i=1}^{n} u_i\,(\xi_{2i} - \delta\,h_{u_i})^2, \qquad (11)
$$

where ξ_{1i} and ξ_{2i} are as in (7). The conditional expectation of the complete-data log-likelihood function has the form

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n}{2}\log(1-\delta^2) + \sum_{i=1}^{n}\log\left(\xi_{1i}^{(k)}\right) - \frac{1}{2(1-\delta^2)}\sum_{i=1}^{n}\widehat{\kappa}_i^{(k)}\,\xi_{2i}^{2(k)}\\
&\quad + \frac{\delta}{1-\delta^2}\sum_{i=1}^{n}\widehat{h}_{u_i}^{(k)}\,\xi_{2i}^{2(k)} - \frac{\delta^2}{2(1-\delta^2)}\sum_{i=1}^{n}\widehat{h^2}_{u_i}^{(k)}, \qquad (12)
\end{aligned}
$$

where the expressions $\widehat{\kappa}_i = \mathrm{E}[U_i|\widehat{\boldsymbol{\theta}}^{(k)}, y_i]$, $\widehat{h}_{u_i} = \mathrm{E}[U_i\,H_i|\widehat{\boldsymbol{\theta}}^{(k)}, y_i]$ and $\widehat{h^2_{u_i}} = \mathrm{E}[U_i\,H_i^2|\widehat{\boldsymbol{\theta}}^{(k)}, y_i]$, $i = 1, \ldots, n$, are obtained from Theorem 2 and $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\alpha}^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\lambda}^{(k)})^\top$ denotes the estimate of $\boldsymbol{\theta}$ at the k-th iteration. The quantities $\widehat{h}_{u_i}$ and $\widehat{h^2_{u_i}}$ are given by

$$
\begin{aligned}
\widehat{h}_{u_i} &= \widehat{\delta}\,\widehat{\xi}_{2i}\,\widehat{\kappa}_i + \widehat{\tau}_i\sqrt{1-\widehat{\delta}^2}, && (13)\\
\widehat{h^2}_{u_i} &= \widehat{\delta}^2\widehat{\xi}_{2i}^2\,\widehat{\kappa}_i + [1-\widehat{\delta}^2] + \widehat{\delta}\,\widehat{\xi}_{2i}\,\widehat{\tau}_i\sqrt{1-\widehat{\delta}^2}, && (14)
\end{aligned}
$$

and $\widehat{\kappa}_i$ depends on the distribution of U. Thus, we have the following ECM algorithm:
E-step. Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, compute $\widehat{\kappa}_i^{(k)}$, $\widehat{h}_{u_i}^{(k)}$ and $\widehat{h^2}_{u_i}^{(k)}$, for $i = 1, \ldots, n$, using the expressions in (13) and (14);

CM-step 1. Fix $\widehat{\boldsymbol{\beta}}^{(k)}$ and update $\widehat{\alpha}^{(k)}$ and $\widehat{\delta}^{(k)}$ as

$$
\begin{aligned}
\widehat{\alpha^2}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\kappa}_i\,\xi_{2i}^2(y_i;1,\widehat{\boldsymbol{\beta}}^{(k)}) + [1-\widehat{\tau}_{h_{u_i}}^{(k)}]\left[\frac{\sum_{i=1}^{n}\widehat{h}_{u_i}^{(k)}\,\xi_{2i}(y_i;1,\widehat{\boldsymbol{\beta}}^{(k)})}{n\widehat{\tau}_{h_{u_i}}^{(k)}}\right]^2,\\
\widehat{\delta}^{(k+1)} &= \frac{1}{\widehat{\alpha}^{(k+1)}}\,\frac{\sum_{i=1}^{n}\widehat{h}_{u_i}^{(k)}\,\xi_{2i}(y_i;1,\widehat{\boldsymbol{\beta}}^{(k)})}{n\widehat{\tau}_{h_{u_i}}^{(k)}},
\end{aligned}
$$

where $\xi_{2i}(y_i;1,\boldsymbol{\beta}) = 2\sinh\left(\frac{y_i - f(\boldsymbol{\beta};\mathbf{x}_i)}{2}\right)$ and $\widehat{\tau}^{(k)}_{h_{u_i}} = (1/n)\sum_{i=1}^{n}\widehat{h^2}^{(k)}_{u_i}$;

CM-step 2. Fix $\widehat{\alpha}^{(k+1)}$ and $\widehat{\delta}^{(k+1)}$ and update $\widehat{\boldsymbol{\beta}}^{(k)}$ using

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \underset{\beta}{\text{argmax}}\ Q(\widehat{\alpha}^{(k+1)}, \boldsymbol{\beta}, \widehat{\delta}^{(k+1)}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

This process is iterated until convergence, i.e., the distance between two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta})$, such as

$$|\ell(\widehat{\boldsymbol{\theta}}^{(r+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(r)})| \text{ or } |\ell(\widehat{\boldsymbol{\theta}}^{(r+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(r)}) - 1|,$$

becomes small enough to the desired level of accuracy.

Remark 2. *It is of interest to mention the following points in the implementation of the EM-algorithm described above:*

1) *From the above algorithm, we get the ML estimate of the parameter δ. Then, from the invariance property of the ML estimates, we have the ML estimate of λ as $\widehat{\lambda} = \widehat{\delta}/\sqrt{1-\widehat{\delta}^2}$;*

2) *The estimates in CM-Step 1 are similar to the estimates in Santana et al. (2011). To ensure the effectiveness of the EM algorithm, suitable starting values are required. This can be done, for example, by using modified moment (MM) estimates as starting values; see Kundu et al. (2013);*

3) *The explicit closed-form expressions for these estimates in CM- Step 1 make the EM algorithm to be an efficient method for computing the ML estimates.*

3.2 The observed information matrix

In this section, we obtain the observed information matrix of the sinh-SNI-RM, defined by

$$\mathbf{J}_o(\boldsymbol{\theta}|\mathbf{y}) = -\partial^2\ell(\boldsymbol{\theta}|\mathbf{y})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top.$$

It is well known that, under some regularity conditions, the covariance matrix of the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ can be approximated by the inverse of $\mathbf{J}_o(\boldsymbol{\Theta}|\mathbf{y})$. Following Basford et al. (1997) and Lin et al. (2007), we evaluate

$$\mathbf{J}_o(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = \sum_{i=1}^{n}\widehat{\mathbf{s}}_i^\top\widehat{\mathbf{s}}_i, \tag{15}$$

where $\widehat{\mathbf{s}}_i = \partial\left(\log f(y_i;\boldsymbol{\theta}_j)\right)/\partial\boldsymbol{\theta}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$. We now consider the vector $\widehat{\mathbf{s}}_i$, which is partitioned into components corresponding to all the parameters in $\boldsymbol{\theta}$, as $\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\alpha}, \widehat{s}_{i,\boldsymbol{\beta}}, \widehat{s}_{i,\lambda})^\top$, whose components are given by

$$\widehat{s}_{i,\alpha} = \frac{D_\alpha(f(y_i;\boldsymbol{\theta}))}{f(y_i;\boldsymbol{\theta})}, \quad \widehat{s}_{i,\boldsymbol{\beta}} = \frac{D_{\boldsymbol{\beta}}(f(y_i;\boldsymbol{\theta}))}{f(y_i;\boldsymbol{\theta})}, \quad \widehat{s}_{i,\lambda} = \frac{D_\lambda(f(y_i;\boldsymbol{\theta}))}{f(y_i;\boldsymbol{\theta})},$$

where $D_{\boldsymbol{\eta}}(f(y_i;\boldsymbol{\theta})) = \partial f(y_i;\boldsymbol{\theta})/\partial\boldsymbol{\eta}$, $\boldsymbol{\eta} = \alpha, \boldsymbol{\beta}, \lambda$. For the sinh-SN, we obtain

$$\begin{aligned} D_{\boldsymbol{\tau}}(f(y_i;\boldsymbol{\theta})) &= \phi(\xi_{2i})\left[\frac{\partial\xi_{1i}}{\partial\boldsymbol{\tau}}\Phi(\lambda\,\xi_{2i}) - \frac{1}{2}\,\xi_{1i}\frac{\partial\xi_{2i}^2}{\partial\boldsymbol{\eta}}\Phi(\lambda\,\xi_{2i}) + \lambda\xi_{2i}\frac{\partial\xi_{2i}}{\partial\boldsymbol{\tau}}\phi(\lambda\,\xi_{2i})\right], \\ &\quad \boldsymbol{\tau} = \alpha \quad \text{or} \quad \boldsymbol{\beta}, \\ D_\lambda(f(y_i;\boldsymbol{\theta})) &= \xi_{1i}\xi_{2i}\phi(\xi_{2i})\,\phi(\lambda\,\xi_{2i}). \end{aligned}$$

Let us use the following notations

$$\begin{aligned} I_i^{\Phi}(w) &= \int_0^\infty u^w \exp\Big\{-\frac{1}{2}u\,\xi_{2i}^2\Big\}\Phi(u^{1/2}\lambda\xi_{2i})dG(u), \\ I_i^{\phi}(w) &= \int_0^\infty u^w \exp\Big\{-\frac{1}{2}u\,\xi_{2i}^2\Big\}\phi(u^{1/2}\lambda\xi_{2i})dG(u), \end{aligned}$$

where $i = 1, \ldots, n$, for simplifying the expressions. Thus, we obtain

$$\begin{aligned} D_{\boldsymbol{\tau}}(f(y_i;\boldsymbol{\theta})) &= \frac{1}{\sqrt{2\pi}}\left[\frac{\partial\xi_{1i}}{\partial\boldsymbol{\tau}}I_i^{\Phi}(1/2) - \xi_{1i}\xi_{2i}\frac{\partial\xi_{2i}}{\partial\boldsymbol{\tau}}I_i^{\Phi}(3/2) + \lambda\xi_{1i}\frac{\partial\xi_{2i}}{\partial\boldsymbol{\tau}}I_i^{\phi}(1)\right], \\ &\quad \boldsymbol{\tau} = \alpha \quad \text{or} \quad \boldsymbol{\beta}, \\ D_\lambda(f(y_i;\boldsymbol{\theta})) &= \frac{\xi_{1i}\xi_{2i}}{\sqrt{2\pi}}I_i^{\phi}(1). \end{aligned}$$

Following Lachos et al. (2010), we have closed-form expressions for the quantities $I_i^{\Phi}(w)$ and $I_i^{\phi}(w)$, $i = 1, \ldots, n$, for sinh-SCN, sinh-SSL and sinh-ST models as follows:

- *The sinh-SN model*

$$\begin{aligned} I_i^{\Phi}(w) &= \sqrt{2\pi}\phi(\xi_{2i})\Phi(\lambda\xi_{2i}), \\ I_i^{\phi}(w) &= \phi\big((\xi_{2i}^2 + \lambda^2\xi_{2i}^2)^{1/2}; 0, 1\big); \end{aligned}$$

- *The sinh-SCN model*

$$\begin{aligned} I_i^{\Phi}(w) &= \sqrt{2\pi}\left\{\nu\gamma^{w-1/2}\phi\left(\xi_{2i}; 0, \frac{1}{\gamma}\right)\Phi(\gamma^{1/2}\lambda\xi_{2i}) + (1-\nu)\phi(\xi_{2i})\Phi(\lambda\xi_{2i})\right\}, \\ I_i^{\phi}(w) &= \nu\gamma^{w-1/2}\phi\left((\xi_{2i}^2 + \lambda^2\xi_{2i}^2)^{1/2}; 0, \frac{1}{\gamma}\right) + (1-\nu)\phi\left((\xi_{2i}^2 + \lambda^2\xi_{2i}^2)^{1/2}; 0, 1\right); \end{aligned}$$

- *The sinh-SSL model*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{2^{2+\nu}\Gamma(w+\nu)}{[\xi_{2i}]^{2w+2\nu}}P_1(w+\nu, \xi_{2i}^2/2)E[\Phi(S_i^{1/2})\lambda\xi_{2i}], \\ I_i^{\phi}(w) &= \frac{\nu 2^{w+\nu}\Gamma(w+\nu)}{\sqrt{2\pi}(\xi_{2i}^2 + \lambda^2\xi_{2i}^2)^{w+\nu}}P_1\left(w+\nu, \frac{\xi_{2i}^2 + \lambda^2\xi_{2i}^2}{2}\right), \end{aligned}$$

  where $S_i \sim \text{Gamma}(w+\nu, \xi_{2i}^2/2)I_{(0,1)}$;

- *The sinh-ST model*

$$\begin{aligned} I_{ir}^{\Phi}(w) &= \frac{2^w\nu^{\nu/2}\Gamma(w+\nu/2)}{\Gamma(\nu/2)(\nu+\xi_{2i}^2)^{\nu/2+w}}T\left(\frac{\lambda\xi_{2i}}{(\xi_{2i}^2+\nu)^{1/2}}\sqrt{2w+\nu}; 2w+\nu\right), \\ I_{ir}^{\phi}(w) &= \frac{2^w\nu^{\nu/2}}{\sqrt{2\pi}\Gamma(\nu/2)}\big(\xi_{2i}^2 + \lambda^2\xi_{2i}^2 + \nu\big)^{-\frac{\nu+2w}{2}}\Gamma\Big(\frac{\nu+2w}{2}\Big). \end{aligned}$$

4 Simulation study

In this section, the quality of the estimation method (ECM algorithm) proposed in the preceding Section and the finite-sample performance of the estimates are evaluated using simulated data obtained by the Monte Carlo procedure for the regression model $Y_i = f(\boldsymbol{\beta}; \mathbf{x}_i) + \varepsilon_i$, $i = 1, \ldots, n$.

4.1 Experiment 1: Recovery of parameters

In this section, we use Monte Carlo simulations to evaluate the performance of the ML estimates of the parameters of the Sihn-SNI-RM obtained by using the proposed EM-algorithm. The samples sizes and the number of Monte Carlo replications considered are $n = 50, 100, 200, 400$ and $M = 1000$, respectively, of the following regression models that are considered for illustrative purposes when the true values of the parameters are $\lambda = 3$ and $\alpha = 0.5, 1$:

$$\text{M1} : f(\boldsymbol{\beta}; \mathbf{x}_i) = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i}; \quad \text{M2} : f(\boldsymbol{\beta}; \mathbf{x}_i) = \beta_1 z_{i1} + \beta_2 zi2 + \beta_3 \exp(\beta_4 x_i),$$

where ε_i follows the sinh-SNI distribution. For Model M1, the linear regression model (LM), we consider $\boldsymbol{\beta} = (1, 2, 3)^\top$. The results of the simulation study are only reported under the sinh-ST ($\nu = 3$) presented in Table 1 because the results were quite similar under other models. For Model M2, the nonlinear regression model (NLM), we consider $\boldsymbol{\beta} = (4, 5, 3, 1.5)^\top$ and the regression models under sinh-SN model, sinh-SCN S model ($\nu = 0.1, \gamma = 0.1$), sinh-SSL ($\nu = 3$) and sinh-ST ($\nu = 3$) models. The corresponding results are in Tables 2-3 wherein the means of the parameter estimates, standard deviations (MC Sd) of the estimates across the Monte Carlo samples (MC Sd), the average (IM SE) values of the approximate standard deviations of the estimates obtained through the method described in Subsection 3.2, and the percentage of times that the confidence intervals cover the true value of the parameter (COV MC) are all presented. From Table 1 and Tables 2-3, we observe that under the sinh-ST model the performance of the EM-algorithm seems to be better for the linear regression model than the nonlinear regression model, and that it does not require a large sample size for the recovery of parameters. We also note that the estimation method of the standard errors provides relatively close results (IM SE and MC Sd) and the COV MC for the parameters is quite stable, indicating that the proposed asymptotic approximation for the variances of the ML estimates (Eq.(15)) is quite reliable.

Table 1: Model M1: Mean and MC Sd are the respective mean estimates and standard deviations under the sinh-ST linear model, while IM SE is the average value of the approximate standard error obtained through the information-based method.

(α)	n	Measure	$\widehat{\alpha}$	$\widehat{\beta}_1$ (1)	$\widehat{\beta}_2$ (2)	$\widehat{\beta}_3$ (3)	$\widehat{\lambda}$ (3)
	50	Mean	0.4728	1.0249	1.9897	3.0099	3.4363
		IM SE	0.0988	0.1482	0.1878	0.1679	2.5514
		MC Sd	0.0913	0.1361	0.1769	0.1597	1.7402
		COV	92.3%	95.3%	95.4%	94.2%	96.9%
	100	Mean	0.4927	1.0041	2.0046	3.0005	3.4737
		IM SE	0.0702	0.1004	0.1218	0.1133	1.5830
		MC Sd	0.0682	0.0965	0.1209	0.1144	1.4324
		COV	94.6%	94.8%	94.6%	94.4%	95.4%
(0.5)	200	Mean	0.4971	1.0056	1.9991	2.9963	3.2637
		IM SE	0.0499	0.0682	0.0834	0.0799	0.9773
		MC Sd	0.0495	0.0690	0.0810	0.0816	1.0487
		COV	94.9%	94.6%	95.5%	94.2%	95.3%
	400	Mean	0.4975	1.0012	1.9998	3.0035	3.0859
		IM SE	0.0354	0.0502	0.0607	0.0590	0.6263
		MC Sd	0.0345	0.0489	0.0616	0.0604	0.6051
		COV	95.0%	95.5%	94.9%	94.1%	96.3%
	50	Mean	0.9443	1.0410	1.9872	3.0177	3.4661
		IM SE	0.2212	0.3020	0.3870	0.3599	2.6455
		MC Sd	0.1875	0.2749	0.3558	0.3398	1.7190
		COV	94.4%	95.2%	95.1%	94.4%	99.1%
	100	Mean	0.9862	1.0118	2.0070	3.0065	3.5109
		IM SE	0.1569	0.2280	0.2568	0.2195	1.6856
		MC Sd	0.1513	0.2275	0.2450	0.2252	1.5215
		COV	94.2%	94.5%	95.5%	93.6%	97.2%
(1.0)	200	Mean	0.9929	1.0035	1.9968	3.0071	3.2769
		IM SE	0.1112	0.1477	0.1600	0.1648	1.0511
		MC Sd	0.1094	0.1494	0.1571	0.1647	1.0667
		COV	94.5%	94.5%	95.1%	95.1%	95.7%
	400	Mean	0.9932	1.0077	1.9999	2.9961	3.0976
		IM SE	0.0789	0.0990	0.1158	0.1152	0.6785
		MC Sd	0.0769	0.1003	0.1199	0.1152	0.6781
		COV	95.7%	94.9%	94.1%	93.7%	96.0%

Table 2: Model M2: Mean and MC Sd are the respective mean estimates and standard deviations under the sinh-SN and sinh-ST nonlinear models, while IM SE is the average value of the approximate standard error obtained through the information-based method.

			Sinh-SN						Sinh-ST					
(α)	n	Measure	$\widehat{\alpha}$	$\widehat{\beta}_1$ (4)	$\widehat{\beta}_2$ (5)	$\widehat{\beta}_3$ (3)	$\widehat{\beta}_4$ (1.5)	$\widehat{\lambda}$ (3)	$\widehat{\alpha}$	$\widehat{\beta}_1$ (4)	$\widehat{\beta}_2$ (5)	$\widehat{\beta}_3$ (3)	$\widehat{\beta}_4$ (1.5)	$\widehat{\lambda}$ (3)
	50	Mean	0.4703	3.9939	5.0008	3.0220	1.4965	3.4097	0.4642	4.0020	4.9953	3.0242	1.4958	3.3525
		IM SE	0.0914	0.1815	0.1894	0.1436	0.0422	2.8559	0.0976	0.1918	0.1686	0.1255	0.0385	2.5233
		MC Sd	0.0694	0.1570	0.1678	0.1232	0.0359	1.6637	0.0838	0.1805	0.1638	0.1171	0.0360	1.5900
		COV	97.8%	96.6%	97.3%	96.9%	97.4%	99.0%	92.2%	94.6%	94.7%	96.0%	94.8%	97.5%
	100	Mean	0.4905	3.9949	5.0002	3.0079	1.4989	3.5505	0.4904	4.0033	5.0016	3.0032	1.5000	3.5394
		IM SE	0.0595	0.1163	0.1125	0.0870	0.0260	1.7671	0.0696	0.1368	0.1148	0.0881	0.0277	1.6592
		MC Sd	0.0538	0.1150	0.1095	0.0832	0.0251	1.4613	0.0698	0.1345	0.1116	0.0894	0.0284	1.5029
		COV	96.4%	94.2%	95.0%	95.4%	95.4000	99.3%	93.2%	94.1%	94.6%	93.7%	92.9%	95.5%
(0.5)	200	Mean	0.4955	3.9985	4.9986	3.0042	1.4990	3.2842	0.4959	4.0044	4.9982	3.0013	1.4998	3.3214
		IM SE	0.0410	0.0788	0.0722	0.0572	0.0176	1.0108	0.0494	0.0798	0.0928	0.0652	0.0202	1.0015
		MC Sd	0.0395	0.0775	0.0685	0.0583	0.0177	0.9988	0.0489	0.0779	0.0964	0.0634	0.0196	1.0324
		COV	96.6%	95.3%	95.8%	94.6%	94.3%	97.5%	95.1%	94.7%	93.9%	95.4%	95.7%	96.4%
	400	Mean	0.4972	4.0019	5.%3	3.0012	1.4998	3.1075	0.4988	4.0047	4.9984	3.0001	1.5000	3.1680
		IM SE	0.0285	0.0515	0.0498	0.0387	0.0117	0.6217	0.0352	0.0605	0.0587	0.0437	0.0136	0.6457
		MC Sd	0.0279	0.0504	0.0495	0.0376	0.0114	0.6118	0.0359	0.0611	0.0615	0.0460	0.0143	0.6924
		COV	95.0%	95.9%	94.4%	95.4%	95.0%	96.2%	94.3%	94.4%	94.0%	93.9%	93.9%	94.6%
	50	Mean	0.9279	3.9889	5.0071	3.0444	1.4911	3.4860	0.9369	4.0030	5.0043	3.0362	1.4974	3.3685
		IM SE	0.2121	0.3383	0.3883	0.2538	0.0753	3.1413	0.2228	0.3840	0.3397	0.3038	0.0892	2.6224
		MC Sd	0.1429	0.3046	0.3331	0.2037	0.0629	1.6695	0.1854	0.3615	0.3085	0.2721	0.0813	1.6179
		COV	98.2%	96.6%	97.7%	96.9%	96.6%	99.0%	93.6%	95.2%	95.4%	95.8%	95.5%	98.7%
	100	Mean	0.9722	3.9910	5.0008	3.0186	1.4974	3.4911	0.9695	4.0068	5.0074	3.0090	1.5006	3.4481
		IM SE	0.1408	0.2132	0.2059	0.1774	0.0565	1.8847	0.1550	0.2416	0.2318	0.1778	0.0540	1.6931
		MC Sd	0.1252	0.1925	0.1957	0.1579	0.0507	1.5182	0.1445	0.2441	0.2404	0.1693	0.0513	1.5291
		COV	96.9%	95.7%	96.5%	96.9%	96.4%	99.0%	93.8%	94.4%	92.9%	95.1%	95.1%	96.8%
(1.0)	200	Mean	0.9898	3.9946	5.0099	3.0043	1.4996	3.2866	0.9968	3.9969	4.9986	3.0072	1.4989	3.3406
		IM SE	0.0965	0.1354	0.1439	0.1215	0.0365	1.1031	0.1107	0.1627	0.1706	0.1190	0.0356	1.0742
		MC Sd	0.0912	0.1292	0.1397	0.1173	0.0354	1.0772	0.1043	0.1686	0.1635	0.1157	0.0349	1.1007
		COV	96.9%	95.7%	95.0%	95.6%	95.4%	98.5%	95.5%	93.8%	96.0%	95.1%	94.2%	96.2%
	400	Mean	0.9958	3.9984	4.9976	3.0074	1.4983	3.1225	0.9967	4.0011	5.0049	2.9999	1.5006	3.1592
		IM SE	0.0663	0.0984	0.0965	0.0764	0.0230	0.6811	0.0783	0.1117	0.1141	0.0883	0.0276	0.6903
		MC Sd	0.0651	0.0945	0.0963	0.0758	0.0228	0.6692	0.0762	0.1116	0.1158	0.0886	0.0276	0.6972
		COV	95.4%	95.6%	95.3%	94.9%	95.3%	96.7%	94.6%	94.9%	94.4%	94.9%	94.8%	95.9%

Table 3: Model M2: Mean and MC Sd are the respective mean estimates and standard deviations under the sinh-SSL and sinh-SCN nonlinear models, while IM SE is the average value of the approximate standard error obtained through the information-based method.

			Sinh-SSL						Sinh-SCN					
(α)	n	Measure	$\widehat{\alpha}$	$\widehat{\beta}_1$ (4)	$\widehat{\beta}_2$ (5)	$\widehat{\beta}_3$ (3)	$\widehat{\beta}_4$ (1.5)	$\widehat{\lambda}$ (3)	$\widehat{\alpha}$	$\widehat{\beta}_1$ (4)	$\widehat{\beta}_2$ (5)	$\widehat{\beta}_3$ (3)	$\widehat{\beta}_4$ (1.5)	$\widehat{\lambda}$ (3)
	50	Mean	0.4682	4.0063	4.9873	3.0232	1.4962	3.3600	0.4647	3.9910	5.0111	3.0243	1.4951	3.2610
		IM SE	0.0945	0.2124	0.1904	0.1556	0.0478	2.8590	0.0981	0.2598	0.1803	0.1516	0.0415	2.6652
		MC Sd	0.0746	0.1882	0.1699	0.1307	0.0410	1.6767	0.0819	0.2368	0.1658	0.1333	0.0373	1.5712
		COV	96.0%	95.5%	96.5%	96.7%	95.1%	99.0%	94.1%	95.4%	95.5%	95.8%	95.2%	99.5%
	100	Mean	0.4869	4.0002	4.9994	3.0088	1.4985	3.4852	0.4883	3.9983	5.0053	3.0085	1.4983	3.5094
		IM SE	0.0634	0.1172	0.1186	0.0938	0.0291	1.7665	0.0671	0.1122	0.1160	0.0886	0.0294	1.7726
		MC Sd	0.0605	0.1142	0.1163	0.0861	0.0270	1.4634	0.0655	0.1080	0.1099	0.0868	0.0286	1.5703
		COV	95.3%	94.8%	94.8%	97.4%	97.0%	98.2%	94.5%	94.6%	95.6%	94.5%	93.5%	97.1%
(0.5)	200	Mean	0.4976	3.9997	4.9978	3.0025	1.4995	3.3423	0.4942	3.9981	4.9999	3.0051	1.4993	3.2396
		IM SE	0.0441	0.0788	0.0795	0.0628	0.0200	1.0517	0.0469	0.0759	0.0879	0.0636	0.0197	1.0087
		MC Sd	0.0432	0.0782	0.0808	0.0628	0.0198	1.0596	0.0451	0.0787	0.0885	0.0658	0.0200	1.0278
		COV	95.5%	94.9%	94.3%	94.2%	95.2%	97.2%	95.5%	94.1%	95.2%	94.3%	94.6%	96.1%
	400	Mean	0.4978	4.0033	5.0026	2.9994	1.5002	3.1155	0.4968	4.0052	4.9963	3.0021	1.4996	3.0954
		IM SE	0.0309	0.0569	0.0596	0.0452	0.0142	0.6391	0.0330	0.0574	0.0559	0.0427	0.0132	0.6377
		MC Sd	0.0304	0.0553	0.0633	0.0455	0.0142	0.6365	0.0319	0.0568	0.0569	0.0428	0.0135	0.6392
		COV	95.4%	95.8%	92.9%	94.0%	94.3%	96.1%	95.1%	95.6%	94.3%	95.9%	95.0%	95.8%
	50	Mean	0.9262	4.0079	5.0084	3.0443	1.4919	3.3563	0.9237	3.9787	4.9981	3.0605	1.4900	3.3663
		IM SE	0.2284	0.3810	0.3459	0.2825	0.0858	3.0427	0.2207	0.3514	0.3352	0.3455	0.1069	2.9597
		MC Sd	0.1616	0.3273	0.2955	0.2112	0.0675	1.6271	0.1707	0.3287	0.3247	0.3105	0.0969	1.6728
		COV	97.9%	97.1%	96.5%	97.8%	96.9%	99.0%	95.0%	94.5%	94.5%	95.9%	94.9%	99.0%
	100	Mean	0.9718	4.0021	5.0002	3.0234	1.4956	3.5038	0.9701	3.9970	5.0011	3.0231	1.4951	3.4113
		IM SE	0.1532	0.2340	0.2420	0.1962	0.0574	1.9372	0.1518	0.2183	0.2106	0.1779	0.0538	1.7694
		MC Sd	0.1326	0.2257	0.2360	0.1831	0.0539	1.5850	0.1348	0.2195	0.2081	0.1624	0.0501	1.4641
		COV	98.3%	95.8%	94.1%	97.2%	96.7%	99.0%	95.6%	94.2%	94.8%	95.5%	95.8%	98.3%
(1.0)	200	Mean	0.9944	4.0062	5.0047	3.0023	1.5006	3.3364	0.9894	4.0055	4.9920	3.0091	1.4989	3.2880
		IM SE	0.1060	0.1578	0.1633	0.1327	0.0412	1.1537	0.1063	0.1631	0.1577	0.1287	0.0387	1.1029
		MC Sd	0.1010	0.1544	0.1628	0.1320	0.0406	1.1513	0.1122	0.1619	0.1584	0.1307	0.0391	1.1454
		COV	96.1%	95.1%	95.4%	94.9%	94.8%	97.9%	92.9%	95.5%	94.8%	94.8%	94.3%	95.2%
	400	Mean	0.9914	3.9969	4.9978	3.0069	1.4988	3.1214	0.9912	4.0017	5.0072	3.0015	1.5000	3.1105
		IM SE	0.0735	0.1108	0.1122	0.0896	0.0276	0.7076	0.0748	0.1034	0.1040	0.0883	0.0269	0.6931
		MC Sd	0.0719	0.1133	0.1142	0.0894	0.0270	0.7149	0.0752	0.1027	0.1063	0.0894	0.0275	0.7008
		COV	95.5%	94.0%	93.9%	94.3%	95.4%	96.9%	94.5%	95.8%	94.4%	95.3%	94.6%	95.6%

4.2 Experiment 2: Asymptotic properties

Our aim here is to demonstrate the asymptotic properties of the EM estimates. We chose sample sizes $n = 100{*}k$, $k = 1, \ldots, 10$, from Model M2 with the choice of parameters $\alpha = 0.5$, $\boldsymbol{\beta} = (4, 5, 3, 1.5)^\top$ and $\lambda = 3$. For each combination of parameters and sample size, we generated 1000 random samples from Model M2 under sinh-ST ($\nu = 3$), sinh-SSL ($\nu = 3$) and sinh-SCN ($\nu = 0.1, \gamma = 0.1$) models. From the EM estimates of model parameters determined, we computed the root mean square error (RMSE), the bias (Bias) and the Monte Carlo standard deviation (MC-SD) for each parameter over the 1000 replicates. If θ_k is a component of $\boldsymbol{\theta}$, then these quantities are defined, respectively, by

$$\text{MC-SD}^2(\widehat{\theta}_k) = \frac{1}{m-1}\sum_{j=1}^{m}\left(\widehat{\theta}_k^{(j)} - \overline{\widehat{\theta}_k}\right)^2 \quad \text{and} \quad \text{RMSE}(\widehat{\theta}_k) = \sqrt{\text{MC-Sd}^2(\widehat{\theta}_k) + \text{Bias}^2(\widehat{\theta}_k)},$$

where $\text{Bias}(\widehat{\theta}_k) = \overline{\widehat{\theta}_k} - \theta_k$, $\overline{\widehat{\theta}_k} = \sum_{j=1}^{m}\widehat{\theta}_k^{(j)}/m$ is the Monte Carlo mean and $\theta_k^{(j)}$ is the estimate of θ_k from the jth sample, with $j = 1, \ldots, m$. Figures 3 and 4 present the Bias and RMSE of the ML estimates of the parameters α, $\boldsymbol{\beta}$, and λ. It is observed that the Bias and RMSE become smaller as the sample size n increases, as one would expect. We also observe that the Bias of the ML estimate of λ are relatively larger than the Bias of the ML estimates of α and $\boldsymbol{\beta}$.

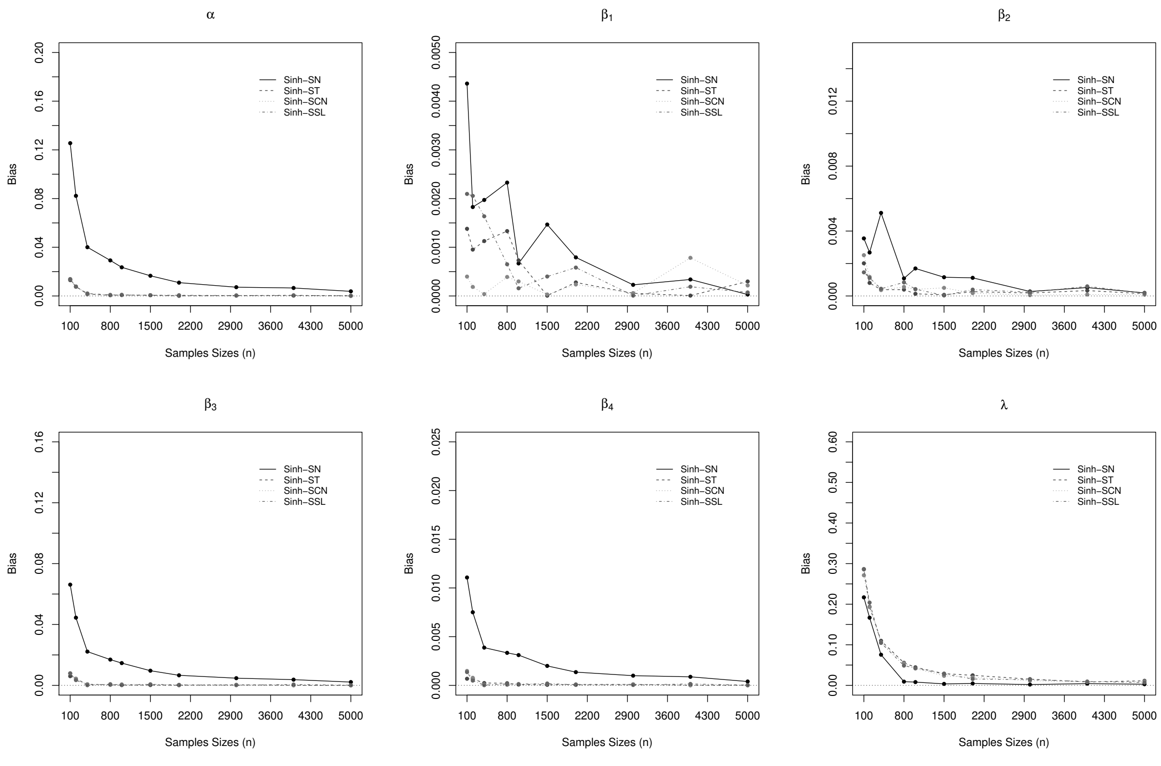


Figure 3: Bias of parameters α, β_1, β_2, β_3, β_4, and λ under sinh-SNI model.

5 Illustrative examples

In this section, we fit the proposed models to two datasets analized earlier by Rieck and Nedelman (1991) and Lin (1994). These examples are used to illustrate the flexibility

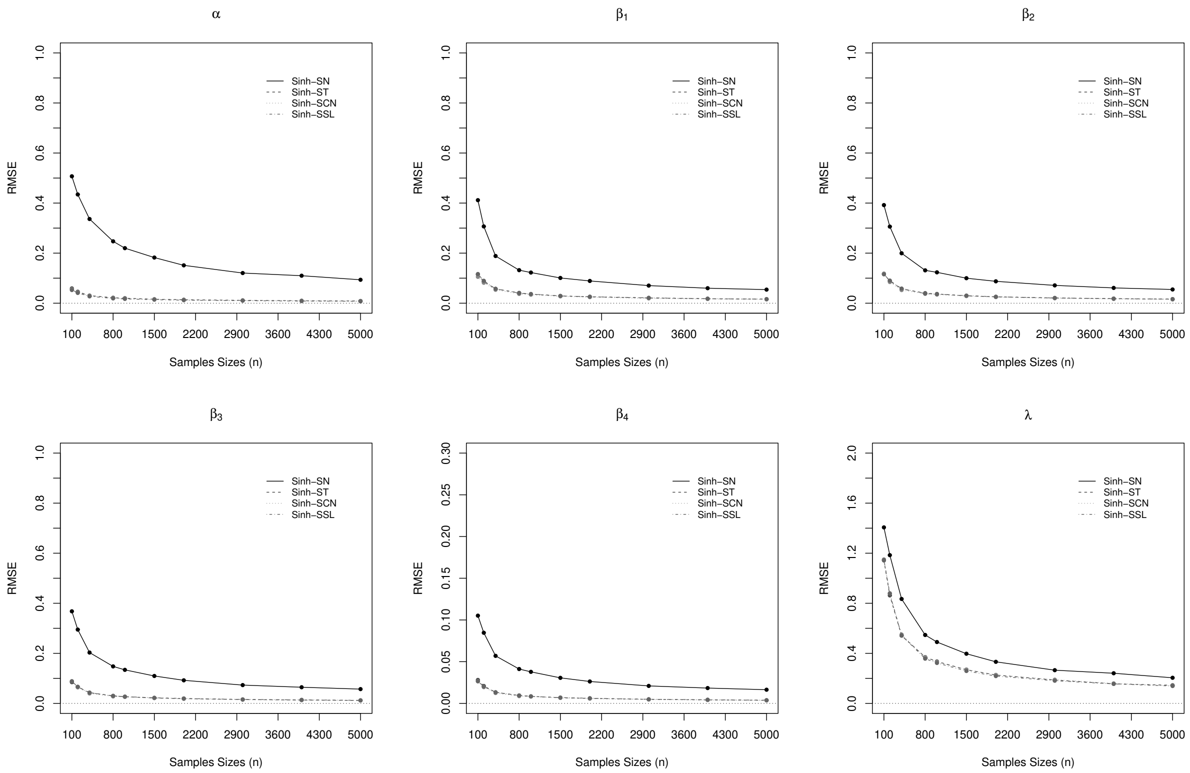


Figure 4: RMSE of MLEs of parameters α, β_1, β_2, β_3, β_4, and λ under sinh-SNI model.

of the proposed linear and nonlinear regression models under sinh-SN, sinh-ST sinh-SCN and sinh-SSL models. Standard errors (SE) are obtained by the method outlined in Section 3.2. The goodness-of-fit is evaluated through

$$d_i = d(Y_i, \widehat{\boldsymbol{\theta}}) = \xi_2^2(Y_i; \widehat{\alpha}, \widehat{\mu}_i, 2)$$

mentioned earlier in Part (iii) of Remark 1, where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$. For model selection criteria, we computed the log-likelihood values $\ell(\boldsymbol{\theta})$, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the efficient determination criterion (EDC).

5.1 Biaxial fatigue dataset

This dataset was first by analyzed by Rieck and Nedelman (1991) in linear regression model. These data were considered by Farias and Lemonte (2011) in a nonlinear regression model and Santana et al. (2011) in a BS linear regression model by using the skew-normal distribution. These data correspond to 46 metal pieces that were tested until failure. The variables considered in this study are the number of cycles until failure (N) and the work per cycle (in MJ / m^3). Lemonte and Cordeiro (2009) proposed the nonlinear regression model

$$Y_i = \beta_1 + \beta_2 \exp\left(-\beta_3/x_i\right) + \varepsilon_i, \quad i = 1, \ldots, 46, \tag{16}$$

where $Y_i = \log(N_i)$ is the logarithm of the number of cycles, x_i is the ith value of the work per cycle, β_1, β_2 and β_3 the regression coefficients, and $\varepsilon_i \sim$ Sinh-N$(\alpha, 0, \sigma = 2)$. Here, we consider $\varepsilon_i \sim$ Sinh-SNI$(\alpha, 0, \sigma = 2; H)$, based on SN-NLM, ST-NLM, SCN-NLM and SSL-NLM. Table 4 provides the ML estimates of the model parameters along with their standard errors (SE). Moreover, the log-likelihood function at the ML estimates and the information selection criteria such as AIC, BIC and EDC criteria (sinh-NI-NLM, in parentheses) are all reported. The criteria values indicate that the sinh-SNI-NLM models with heavy tails provide a significantly better fit than the sinh-NI-NLM model. Next, upon substituting the ML estimate of $\boldsymbol{\theta}$ in $d(Y_i)$, the goodness-of-fit presented in Figure 5 displays simulated envelopes (lines representing the 5th percentile, the mean and the 95th percentile of 200 simulated points for each observation). This figure provides strong evidence that the sinh-SNI-NLM provides a better fit to these data than the sinh-NI-NLM model, with the best model being the SN-NLM.

Table 4: Biaxial dataset: ML estimates for the four selected sinh-SNI-NLMs, the log-likelihood values and the information criteria values, with sinh-NI-NLMs (in parentheses).

Parameter	Sinh-SN		Sinh-ST		Sinh-SCN		Sinh-SSL	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
α	0.7026	0.1185	0.6673	0.1223	0.6110	0.1080	0.67350	0.11419
β_1	10.0779	1.0045	10.1018	0.9945	10.0910	0.9978	10.07918	1.00373
β_2	-5.6500	0.6608	-5.6817	0.6667	-5.6675	0.6632	-5.65171	0.66098
β_3	-18.8645	7.3732	-18.7276	7.1313	-18.7920	7.2333	-18.85769	7.35884
λ	-6.9827	8.3388	-6.4896	7.4741	-6.7127	7.8613	-6.95720	8.29373
ν	-	-	16.4	-	0.5	-	12.1	-
γ	-	-	-	-	0.6	-	-	-
$l(\widehat{\boldsymbol{\theta}})$	-19.3556 (-22.3862)		-19.8041 (-22.8317)		-19.5798 (-22.5373)		-19.3811 (-22.2623)	
AIC	48.7112 (52.7724)		49.6082 (53.6635)		49.1596 (53.0746)		48.7622 (52.5246)	
BIC	57.8544 (60.0869)		58.7515 (60.9781)		58.3028 (60.3891)		57.9054 (59.8391)	
EDC	45.4935 (50.1982)		46.3906 (51.0894)		45.9419 (50.5004)		45.5445 (49.9504)	

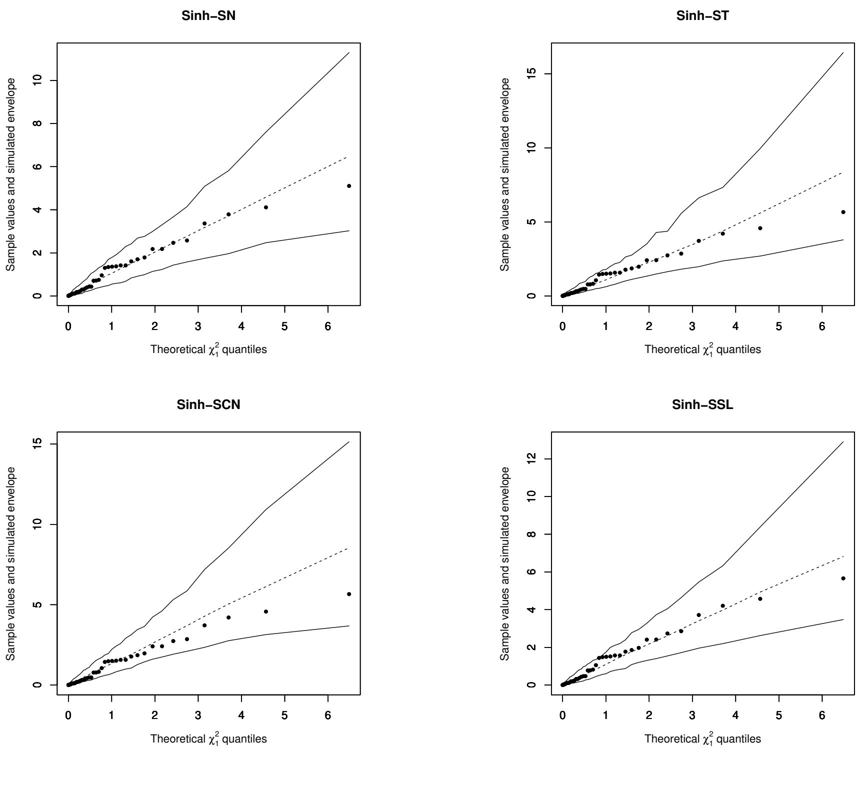
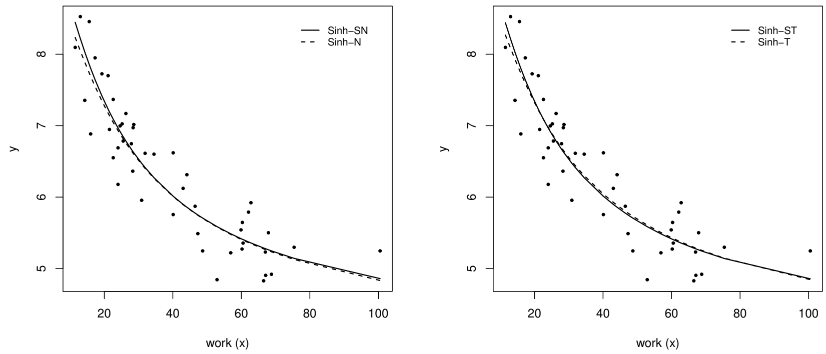


Figure 5: Biaxial dataset: Simulated envelopes for sinh-SNI-NLM based on SN, ST, SCN and SSL distributions.

Figure 6 displays the scatter plots of the dataset and the fitted nonlinear models based on SNI and NI distributions. From this figure, we once again see strong evidence for the sinh-SNI-NLM model providing a better fit to these data.



(a) (b)

Figure 6: Biaxial dataset: Scatterplots and fitted lines based on sinh-SN-NLM and sinh-ST-NLM models.

5.2 Martin Marietta dataset

Here, we consider the Martin Marietta dataset analyzed by Ferreira et al. (2015) and Butler et al. (1990). This data set was also studied by Azzalini and Capitanio (2003) under asymmetrical models. On the basis of arguments presented by them, we consider here the linear regression

$$Y_i = \beta_0 + \beta_1 \mathrm{CRSP}_i \varepsilon_i, \quad i = 1, \ldots, 60, \tag{17}$$

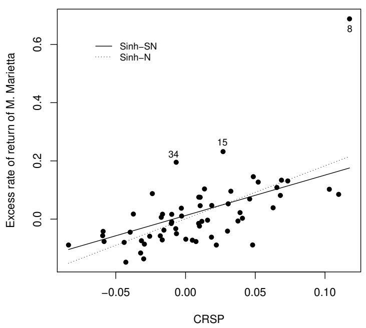
where the response Y is the excess rate of return of the Martin Marietta company and CRSP is an index of the excess rate of return for the New York market as a whole. We assume that the error terms are distributed as $\varepsilon_i \sim \text{sinh-SNI}(\alpha, 0, \sigma = 2; H)$. Data over a period of $n = 60$ consecutive months are available.

Table 5 displays the ML estimates of the model parameters under the sinh-SN, sinh-ST, sinh-SCN and sinh-SSL models and their corresponding standard errors (SE). For model comparison, we computed the $\ell(\boldsymbol{\theta})$ values, and the AIC, BIC and EDC criteria (sinh-NI model, in parentheses) values. The criteria values reveal that the regression model under sinh-SNI models with heavy tails present a significantly better fit than the symmetric sinh-NI models. We also note that the fit under the sinh-SCN and sinh-SSL models are significantly better than the sinh-SN model.

Table 5: Martin Marietta dataset: Results from fitting four selected sinh-SNI models (information criteria values under sinh-NI model, in parentheses).

Parameter	Sinh-SN		Sinh-ST		Sinh-SCN		Sinh-SSL	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
α	0.1373	0.0132	0.0625	0.0138	0.0670	0.0178	0.0493	0.0121
β_0	-0.0937	0.0143	-0.0472	0.0210	-0.0484	0.0264	-0.0476	0.0237
β_1	1.3891	0.3839	1.2455	0.2138	1.2665	0.2184	1.2631	0.2182
λ	3.9496	2.0794	1.0841	0.8518	1.0594	0.9849	1.0717	0.9193
ν	-	-	3.1819	-	0.0836	-	1.1903	-
γ	-	-	-	-	0.0526	-	-	-
$\ell(\widehat{\boldsymbol{\theta}})$	65.7676 (57.2889)		73.0699 (71.8111)		74.0996 (72.9468)		73.6296 (72.4625)	
AIC	-123.5353 (-108.5779)		-138.1398 (-137.6221)		-140.1992 (-139.8936)		-139.2592 (-138.925)	
BIC	-115.1579 (-102.2948)		-129.7625 (-131.3391)		-131.8218 (-133.6106)		-130.8818 (-132.642)	
EDC	-125.3385 (-109.9303)		-139.9431 (-138.9746)		-142.0024 (-141.2461)		-141.0624 (-140.2775)	

Figure 7 displays the scatter plots of the dataset and the fitted linear models based on the sinh-SN and sinh-SCN models. These fitted linear models are similar to those obtained by Ferreira et al. (2015). From this figure, we can note some possible outlier observations, such as #8, #15 and #34, under the fitted models. But, the effect of these observations are attenuated through the use of the sinh-SNI distributions as we will see later on.



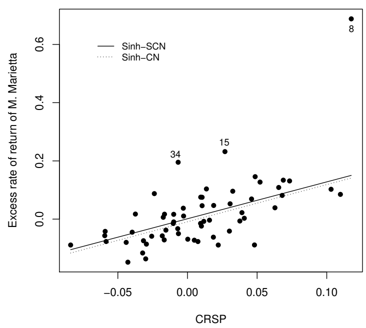


Figure 7: Martin Marietta dataset: Scatterplots and fitted lines under the sinh-SN (left) and sinh-SCN (right) models.

In addition, Figure 8 displays simulated envelopes (lines representing the 5th percentile, the mean and the 95th percentile of 200 simulated points for each observation) based on d_i values which once again strongly supports that the models with heavy tails produce a better fit to the current data, with the sinh-SCN and sinh-SSL models being significantly better.

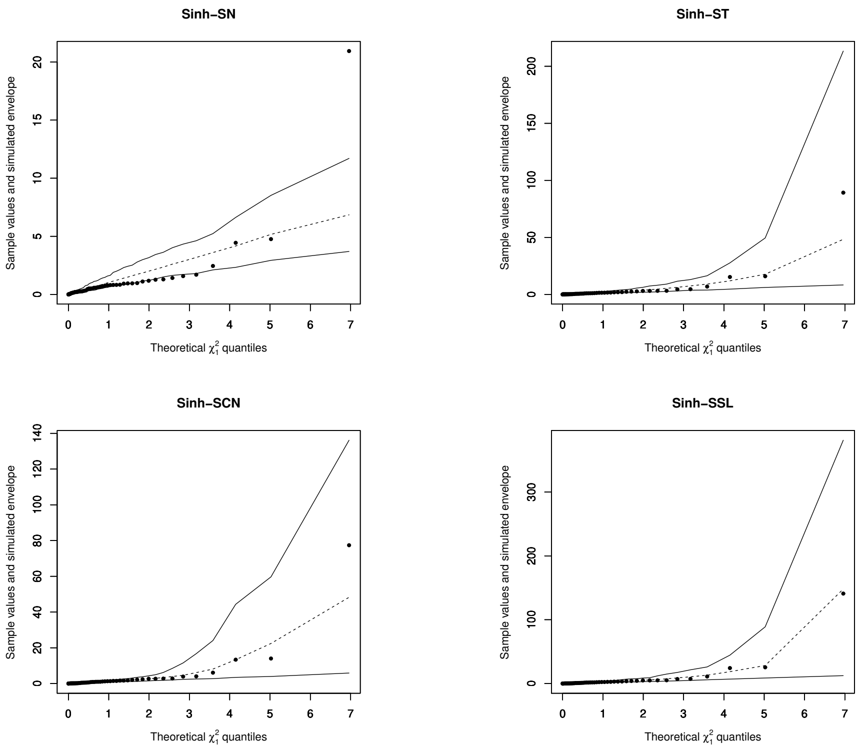


Figure 8: Martin Marietta dataset: Simulated envelopes under the four proposed models.

To identify outlying observations, we use the Mahalanobis distance d_i. Figure 9 displays such distances for the four fitted models, where the cutoff lines correspond to the quantile $\tau = 0.95$. From the figure, we note that the observations #8, #15 and #34 are detected as possible outliers under the fitted models. From the EM-algorithm, the estimated weights ($\widehat{\kappa}_i$, $i = 1, \ldots, 60$) for these observations are the smallest ones for all fitted asymmetric models with heavier tails as can be seen in Figure 10, thus confirming the robustness aspects of the ML estimates against outlying observations of the heavy-tailed sinh-SNI models, where segmented lines correspond to the sinh-normal model in which case $\kappa_i = 1$, for all i. Here, we note that larger d_i values imply a smaller κ_i values, and so in the process of estimation of $\boldsymbol{\theta}$, little weight is given to the outliers. To reveal the impact of the detected influential observations, we follow the suggestion of Lee et al. (2006) and use the quantities TRC and MRC defined as follows:

$$\mathrm{TRC} = \sum_{j=1}^{n_p} \left| \frac{\widehat{\theta}_j - \widehat{\theta}_{[i]j}}{\widehat{\theta}_j} \right| \quad \text{and} \quad \mathrm{MRC} = \max_{j=1,\ldots,n_p} \left| \frac{\widehat{\theta}_j - \widehat{\theta}_{[i]j}}{\widehat{\theta}_j} \right|, \tag{18}$$

where n_p is the dimension of $\boldsymbol{\theta}$ and the subscript $[i]$ stands for the ML estimate of $\boldsymbol{\theta}$ when the ith observation is deleted. Table 6 shows the comparisons of these measures for the fitted models without the observation #8 where the biggest change takes place under the Sinh-SN distribution. As expected, the results indicate that the ML estimates are less sensitive in the presence of atypical data when distributions with heavier tails are assumed than the sinh-SN model.

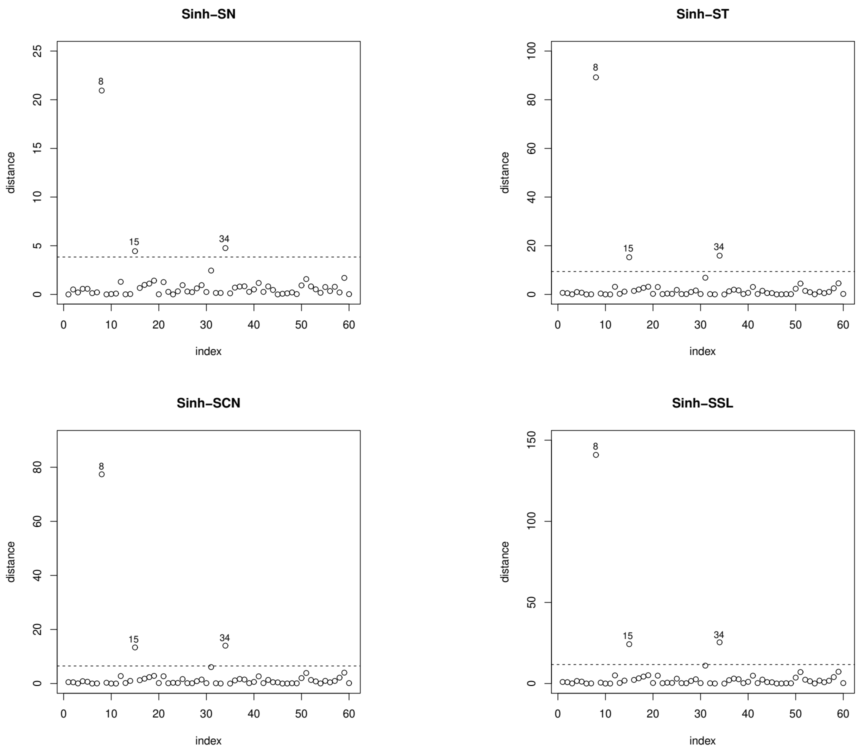


Figure 9: Martin Marietta dataset: Index plots of the distances (d_i) for the fitted models.

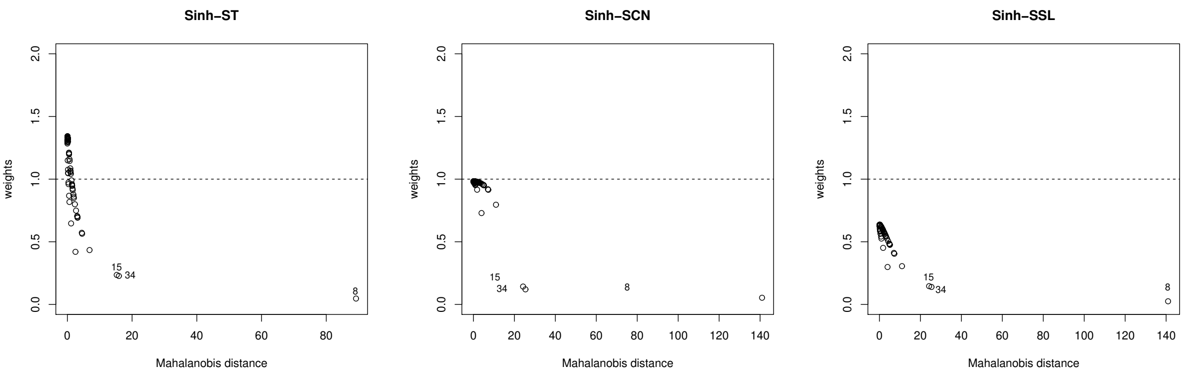


Figure 10: Martin Marietta dataset: Estimated κ_i for the ST, SCN and SSL models.

Table 6: Martin Marietta dataset: Comparison of the relative changes in the ML estimates in term of TRC and MRC for the four selected sihn-SNI models.

	TRC	MRC
Sinh-SN	11.6174	9.8691
Sinh-ST	2.3682	1.2204
Sinh-SCN	1.9838	0.9829
Sinh-SSL	2.3232	1.2041

6 Concluding remarks

Nonlinear regression analysis is a very useful methodology that has been used extensively in several fields, and this paper provides useful results that can be used in survival and reliability studies when asymmetric distributions are assumed. Specifically, we first introduce a general class of generalized sinh-Normal distributions, called sinh-SNI distributions, that is based on the skew-normal/Independent distribution instead of the normal distribution. So, this proposal generalizes some results presented earlier by Rieck and Nedelman (1991), Balakrihnan et al. (2009) and Vilca et al. (2015), among others. Some special cases of this general model are the asymmetric sinh-SN, sinh-ST, sinh-SCN and sinh-SSL distributions, and an important case is the log-Birnbaum-Saunders distribution that has been used in the linear regression model setup. Second, we propose Sinh-SNI nonlinear regression models which generalize linear and nonlinear regression models that have been developed so far by considering symmetric generalized sinh-normal distributions developed by some authors including Rieck and Nedelman (1991), Lemonte and Cordeiro (2009), Paula et al. (2011) and Vilca et al. (2015). Furthermore, we have presented an EM-algorithm for the ML estimation of model parameters, in which the CM-Step provides closed-form expressions for the estimates of α and λ that allow us to obtain qualitatively robust ML estimates in an efficient manner, as seen in the numerical illustrations. This generalization results in making the inference robust to the presence of outlying observations, as demonstrated in Section 5, and is quite a flexible and useful distribution for modeling purposes.

References

Andrews, D. F.; Mallows, C. L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc., Ser. B* 36, 99-102.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* 12(2), 171-178.

Azzalini, A.; Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. Roy. Statist. Soc., Ser. B* 65, 367-389.

Balakrishnan, N.; Leiva, V.; Sanhueza, A.; Vilca, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Stat. Oper. Res. T.* 33, 171-192.

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

Birnbaum, Z. W.; Saunders, S. C. (1969a). A new family of life distributions. *J. Appl. Prob.* 6, 637-652.

Birnbaum, Z. W.; Saunders, S. C. (1969b) Estimation for a family of life distributions with applications to fatigue. *J. Appl. Prob.* 6, 328-347.

Branco, M. D.; Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* 79, 99-113.

Butler, R. L.; McDonald, J. B.; Nelson, R. D.; White, S. B. (1990). Robust and partly adaptive estimation of regression models. *Rev. Econ. Statist.* 72, 321-327.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc., Ser. B* 39, 1-38.

Farias, R. B. A.; Lemonte, A. J. (2011). Bayesian inference for the Birnbaum-Saunders nonlinear regression model. *Stat. Methods Appl.* 20, 423-438.

Ferreira, C. S.; Lachos, V. H.; Bolfarine, H. (2015). Inference and diagnostics in skew scale mixtures of normal regression models. *J. Stat. Comput. Simul.* 85(3), 517-537.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* 36, 146-176.

Johnson, N. L.; Kotz, S; Balakrishnan, N. (1994). *Continuous Univariate Distributions-Vol. 1*, Second edition. John Wiley & Sons, New York.

Lachos, V. H.; Ghosh, P.; Arellano-Valle, R. B. (2010). Likelihood based inference for skew normal independent linear mixed models. *Stat. Sin.* 20, 303-322.

Lange, K. L.; Little, J. A.; Taylor, M. G. J. (1989). Robust statistical modeling using the t distribution. *J. Amer. Stat. Assoc.* 84, 881-896.

Lange, K.; Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *J. Comp. Graph. Stat.* 2, 175-198.

Lee, S. Y.; Lu, B.; Song, X. Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Comp. Stat. Data Anal.* 5, 1356-1377.

Leiva, V.; Vilca, F.; Balakrishnan, N.; Sanhueza, A. (2010). A skewed Sinh-Normal distribution and its properties and application to air pollution. *Comm. Stat. Theor. Meth.* 39, 426-443.

Lemonte, A. J.; Cordeiro, G. M. (2009). Improved maximum likelihood estimation in Birnbaum-Saunders nonlinear regressions. *Comp. Stat. Data Anal.* 53, 4441-4452.

Lemonte, A. J.; Patriota, A. G. (2011). Influence diagnostics in Birnbaum-Saunders nonlinear regression models. *J. Appl. Stat.* 38, 871-884.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Stat. Med.* 13, 2233-2247.

Lin, T. I.; Lee, J. C.; Yen, S. Y. (2007). Finite mixture modelling using the skew normal distribution. *Stat. Sin.* 17, 909-927.

Maehara, R. P.; Vilca, F.; Bolfarine, H. (2017). Modeling extreme percentiles in Birnbaum-Saunders distributions based on an asymmetric heavy-tailed distribution. *Under review.*

Paula, G. A.; Leiva, V.; Barros, M.; Liu, S. (2011). Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance. *Appl. Stoc. Models Bus. Ind.* 28, 16-34.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling.* Marcel Dekker, New York.

Rieck, J. R. (1989). *Statistical Analysis for the Birnbaum-Saunders Fatigue Life Distribution.* Ph.D dissertation, Clemson University, South Carolina.

Rieck, J. R.; Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics* 33, 51-60.

Santana, L.; Vilca, F.; Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *J. Appl. Stat.* 38, 1633-1649.

Spiegelhalter, D. J.; Best, N. G.; Carlin, B. P.; Van Der Linde, A. (2002). Bayesian measures of complexity and fit. *J. Roy. Statist. Soc., Ser. B* 64, 1-34.

Vilca, F.; Santana, L.; Leiva, V.; Balakrishnan, N. (2011). Estimation of extreme percentile in Birnbaum-Saunders distributions. *Comp. Stat. Data Anal.* 55, 1665-1678.

Vilca, F.; Zeller, C. B.; Cordeiro, G. M. (2015). The sinh-normal/independent nonlinear regression model. *J. Appl. Stat.* 42, 1659-1676.

Vilca, F.; Azevedo, C. L. N.; Balakrishnan, N. (2017): Bayesian inference for sinh-normal/independent nonlinear regression models. *J. Appl. Stat.* 44, 2052-2074.

Appendix A: Special cases of sinh-SNI models

i) *The sinh-SCN distribution*
This distribution is based on the SCN model, which is called sinh-skew-contaminated normal (sinh-SCN) distribution. Here, U is a discrete random variable with pdf $h(u;\nu,\gamma)=\nu\mathbb{I}_{\{\gamma\}}(u)+(1-\nu)\mathbb{I}_{\{1\}}(u)$, with $0<\nu<1, 0<\gamma<1$, where $\mathbb{I}_A(\cdot)$ denotes the indicator function of the set A. Then, the pdf of Y is

$$f_Y(y) \;=\; \frac{2}{\sigma}\Big[\nu\phi\big(\xi_{2y};0,\frac{1}{\gamma}\big)\Phi\big(\gamma^{1/2}\,\lambda\xi_{2y}\big)+(1-\nu)\phi\big(\xi_{2y}\big)\Phi\big(\lambda\xi_{2y}\big)\Big]\,\xi_{1y}.$$

This distribution is denoted by $Y\sim$ Sinh-SNC$(\alpha,\mu,\lambda;\nu,\gamma)$. In this case,

$$\kappa_r=\frac{2\,\xi_{1y}}{\sigma\, f_Y(y)}\left[\nu\gamma^r\phi\left(\xi_{2y};0,\gamma^{-1}\right)\Phi\left(\gamma^{1/2}\,\lambda\xi_{2y}\right)+(1-\nu)\phi\left(\xi_{2y}\right)\Phi\left(\lambda\xi_{2y}\right)\right],$$

$$\tau_r=\frac{2\,\xi_{1y}}{\sigma\, f_Y(y)}\left[\nu\gamma^{r/2}\phi\left(\xi_{2y};0,\gamma^{-1}\right)\Phi\left(\gamma^{1/2}\,\lambda\xi_{2y}\right)+(1-\nu)\phi\left(\xi_{2y}\right)\Phi\left(\lambda\xi_{2y}\right)\right].$$

Finally, $d(Y)$ has its cdf as $Pr(d(Y)\le\upsilon)=\nu Pr(\chi^2_1\le\gamma\upsilon)+(1-\nu)Pr(\chi^2_1\le\upsilon)$;

ii) *The sinh-SSL distribution*
In this case, $U\sim Beta(\nu,1)$ distribution, and the resulting distribution is called the sinh-skew-slash (sinh-SSL) distribution, denoted by $Y\sim$ Sinh-SSL$(\alpha,\mu,\lambda;\nu)$, and its pdf is given by

$$f_Y(y)=\frac{2\,\nu}{\sigma}\int_0^1 u^{\nu-1}\phi(\xi_{2y};0,1/u)\Phi(u^{1/2}\lambda\xi_{2y})du\,\xi_{1y}. \tag{19}$$

In this case,

$$\kappa_r=\frac{2^{\nu+r+1}\nu\,\Gamma\left(\frac{2\nu+2r+1}{2}\right)\,\xi_{1y}}{f_Y(y)\sqrt{\pi}\,\sigma}P_1\Big(\frac{2\nu+2r+1}{2},\frac{\xi^2_{2y}}{2}\Big)\xi_{2y}^{-(2\nu+2r+1)}\mathrm{E}\left[\Phi\left(S^{1/2}\lambda\xi_{2y}\right)\right],$$

$$\tau_r=\frac{2^{\nu+r/2+1/2}\,\nu\,\Gamma\left(\frac{2\nu+r+1}{2}\right)\,\xi_{1y}}{f_Y(y)\sqrt{\pi}^2\,\sigma}\left(\xi^2_{2y}+\lambda^2\xi^2_{2y}\right)^{-\frac{2\nu+r+1}{2}}P_1\left(\frac{2\nu+r+1}{2},\frac{\xi^2_{2y}+\lambda^2\xi^2_{2y}}{2}\right),$$

where $P_x\left(a,b\right)$ denotes the cdf of the $Gamma(a,b)$ distribution evaluated at x and $S\sim Gamma\left(\frac{2\nu+2r+1}{2},\frac{\xi^2_{2y}}{2}\right)\mathbb{I}_{(0,1)}$, a truncated gamma distribution on $(0,1)$. Finally, the cdf of $d(Y)$ is $Pr(d(Y)\le\upsilon)=Pr(\chi^2_1\le\upsilon)-\dfrac{2^\nu\Gamma(\nu+1/2)}{\upsilon^\nu\Gamma(1/2)}Pr(\chi^2_{2\nu+1}\le\upsilon)$;

iii) *The sinh-ST distribution*
This distribution is obtained when $U\sim Gamma(\nu/2,\nu/2)$, and the resulting distribution is called the sinh-skew-Student-t (sinh-ST) distribution. The pdf of Y is

$$f_Y(y)=\frac{2\,t}{\sigma}(\xi_{2y};\nu)T\left(\sqrt{\frac{\nu+1}{\xi^2_{2y}+\nu}}\,\lambda\xi_{2y};\nu+1\right)\xi_{1y}, \tag{20}$$

where $t(.;\nu)$ and $T(.;\nu)$ denote, respectively, the pdf and cdf of the standard Student-t distribution. In this case, the conditional expectations are given by

$$\kappa_r=\frac{2^{r+1}\,\nu^{\nu/2}\Gamma\left(\frac{\nu+2r+1}{2}\right)\left(\xi^2_{2y}+\nu\right)^{-\frac{\nu+2r+1}{2}}\,\xi_{1y}}{f_Y(y)\Gamma(\nu/2)\sqrt{(\pi)}\,\sigma}T\left(\sqrt{\frac{\nu+2r+1}{\xi^2_{2y}+\nu}}\,\lambda\xi_{2y};\nu+2r+1\right)$$

and

$$\tau_r = \frac{2^{(r+1)/2}\, \nu^{\nu/2}\, \Gamma\left(\frac{\nu+r+1}{2}\right) \left(\xi_{2y}^2 + \nu + \lambda^2 \xi_{2y}^2\right)^{-\frac{\nu+r+1}{2}} \xi_{1y}}{f_Y(y)\Gamma\left(\nu/2\right) \sqrt{(\pi)}^2\, \sigma}.$$

Finally, we have $d(Y) \sim F(1, \nu)$.

Appendix B: Derivatives of ξ_{1i} and ξ_{2i}^2

$$\begin{aligned}
\frac{\partial \xi_{i1}}{\partial \alpha} &= -\frac{1}{\alpha}\, \xi_{1i},\ \frac{\partial \xi_{1i}}{\partial \boldsymbol{\beta}} = -\frac{1}{2}\, \xi_{2i}\, \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial \xi_{2i}^2}{\partial \alpha} = -\frac{2}{\alpha} \xi_{2i}^2,\ \frac{\partial \xi_{i2}^2}{\partial \boldsymbol{\beta}} = -\xi_{1i}\, \xi_{2i}\, \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}, \\
\frac{\partial^2 \xi_{1i}}{\partial \alpha \partial \alpha} &= \frac{2}{\alpha^2}\, \xi_{1i}, \quad \frac{\partial^2 \xi_{1i}}{\partial \alpha \partial \boldsymbol{\beta}^\top} = \frac{1}{4}\, \xi_{2i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^\top}, \quad \frac{\partial^2 \xi_{1i}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \frac{1}{4}\, \xi_{1i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^\top} - \frac{1}{2}\, \xi_{2i} \frac{\partial^2 \mu_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \xi_{2i}^2}{\partial \alpha} &= -\frac{2}{\alpha} \xi_{2i}^2,\ \frac{\partial \xi_{i2}^2}{\partial \boldsymbol{\beta}} = -\xi_{1i}\, \xi_{2i}\, \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}, \\
\frac{\partial^2 \xi_{2i}^2}{\partial \alpha \partial \alpha} &= \frac{6}{\alpha^2} \xi_{2i}^2, \quad \frac{\partial^2 \xi_{i2}^2}{\partial \alpha \partial \boldsymbol{\beta}^\top} = \frac{2}{\alpha}\, \xi_{1i}\, \xi_{2i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^\top}, \\
\frac{\partial^2 \xi_{1i}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{1}{2} \left(\xi_{1i}^2 + \xi_{2i}^2\right) \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^\top} - \xi_{1i}\, \xi_{2i}\, \frac{\partial^2 \mu_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}.
\end{aligned}$$