REVSTAT – Statistical Journal
Volume 19, Number 4, October 2021, 553–573

A REGRESSION MODEL FOR POSITIVE DATA BASED ON THE SLASHED HALF-NORMAL DISTRIBUTION

Authors: Yolanda M. Gómez
– Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile
yolanda.gomez@uda.cl

Diego I. Gallardo
– Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile
diego.gallardo@uda.cl

Mário de Castro
– Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brazil
mcastro@icmc.usp.br

Received: September 2018 Revised: September 2019 Accepted: September 2019

Abstract:

- In this paper, we discuss several aspects about the slashed half-normal distribution. We reparameterize the model based on the mean and we perform comparisons with well-known regression models for positive data. Maximum likelihood estimation of the parameters is carried out through the *expectation-maximization* algorithm. Some properties of the estimators and two kinds of residuals are assessed in a simulation study. Two real datasets illustrated the proposed model as well as other three models for the sake of comparison.

Keywords:

- *EM algorithm; gamma distribution; half-normal distribution; slashed distribution.*

AMS Subject Classification:

- 62F86, 60E05.

1. INTRODUCTION

The half-normal (HN) distribution is a very important model in the study of skewed distributions. For instance, it is used in the stochastic representation of the skew-normal distribution in Azzalini [4, 5] and Henze [15]. Several papers in the literature have paid attention to the half-normal distribution. For instance, Chou and Liu [7] studied its properties and its uses in quality control. Pewsey [22, 23] studied asymptotic inference and maximum likelihood estimation for the general location-scale half-normal distribution. For analysis and applications from a Bayesian point of view, the reader is referred to Wiper *et al.* [32] and Khan and Islam [17]. Also, the `hnp` R package [20], generates half-normal plots with simulated envelopes using different diagnostics tools from a range of different fitted models. Even though the HN distribution accommodates only decreasing hazard rates, this distribution has been used to model positive data and is becoming an important model in reliability theory,. Some of the generalizations of this distribution can be found in Cooray and Ananda [8], Cordeiro *et al.* [9], Olmos *et al.* [21], Gómez and Bolfarine [13], Bourguignon *et al.* [6] and Asgharzadeh *et al.* [1], among others. Particularly, we focus on the extension proposed in Olmos *et al.* [21], named slash half-normal (SHN) distribution, where the goal is to increase the kurtosis with respect to its parent half-normal distribution, and hence be more useful for modeling positive datasets that may have a heavy right tail. In this work, we propose a reparameterization for this model based on the mean. We use this parameterization because it is convenient for proposing a regression model.

The article is organized as follows. In Section 2, we describe the reparametrized SHN regression model and compare it with some existing models. In Section 3, we describe parameter estimation by the maximum likelihood (ML) method using the *expectation-maximization (EM)* algorithm. Goodness of fit through residuals is discussed in Section 4. In Section 5, we carry out two simulation studies to assess the performance of the proposed estimators and the two kinds of residuals. In Section 6, we apply the proposed model to analyze two datasets on the diet of the hunter-gatherer and concentration of minerals in soil samples. Concluding remarks are given in Section 7.

2. THE PROPOSAL

In this section, we present the proposed reparameterization for the SHN model in terms of the mean. We also present three common distributions to accommodate positive data that also are reparametrized in terms of the mean: the gamma, Weibull and Birnbaum–Saunders models.

2.1. Reparametrized slashed half-normal model

The SHN model (Olmos *et al.* (2012) [21]) is built in the following way. If $X \sim \mathrm{HN}(\sigma)$ $(\sigma > 0)$ and $Z \sim \mathrm{Beta}(\alpha, 1)$ are independent random variables, then

$$Y = \frac{X}{Z} \sim \mathrm{SHN}(\sigma, \alpha)\,, \tag{2.1}$$

where $\alpha > 0$ is a shape parameter that mainly controls the *right* tail of the distribution. Lower values of α $(0 < \alpha < 1)$ lead to a heavier tail (see Figure 1 in Olmos *et al.* [21]). However, in practice we have found estimates for α greater than 1 (see the two examples in Olmos *et al.* [21] and our applications). For this reason, the potential advantages of the parameterization of the model in terms of the mean (mainly related to the interpretation of the coefficients in a regression model) justify the restriction $\alpha > 1$. Such kind of restriction is not uncommon in the literature. Without going further, the popular Student's t distribution has a finite mean if the degrees of freedom are greater than 1. We propose a reparameterization of the SHN model based on $\mu = \sqrt{2/\pi}\,\alpha\sigma/(\alpha-1)$. The probability density function of the reparametrized SHN, henceforth RSHN(μ,α), is given by

$$(2.2)\qquad f_{\mathrm{RSHN}}(y;\mu,\alpha) \;=\; \alpha\sqrt{\frac{2^\alpha}{\pi}}\left[\sqrt{\frac{\pi}{2}}\,\frac{\mu(\alpha-1)}{\alpha}\right]^{\alpha}\Gamma\!\left(\frac{\alpha+1}{2}\right) y^{-(\alpha+1)}\; G\!\left[\frac{\alpha^2 y^2}{\pi\,\mu^2\,(\alpha-1)^2},\,\frac{\alpha+1}{2}\right],$$

for $y > 0$, where $\Gamma(\cdot)$ denotes the gamma function and $G(y,a) = \int_0^y u^{a-1}e^{-u}\,du/\Gamma(a)$ is the cumulative distribution function (cdf) of the gamma distribution with rate parameter equal to 1. Based on results in Olmos *et al.* [21], we have $\mathbb{E}(Y) = \mu$, for $\alpha > 1$,

$$\mathrm{Var}(Y) \;=\; \frac{\mu^2}{2}\left[\pi-2+\frac{\pi}{\alpha(\alpha-2)}\right], \qquad \text{for}\quad \alpha > 2\,,$$

$$\sqrt{\nu_3} \;=\; \frac{\pi\,\sqrt{2(\alpha-2)}\,\left[\frac{4}{\pi}\,\alpha^2\,(\alpha-2)\,(\alpha-3)-(\alpha-1)^2\,(\alpha-4)\,(\alpha+1)\right]}{\sqrt{\alpha}\,(\alpha-3)\left[(\pi-2)\,\alpha\,(\alpha-2)+\pi\right]^{3/2}}, \qquad \text{for}\quad \alpha > 3\,,$$

and

$$\nu_4 \;=\; \frac{3\,\alpha\,(\alpha-2)^2\,(\alpha-3)\left[\pi^2\,(\alpha-1)^4-4\alpha^3\,(\alpha-4)\right]-4\pi\,\alpha^2\,(\alpha-1)^2\,(\alpha-2)\,(\alpha-4)\,(\alpha^2-3\alpha+8)}{\alpha^2\,(\alpha-3)\,(\alpha-4)\left[(\pi-2)\,\alpha(\alpha-2)+\pi\right]^2}\,,$$

for $\alpha > 4$, where $\sqrt{\nu_3}$ and ν_4 denote the skewness and kurtosis coefficients, respectively. Note that this parameterization is very convenient because the parameter μ is related only to the mean and the variance of the distribution.

2.2. Reparametrized gamma distribution

For $Y \sim \mathrm{RG}(\mu,\phi)$ (the gamma model parametrized in terms of the mean), we have

$$\mathbb{E}(Y) = \mu\,, \qquad \mathrm{Var}(Y) = \frac{\mu^2}{\phi}\,, \qquad \sqrt{\nu_3} = \frac{2}{\sqrt{\phi}} \qquad \text{and} \qquad \nu_4 = 3+\frac{6}{\phi}\,.$$

The RSHN model is a competing distribution for the gamma distribution because the coefficient of variation (cv), skewness and kurtosis coefficients do not depend on μ in both models. Figure 1(a) shows the values of ϕ in the RG(μ,ϕ) model and α in the RSHN(μ,α) model that lead to the same values of cv. Figure 1(b) displays the kurtosis coefficient for those pairs (ϕ,α) corresponding to the same value of cv. It is clear that the gamma model is more flexible in the sense that it allows to obtain any positive value for the cv, whereas the RSHN distribution only supports values for cv greater than $\left[(\pi-2)/2\right]^{1/2} \approx 0.756$, i.e., greater than the cv of the half-normal distribution. However, there is a range of values of α such that, for the same value of the cv, the RSHN distribution has a greater kurtosis coefficient than the gamma distribution. In short, in the RSHN model the variance is proportional to the square of the mean (similar to the gamma model), but the RSHN model has a greater kurtosis coefficient for a certain range of values of α.

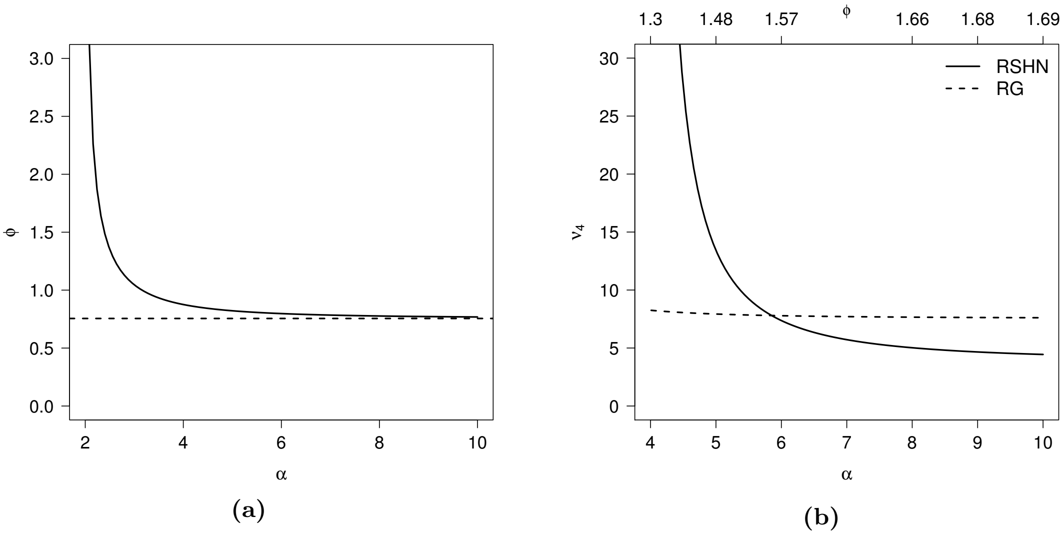


Figure 1: (a) Values for ϕ and α in the $\mathrm{RG}(\mu,\phi)$ and $\mathrm{RSHN}(\mu,\alpha)$ distributions that produce the same coefficient of variation and (b) their respective kurtosis coefficients.

2.3. Reparametrized Weibull and Birnbaum–Saunders distributions

The reparametrized form of the Weibull distribution with parameters $\mu > 0$ and $\delta > 0$ has probability density function

$$f_{\mathrm{RW}}(y;\mu,\delta) = \frac{\phi}{\gamma}\left(\frac{y}{\gamma}\right)^{\delta-1}\exp\left[-\left(\frac{y}{\gamma}\right)^{\delta}\right], \qquad \text{for} \quad y > 0\,,$$

where $\gamma = \mu\big/\Gamma(1/\delta+1)$, so that

$$\mathbb{E}(Y) = \mu \qquad \text{and} \qquad \mathrm{Var}(Y) = \mu^2\left\{\frac{\Gamma(2/\delta+1)}{\left[\Gamma(1/\delta+1)\right]^2} - 1\right\}.$$

We denote as $Y \sim \mathrm{RW}(\mu,\delta)$.

In the same way, Santos-Neto *et al.* [31] also reparametrized the Birnbaum–Saunders distribution in terms of the mean. With parameters $\mu > 0$ and $\xi > 0$, the probability density function is given by

$$f_{\mathrm{RBS}}(y;\mu,\xi) = \frac{\exp(\xi/2)\sqrt{\xi+1}}{4\sqrt{\pi\mu}\,y^{3/2}}\left(y + \frac{\xi\mu}{\xi+\mu}\right)\exp\left\{-\frac{\xi}{4}\left[\frac{y(\xi+1)}{\xi\mu} + \frac{\xi\mu}{y(\xi+1)}\right]\right\},$$

for $y > 0$, so that $\mathbb{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu^2(2\xi+5)/(\xi+1)^2$. We use the notation $Y \sim \mathrm{RBS}(\mu,\xi)$. The RW and RG (Section 2.2) will be compared with the RSHN model fitted to real datasets in Section 6.

Remark 2.1. The RG and RW models are more flexible than the RBS and RSHN models in the sense that, for a given value of μ, they allow to obtain any positive value for the variance, whereas the RBS and RSHN models have some restrictions. However, even when all the models produce the same mean and variance, the skewness and kurtosis are not the same. Moreover, such terms do not depend on μ. Table 1 shows four models with the same mean and variance. However, the skewness and kurtosis coefficients are different.

Table 1: Examples of models with the same mean and variance.

Moment or coefficient	Model			
	$\mathrm{RG}(\mu, 1.333)$	$\mathrm{RW}(\mu, 1.158)$	$\mathrm{RBS}(\mu, 3.692)$	$\mathrm{RSHN}(\mu, 4.125)$
Mean	μ	μ	μ	μ
Variance	$0.75\,\mu^2$	$0.75\,\mu^2$	$0.75\,\mu^2$	$0.75\,\mu^2$
Skewness	1.732	1.390	12.662	1.791
Kurtosis	7.500	6.868	59.641	120.807

Remark 2.2. The mean and the variance of the RG, RW, RBS and RSHN models are μ and $\mu^2 w^2(\eta)$, where η represents ϕ, δ, ξ or α in each model, respectively, and $w(\cdot)$ is a positive function representing the coefficient of variation. This function is presented in Table 2. The computational implementation to model mean and dispersion parameters with a set of covariates linked to both components in RG and RW models is implemented in the `gamlss.dist` package in `R` (see Rigby and Stasinopoulos [28, 29]), while the RBS model is discussed in Santos-Neto *et al.* [30]. A similar scheme to model mean and dispersion might be considered for the RSHN distribution. However, we only consider a model for the mean parameter in this work.

Table 2: Summary for some models with quadratic variance function.

Model	$\mathrm{RG}(\mu,\phi)$	$\mathrm{RW}(\mu,\delta)$	$\mathrm{RBS}(\mu,\xi)$	$\mathrm{RSHN}(\mu,\alpha)$
$w(\eta)$	$\dfrac{1}{\sqrt{\phi}}$	$\sqrt{\dfrac{\Gamma(2/\delta+1)}{[\Gamma(1/\delta+1)]^2}-1}$	$\dfrac{\sqrt{(2\xi+5)}}{\xi+1}$	$\sqrt{\dfrac{1}{2}\left(\pi-2+\dfrac{\pi}{\alpha(\alpha-2)}\right)}$

3. ESTIMATION

In this section, we discuss some details about the estimation procedure based on the ML method. We also consider an EM type algorithm to obtain a more stable estimation procedure. Henceforth, we consider a set of p observed covariates for each individual, say $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^\top$. Since $\mu = \mathbb{E}(Y)$ is a positive parameter, we adopt the logarithmic link function $\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, $i = 1, ..., n$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients.

3.1. General context

In Olmos *et al.* [21], parameter estimation (without covariates) was carried out based on the direct maximization of the log-likelihood function using as initial values the method of moments estimates of the parameters. In our model, assuming the intercept is included, naive estimators for β_0 and α can be obtained ignoring the covariates, i.e., $\beta_1 = \cdots = \beta_p = 0$. In this case, such estimators are given by

$$\widehat{\beta}_{0\mathrm{M}} = \log\bigl(\overline{Y}\bigr) \quad \text{and} \quad \widehat{\alpha}_{\mathrm{M}} = \frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{\pi}{2\,A_y - 2 + \pi}}\,, \qquad \text{if} \quad \overline{Y^2} > \frac{\pi}{2}\overline{Y}^2, \tag{3.1}$$

where $A_y = \overline{Y^2}/\overline{Y}^2$ and $\overline{Y^2}$ is the sample mean of the squared observations.

The log-likelihood function of $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \alpha)^\top$ in a random sample with observations $y_1, ..., y_n$ is given by

$$\ell(\boldsymbol{\psi}) \;=\; c(\alpha) + \alpha \log(\mu) - (\alpha+1)\sum_{i=1}^{n}\log(y_i) + \sum_{i=1}^{n}\log\left\{G\left[\frac{\alpha^2 y^2}{\pi\mu^2(\alpha-1)^2}, \frac{\alpha+1}{2}\right]\right\}, \tag{3.2}$$

where $c(\alpha) = -n(\alpha-1)\log(\alpha) - n\,\alpha\log(2)/2 + (\alpha-1/2)\log(\pi) + \alpha\log(\alpha-1) + n\log\bigl[\Gamma(\alpha/2+1/2)\bigr]$. However, direct maximization of (3.2) is not simple and may suffer from numerical instabilities. In Section 3.2, we propose a stable estimation procedure for this model based on the stochastic representation in (2.1). We develop in the sequel an EM algorithm (Dempster *et al.* [10]) for parameter estimation.

3.2. ECM and ECME algorithms

To facilitate the estimation process, we include latent variables $Z_1, ..., Z_n$ through the following hierarchical representation of the RSHN model:

$$Y_i\,|\,Z_i \;=\; z_i\,, \qquad \mu_i \sim \mathrm{HN}\left[\sqrt{\frac{\pi}{2}}\,\frac{\mu_i(\alpha-1)}{\alpha\,z_i}\right] \qquad \text{and} \qquad Z_i \sim \mathrm{Beta}(\alpha,1)\,.$$

Thus, the complete likelihood function for $\boldsymbol{\psi}$ is given by

$$L_{\mathrm{c}}(\boldsymbol{\psi}) \;=\; \left(\sqrt{\frac{2}{\pi}}\,\frac{\alpha^2}{\alpha-1}\right)^{n} \exp\left\{-\sum_{i=1}^{n}\bigl[\log(\mu_i) - \alpha\log(z_i)\bigr] - \frac{\alpha^2}{\pi(\alpha-1)^2}\sum_{i=1}^{n}\frac{y_i^2 Z_i^2}{\mu_i^2}\right\}.$$

Consequently, up to a constant, the complete log-likelihood function for $\boldsymbol{\psi}$ is

$$\ell_{\mathrm{c}}(\boldsymbol{\psi}) \;=\; -\frac{\alpha^2}{\pi(\alpha-1)^2}\sum_{i=1}^{n}\frac{y_i^2 z_i^2}{\mu_i^2} - \sum_{i=1}^{n}\bigl[\log(\mu_i) - \alpha\log(z_i)\bigr] + n\bigl[2\log(\alpha) - \log(\alpha-1)\bigr].$$

Let $\widehat{z_i^2} = \mathbb{E}\bigl(Z_i^2\,|\,\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}\bigr)$, $\widehat{\log(z_i)} = \mathbb{E}\bigl(\log(Z_i)\,|\,\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}\bigr)$ and $Q\bigl(\boldsymbol{\psi}\,|\,\widehat{\boldsymbol{\psi}}\bigr) = \mathbb{E}\bigl(\ell_{\mathrm{c}}(\boldsymbol{\psi})\,|\,\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}\bigr)$. With these definitions,

$$Q\bigl(\boldsymbol{\psi}\,|\,\widehat{\boldsymbol{\psi}}\bigr) \;=\; -\frac{\alpha^2}{\pi(\alpha-1)^2}\sum_{i=1}^{n}\frac{y_i^2\,\widehat{z_i^2}}{\mu_i^2} - \sum_{i=1}^{n}\bigl[\log(\mu_i) - \alpha\,\widehat{\log(z_i)}\bigr] + n\bigl[2\log(\alpha) - \log(\alpha-1)\bigr].$$

In addition,

$$f\big(z_i\,|\,Y_i=y_i\big) \;\propto\; (z_i^2)^{\left(\frac{\alpha}{2}+1\right)-1}\exp\left[-\frac{\alpha^2\,y_i^2\,z_i^2}{\pi\,\mu_i^2\,(\alpha-1)^2}\right] I_{(0,1)}(z_i)\,,$$

where $I_A(a)=1$ if $a\in A$ and 0 otherwise. Define $W_i=Z_i^2$, $i=1,...,n$. It is straightforward to show that

$$f\big(w_i\,|\,Y_i=y_i\big) \;\propto\; w_i^{\frac{\alpha+1}{2}-1}\exp\left[-\frac{\pi\,\mu_i^2\,(\alpha-1)\,y_i^2\,w_i}{\alpha}\right] I_{(0,1)}(w_i)\,,$$

so that

$$W_i\,|\,Y_i \;=\; y_i \;\sim\; \text{Gamma}\left[\frac{\alpha+1}{2},\,\frac{\pi\,\mu_i^2\,(\alpha-1)\,y_i^2}{\alpha}\right] I_{(0,1)}\,,$$

i.e., the truncated gamma distribution on the $(0,1)$ interval. Thus,

$$\widehat{z}_i^2 \;=\; \frac{\pi\,\mu_i\,(\alpha+1)\,(\alpha-1)^2\,G\left[\dfrac{\alpha^2\,y_i^2}{\pi\,\mu_i^2\,(\alpha-1)^2},\,\dfrac{\alpha+3}{2}\right]}{y_i^2\,G\left[\dfrac{\alpha^2\,y_i^2}{\pi\,\mu_i^2\,(\alpha-1)^2},\,\dfrac{\alpha+1}{2}\right]}\,.$$

However, a closed form expression for $\widehat{\log(z_i)}$ is not available, but it can be computed numerically noticing that $\mathbb{E}\big[\log(Z_i)\big]=\mathbb{E}\big[\log(W_i)\big]/2=C_{i1}(\boldsymbol{\psi})/\big[2\,C_{i0}(\boldsymbol{\psi})\big]$, where

$$C_{ij}(\boldsymbol{\psi}) \;=\; \int_0^1 \big[\log(w)\big]^j\,w^{\frac{\alpha+1}{2}-1}\exp\left[-\frac{\pi\,\mu_i^2\,(\alpha-1)\,y_i^2\,w}{\alpha}\right]\mathrm{d}w\,, \tag{3.3}$$

for $\alpha>1$ and $j=0,1$. Note that if $W_i^*\sim\text{Gamma}(a_i,b_i)$, $a_i,b_i>0$, then $\mathbb{E}\big[\log(W_i^*)\big]=\eta(a_i)-\log(b_i)$, with $\eta(\cdot)$ denoting the digamma function. For this reason, the convergence of $C_{i1}(\boldsymbol{\psi})$ is guaranteed because $C_{i1}(\boldsymbol{\psi})<\mathbb{E}\big[\log(W_i^*)\big]<\infty$, taking a_i and b_i conveniently. Therefore, the k-th iteration of the ECM algorithm takes the form:

- **E step**. For $i=1,...,n$, use $\widehat{\boldsymbol{\psi}}^{(k-1)}$, the estimate of $\boldsymbol{\psi}$ at the $(k-1)$-th iteration of the algorithm, to compute

  $$\widehat{z}_i^{2(k)} \;=\; \frac{\pi\,\widehat{\mu}_i^{(k-1)}\big(\widehat{\alpha}^{(k-1)}+1\big)\big(\widehat{\alpha}^{(k-1)}-1\big)^2\,G\left[\dfrac{\widehat{\alpha}^{2(k-1)}\,y_i^2}{\pi\,\widehat{\mu}_i^{2(k-1)}\big(\widehat{\alpha}^{(k-1)}-1\big)^2},\,\dfrac{\widehat{\alpha}^{(k-1)}+3}{2}\right]}{y_i^2\,G\left[\dfrac{\widehat{\alpha}^{2(k-1)}\,y_i^2}{\pi\,\widehat{\mu}_i^{2(k-1)}\big(\widehat{\alpha}^{(k-1)}-1\big)^2},\,\dfrac{\widehat{\alpha}^{(k-1)}+1}{2}\right]}$$

  and $\widehat{\log(z_i)}^{(k)}=C_{i1}\big(\widehat{\boldsymbol{\psi}}^{(k)}\big)/\big[2\,C_{i0}\big(\widehat{\boldsymbol{\psi}}^{(k)}\big)\big]$, where $\widehat{\mu}_i^{(k-1)}=\exp\big(\boldsymbol{x}_i^\top\,\widehat{\boldsymbol{\beta}}^{(k-1)}\big)$ and $C_{ij}(\boldsymbol{\psi})$, for $j=0,1$, is given in (3.3).

- **CM step I**. Given $\widehat{\alpha}^{(k-1)}$ and $\widehat{\boldsymbol{z}}^{2(k)}=\big(\widehat{z}_1^{2(k)},...,\widehat{z}_n^{2(k)}\big)^\top$, maximize the expression

  $$-\frac{\widehat{\alpha}^{2(k-1)}}{\pi\big(\widehat{\alpha}^{(k-1)}-1\big)^2}\sum_{i=1}^n\frac{y_i^2\,\widehat{z}_i^{2(k)}}{\exp\big(2\,\boldsymbol{x}_i^\top\boldsymbol{\beta}\big)} \;-\;\sum_{i=1}^n\boldsymbol{x}_i^\top\boldsymbol{\beta}$$

  with respect to $\boldsymbol{\beta}$ to obtain $\widehat{\boldsymbol{\beta}}^{(k)}$.

- **CM step II**. Given $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\log(\boldsymbol{z})}^{(k)}=\Big(\widehat{\log(z_1)}^{(k)},...,\widehat{\log(z_n)}^{(k)}\Big)^\top$, maximize the expression

  $$-\frac{\alpha^2}{\pi\,(\alpha-1)^2}\sum_{i=1}^n\frac{y_i^2\,\widehat{z}_i^{2(k)}}{\widehat{\mu}_i^{2(k)}} \;+\;\alpha\sum_{i=1}^n\widehat{\log(z_i)}^{(k)}+\;n\big[2\,\log(\alpha)-\log(\alpha-1)\big]$$

  with respect to α, subject to $\alpha>1$, to obtain $\widehat{\alpha}^{(k)}$.

The maximization procedures in the CM steps can be performed using extant software, e.g., with the `optim` function in the R language [24]. The E and CM steps are repeatedly cycled until a suitable convergence rule is satisfied, e.g., the difference in successive values of the estimates given by the Euclidean norm $\left\|\boldsymbol{\psi}^{(k+1)}-\boldsymbol{\psi}^{(k)}\right\|$ is less than a tolerance value.

In practice, the implementation of the ECM algorithm in this form can be computationally expensive, mainly due to the computation of $\widehat{\log(z_i)}$, $i=1,\ldots,n$, in the E step. To avoid this problem and following the same idea used in [19], we can replace the CM step II by the following step:

- **CME step II**. Given $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}$, update the estimate of α by maximizing the expression $\sum_{i=1}^{n}\log\left[f_{\text{RSHN}}\left(y_i;\widehat{\mu_i}^{(k)},\alpha\right)\right]$ with respect to α, subject to $\alpha>1$, where f_{RSHN} is presented in (2.2). In other words, α is updated based on the maximization of the observed log-likelihood function with $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}$. This step involves a unidimensional maximization, which can be performed using, for instance, the Brent method available in the `optim` function in R.

Finally, the covariance matrix of $\widehat{\boldsymbol{\psi}}$ can be estimated based on the Hessian matrix of the observed log-likelihood function. The `numDeriv` R package [12] provides an accurate numerical approximation for this matrix. In Sections 5 and 6, this estimate of the covariance matrix of $\widehat{\boldsymbol{\psi}}$ is used to build approximate confidence intervals and to compute standard errors. Computational codes are available in *supplementary material*.

Remark 3.1. For the case without covariates, the CM step I is reduced to

CM step I. Update μ as follows: $$\widehat{\mu}^{(k)}=\frac{\widehat{\alpha}^{(k)}}{\widehat{\alpha}^{(k)}-1}\left(\frac{2}{n\pi}\sum_{i=1}^{n}z_i^2\,\widehat{y_i}^{2(k)}\right)^{1/2}.$$

Remark 3.2. In the RSHN regression model, when the intercept term is included in the model, an initial value to $\boldsymbol{\psi}$ can be obtained based on the moment estimators presented in (3.1). Such initial value can be considered as $\widehat{\boldsymbol{\psi}}^{(0)}=\left(\widehat{\beta}_{0\mathrm{M}},0,\ldots,0,\widehat{\alpha}_{\mathrm{M}}\right)$.

4. RESIDUAL DIAGNOSTICS FOR THE RSHN MODEL

In this section, we discuss some aspects related to the deviance and quantile residuals for the RSHN model.

4.1. Deviance residuals

Residual diagnostics for the RSHN model can be carried out using the deviance residuals defined as $r_{D_i}=\operatorname{sign}(Y_i-\widehat{\mu}_i)\sqrt{2}\left[\ell(\widetilde{\mu}_i,\widehat{\alpha})-\ell(\widehat{\mu}_i,\widehat{\alpha})\right]^{1/2}$, where $\ell(\cdot)$ denotes the log-likelihood function, $\widetilde{\mu}_i$ is the ML estimator of $\mu_i=\exp\left(\boldsymbol{x}_i^{\top}\boldsymbol{\beta}\right)$ under the saturated model and $\widehat{\mu}_i$ is the ML estimator of μ_i under the working model (with $p<n$ regression coefficients).

For the RSHN regression model, with $\widetilde{\mu}_i = Y_i$ and $\ell(\cdot)$ coming from (3.2), these residuals are given by

$$\begin{aligned} r_{D_i} \;=\; & \operatorname{sign}(Y_i - \widehat{\mu}_i)\,\sqrt{2}\,\Bigg(\widehat{\alpha}\,\log(Y_i/\widehat{\mu}_i) + \log\Bigg\{G\Bigg[\frac{\widehat{\alpha}^2}{\pi(\widehat{\alpha}-1)}, \frac{\widehat{\alpha}+1}{2}\Bigg]\Bigg\} \\ & - \log\Bigg\{G\Bigg[\frac{Y_i^2\,\widehat{\alpha}^2}{\pi\widehat{\mu}_i^2\,(\widehat{\alpha}-1)}, \frac{\widehat{\alpha}+1}{2}\Bigg]\Bigg\}\Bigg)^{1/2}, \qquad \text{for} \quad i = 1, ..., n\,, \end{aligned}$$

where $G(\cdot)$ is given in (2.2). If the model is correct, the approximate distribution of r_{D_i}, $i = 1, ..., n$, is the standard normal. The normality of the residuals can be tested based on different tests such as the Shapiro–Wilk (SW), Anderson–Darling (AD) and Cramér–von Mises (CVM) tests [33]. Moreover, simulated envelopes (Atkinson [3]) are also useful to assess the fitting of the models.

4.2. Quantile residuals

A second alternative for residual analysis can be based on the normalized quantile residuals (Dunn and Smyth [11]). These residuals are defined as

$$r_{Q_i} = \Phi^{-1}\big[F(Y_i; \widehat{\boldsymbol{\psi}})\big]\,, \qquad i = 1, ..., n\,,$$

where $F(\cdot\,; \boldsymbol{\psi})$ is the cdf of the response variable and $\Phi^{-1}(\cdot)$ denotes the quantile function of the standard normal distribution. Except for the uncertainty due to estimation of the parameters, if the model is correct, r_{Q_i}, $i = 1, ..., n$, constitute a random sample from the standard normal distribution. For the RSHN model, we have

$$r_{Q_i} \;=\; \widehat{\alpha}\,\sqrt{\frac{2^{\widehat{\alpha}}}{\pi}}\,\Bigg[\sqrt{\frac{\pi}{2}}\,\frac{\widehat{\mu}_i\,(\widehat{\alpha}-1)}{\widehat{\alpha}}\Bigg]^{\widehat{\alpha}}\,\Gamma\bigg(\frac{\widehat{\alpha}+1}{2}\bigg)\int_0^{Y_i} u_i^{-(\widehat{\alpha}+1)}\,G\Bigg[\frac{\widehat{\alpha}^2\,u_i^2}{\pi\widehat{\mu}_i^2\,(\widehat{\alpha}-1)^2}, \frac{\widehat{\alpha}+1}{2}\Bigg]\,\mathrm{d}u_i\,,$$

where the integral can be computed numerically using, for instance, the `integrate` function in R.

5. SIMULATION STUDIES

In this section, we present two simulation studies. The first is devoted to assess the performance of the ML estimator for the RSHN model in finite samples when the model is well specified. The main goal of the second study is similar to the one in Leiva *et al.* [18], with the aim of assess the behavior of the deviance and normalized quantile residuals when the model is either well or misspecified.

5.1. Parameters recovery

We stress that in Olmos *et al.* [21], the authors did not carry out a simulation study, so that it is of interest to address this issue. To draw synthetic datasets from the RSHN model, we fix $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ (two covariates) and α at the true values in Table 3.

Table 3: Bias, average of the asymptotic standard error (SE), square root of the simulated mean squared error (RMSE) and coverage probability of the 95% asymptotic confidence intervals (CP) of the estimators under the RSHN regression model with 1,000 replications.

Parameter	True value	$n = 50$				$n = 100$				$n = 200$			
		Bias	SE	RMSE	CP	Bias	SE	RMSE	CP	Bias	SE	RMSE	CP
α	2.5	1.962	1.987	1.659	0.906	1.655	1.311	1.178	0.913	1.001	0.926	0.879	0.931
β_0	0.5	−0.039	0.352	0.304	0.924	−0.028	0.287	0.231	0.928	−0.011	0.171	0.159	0.955
β_1	0.5	−0.001	0.272	0.245	0.933	−0.001	0.192	0.151	0.936	−0.001	0.125	0.111	0.947
β_2	0.05	−0.001	0.009	0.005	0.934	0.000	0.007	0.003	0.939	0.000	0.003	0.002	0.947
α	2.5	2.139	2.152	1.993	0.912	1.683	1.559	1.313	0.921	1.149	1.082	1.032	0.932
β_0	1.0	0.041	0.281	0.265	0.935	0.031	0.256	0.225	0.937	0.010	0.181	0.169	0.945
β_1	0.5	−0.031	0.223	0.201	0.936	−0.029	0.169	0.147	0.941	−0.024	0.127	0.119	0.941
β_2	0.05	−0.005	0.010	0.006	0.931	−0.004	0.007	0.004	0.937	−0.004	0.005	0.004	0.942
α	2.5	2.389	1.559	1.379	0.918	1.446	1.333	1.052	0.922	0.982	0.790	0.754	0.935
β_0	0.5	−0.089	0.369	0.311	0.912	−0.045	0.246	0.201	0.934	−0.021	0.171	0.152	0.952
β_1	0.5	0.031	0.249	0.219	0.924	0.005	0.178	0.152	0.931	0.003	0.130	0.111	0.947
β_2	0.025	0.001	0.010	0.005	0.926	0.000	0.008	0.003	0.931	0.000	0.003	0.002	0.939
α	2.5	2.424	1.587	1.401	0.914	1.452	1.156	1.038	0.921	0.951	0.891	0.858	0.941
β_0	1.0	−0.079	0.402	0.351	0.924	−0.059	0.271	0.217	0.943	−0.013	0.178	0.156	0.953
β_1	0.5	0.012	0.251	0.210	0.931	0.009	0.180	0.154	0.933	0.002	0.135	0.112	0.941
β_2	0.025	0.000	0.009	0.005	0.924	0.000	0.007	0.003	0.931	0.000	0.003	0.002	0.941
α	3.0	2.094	1.852	1.650	0.918	1.912	1.210	1.003	0.919	0.929	0.974	0.936	0.937
β_0	0.5	0.049	0.336	0.281	0.944	0.043	0.251	0.202	0.945	0.038	0.161	0.143	0.949
β_1	0.5	−0.005	0.242	0.200	0.942	−0.002	0.186	0.142	0.943	0.000	0.125	0.101	0.947
β_2	0.05	−0.001	0.009	0.004	0.922	−0.001	0.007	0.003	0.939	0.000	0.003	0.002	0.942
α	3.0	2.150	2.014	1.833	0.908	1.850	1.319	1.142	0.923	0.839	0.981	0.954	0.931
β_0	1.0	0.090	0.369	0.316	0.914	0.050	0.299	0.245	0.929	0.040	0.214	0.190	0.943
β_1	0.5	−0.049	0.241	0.206	0.932	−0.046	0.187	0.143	0.933	−0.038	0.111	0.097	0.944
β_2	0.05	−0.006	0.009	0.006	0.917	−0.005	0.008	0.005	0.933	−0.004	0.005	0.004	0.941
α	3.0	2.202	1.263	1.029	0.902	1.456	1.099	0.878	0.914	1.141	0.725	0.697	0.935
β_0	0.5	0.057	0.349	0.277	0.930	0.036	0.243	0.192	0.944	0.028	0.151	0.136	0.949
β_1	0.5	0.037	0.271	0.203	0.929	0.019	0.160	0.135	0.935	0.013	0.123	0.101	0.943
β_2	0.025	0.000	0.009	0.004	0.957	0.000	0.007	0.003	0.952	0.000	0.003	0.002	0.950
α	3.0	2.378	1.295	1.075	0.912	1.670	1.091	0.914	0.925	0.947	0.619	0.600	0.941
β_0	1.0	0.035	0.356	0.287	0.948	0.031	0.251	0.193	0.949	0.022	0.178	0.152	0.950
β_1	0.5	0.011	0.261	0.198	0.930	0.005	0.184	0.139	0.932	0.001	0.119	0.097	0.943
β_2	0.025	0.000	0.008	0.004	0.939	0.000	0.007	0.003	0.946	0.000	0.003	0.002	0.949
α	5.0	2.419	2.514	2.297	0.902	1.926	1.894	1.640	0.930	1.503	1.212	1.199	0.937
β_0	0.5	0.060	0.351	0.274	0.962	0.043	0.231	0.187	0.958	0.030	0.134	0.117	0.952
β_1	0.5	−0.007	0.246	0.177	0.959	−0.002	0.157	0.116	0.957	−0.001	0.099	0.086	0.956
β_2	0.05	−0.001	0.008	0.004	0.961	−0.001	0.007	0.003	0.960	0.000	0.003	0.002	0.957
α	5.0	2.134	2.152	1.958	0.904	1.069	1.419	1.275	0.912	0.825	0.974	0.951	0.939
β_0	1.0	0.082	0.362	0.270	0.910	0.078	0.266	0.213	0.934	0.044	0.200	0.181	0.947
β_1	0.5	−0.019	0.253	0.183	0.957	−0.015	0.184	0.136	0.954	−0.005	0.119	0.092	0.952
β_2	0.050	−0.005	0.009	0.005	0.902	−0.005	0.008	0.005	0.922	−0.004	0.005	0.004	0.939
α	5.0	1.354	2.055	1.728	0.902	1.029	1.462	1.284	0.919	0.899	1.034	0.995	0.932
β_0	0.5	0.025	0.314	0.246	0.959	0.018	0.233	0.179	0.954	0.013	0.145	0.126	0.953
β_1	0.5	0.014	0.256	0.186	0.958	0.008	0.176	0.128	0.957	0.008	0.099	0.084	0.944
β_2	0.025	0.000	0.008	0.004	0.958	0.000	0.007	0.003	0.956	0.000	0.003	0.002	0.952
α	5.0	1.768	2.263	1.928	0.922	1.483	1.500	1.396	0.930	1.156	1.127	1.091	0.938
β_0	1.0	−0.007	0.325	0.257	0.962	−0.005	0.221	0.177	0.955	−0.003	0.152	0.131	0.952
β_1	0.5	0.006	0.254	0.186	0.960	0.002	0.187	0.136	0.956	0.000	0.100	0.084	0.954
β_2	0.025	0.000	0.009	0.004	0.939	0.000	0.007	0.002	0.940	0.000	0.003	0.002	0.942

In practice, covariates may have any kind of association. Therefore, we assume that the values of one covariate depends on the other. In short, for $i=1,...,n$, the steps to generate datasets are the following:

- Draw $x_{1i} \sim \mathrm{U}(10,90)$ (the uniform distribution).
- Draw $x_{2i} \sim \mathrm{Bernoulli}(\theta_i)$, where $\theta_i = \exp(2-0.025\,x_{1i})\big/\big[1+\exp(2-0.025\,x_{1i})\big]$, i.e., $x_{2i}=1$ with probability θ_i that varies between 0.438 and 0.852 depending on the value of x_{1i}.
- Compute $\mu_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ and draw $W_i \sim \mathrm{HN}(\sigma_i)$ independent from $Z_i \sim \mathrm{Beta}(\alpha,1)$, where $\sigma_i = \sqrt{2}\,\mu_i\,\alpha\big/\big[\sqrt{\pi}\,(\alpha-1)\big]$.
- Compute $Y_i = W_i/Z_i$.

Once generated, the values of $\boldsymbol{x_i}$, $i=1,...,n$, are kept fixed throughout the simulations. For each generated sample, we apply the scheme described in Section 3.2 to estimate $\boldsymbol{\beta}$ and α, while the standard errors of the estimates are computed from the Hessian matrix in Section 3.2. We report the average bias of the estimates (Bias), the average of the asymptotic standard error (SE), the square root of the simulated mean squared error (RMSE) and the coverage probability of the 95% asymptotic confidence intervals (CP).

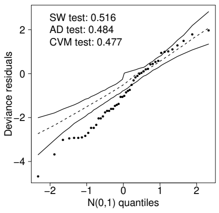
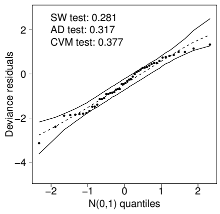
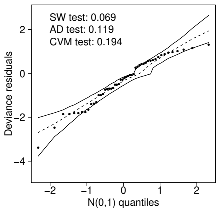
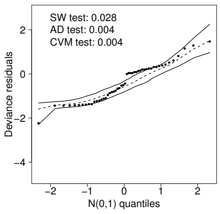
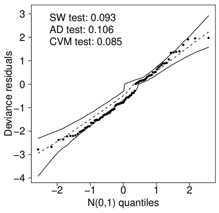
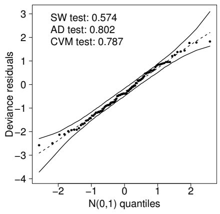
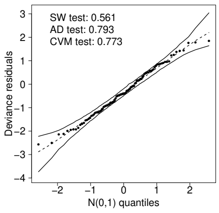
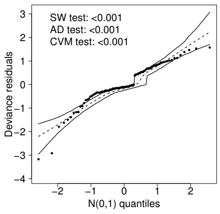
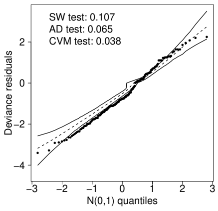
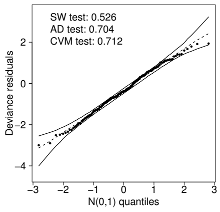
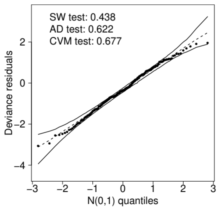
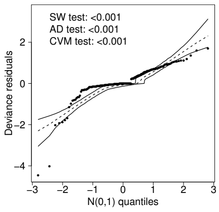
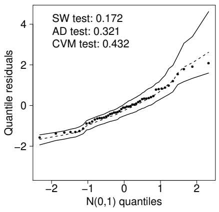
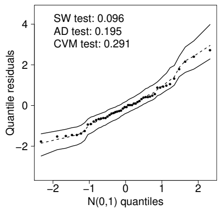
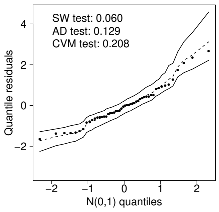
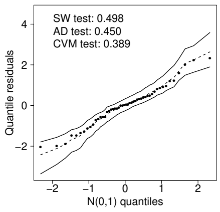
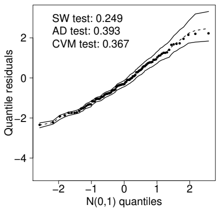
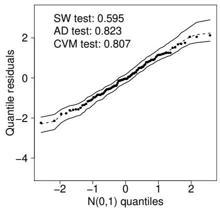
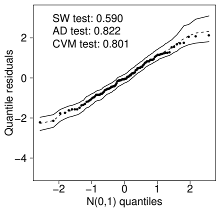
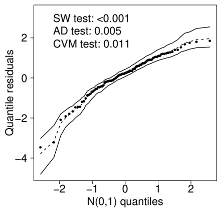
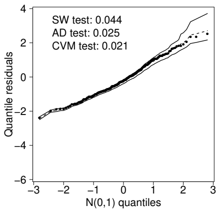
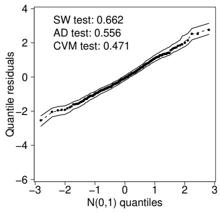
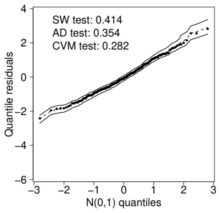
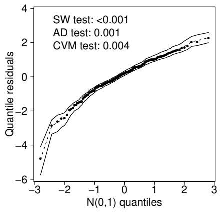
We considered four different regression coefficients $\boldsymbol{\beta}$, namely, $(0.5,0.5,0.05)$, $(1.0,0.5,0.05)$, $(0.5,0.5,0.025)$ and $(1.0,0.5,0.025)$. Such values guarantee that the drawn values of y_i belong to the interval $(1.649,4.711)$ in all the cases. We also considered $\alpha \in \{2.5,3.0,5.0\}$ (that guarantees a finite value for the variance of y_i) and $n \in \{50,100,200\}$. The results presented in Table 3 were obtained from 1000 replications. Note that in all cases, the absolute value of bias and the RMSE decrease when n increases, suggesting that the estimators are consistent, and the coverage probabilities are close to the nominal value, as expected. Except for the estimator of α, we see that SE and RMSE get closer when the sample size increases, as expected from the asymptotic properties of the estimators. However, even for $n=200$ the bias of $\widehat{\alpha}$ is substantial. This result is in agreement with other slashed distributions in the literature (see, for instance, Astorga [2] and Reyes *et al.* [27, 26, 25]). This should not be a serious concern because in practice the most important inferences pertain to the mean of the response variable, which depends only on the regression coefficients vector $\boldsymbol{\beta}$. Additionally, since the coverage probability of the confidence interval for α ranges from 0.902 to 0.941, we see that the interval estimator behaves better than the point estimator.

5.2. Deviance and quantile residuals

In order to assess the performance of the distribution of the deviance and quantile residuals, we take samples drawn from the $\mathrm{RG}(\mu_i,\phi=1)$ model (which also corresponds to the $\mathrm{RW}(\mu_i,\delta=1)$ model) and $\mathrm{RSHN}(\mu_i,\alpha=2.1)$ models, where $\mu_i=\exp(\beta_0+\beta_1 x_i)$, and x_i was drawn from the $\mathrm{U}(0,10)$ distribution. For each sample, we fit the RSHN, RG, RW and RBS regression models and present the quantile-quantile (QQ) plots with simulated envelopes based on 1000 replicates for the deviance and quantile residuals. We consider three sample sizes: $n=50$, $n=100$ and $n=200$. We also present the p-value for the SW, AD and CVM normality tests. Tables 4 and 5 show the QQ plots.

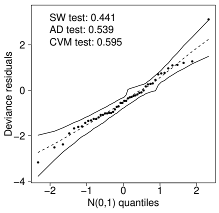
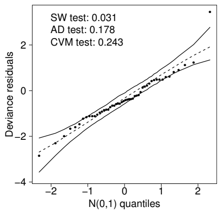
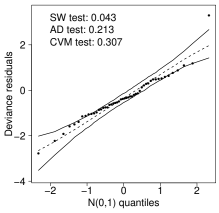
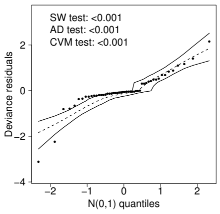
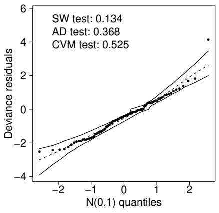
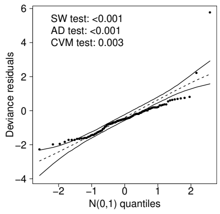
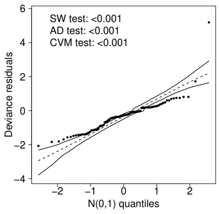
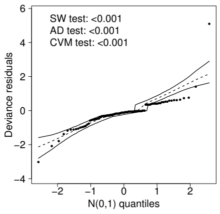
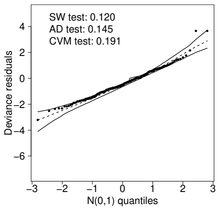
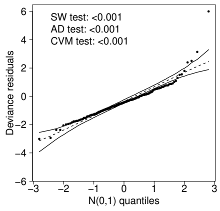
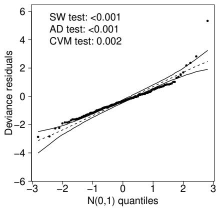
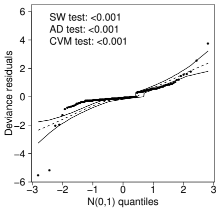
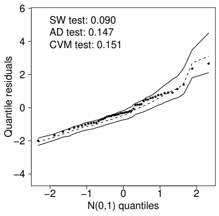
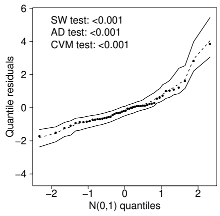
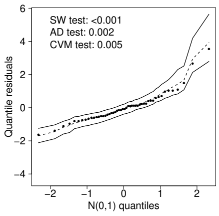
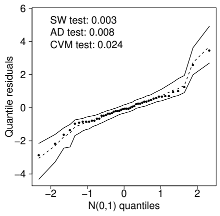
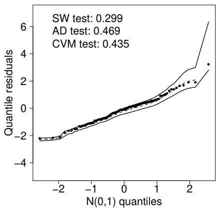
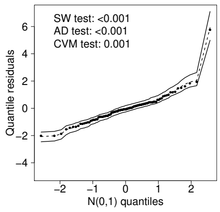
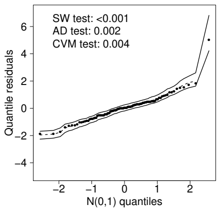
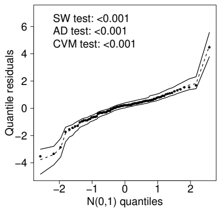
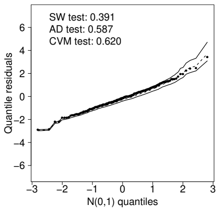
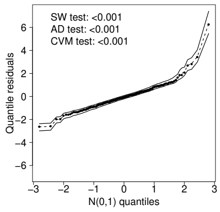
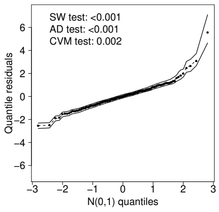
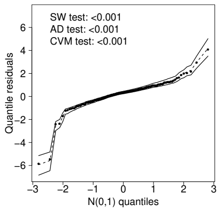
As expected, when the true model is the RG model, the QQ plots related to the RG and RW models present an approximately linear behavior and a good agreement with the standard normal distribution for the three sample sizes for both, deviance and quantile residuals.

Table 4: QQ plots with simulated envelopes for the deviance and quantile residuals when RG($\mu_i, \phi = 1$) is the true model.

Residual	n	Fitted model			
		RSHN	RG	RW	RBS
Deviance	50				
	100				
	200				
Quantile	50				
	100				
	200				

Moreover, the three normality tests do not reject the hypothesis of normality under the common significance levels. In counterpart, in this case the RSHN regression models yields unsatisfactory results and the normality assumption of the residuals is questionable.

Table 5: QQ plots with simulated envelopes for the deviance and quantile residuals when RSHN$(\mu_i, \alpha = 2.1)$ is the true model.

Residual	n	Fitted model			
		RSHN	RG	RW	RBS
Deviance	50				
	100				
	200				
Quantile	50				
	100				
	200				

When the true model is the RSHN model, as expected, the QQ plots for the deviance and quantile residuals of the RSHN model present a good agreement with the standard normal for all sample sizes. In addition, the deviance residuals for the RG and RW models only provides fair results when $n = 50$. This result suggest that the RG and RW regression models are very competitive in small sample sizes, even when the true model is not the RG model or the RW model. Finally, the deviance and quantile residuals of the RBS regression model are far away from the identity line in all the cases, suggesting poor results when the true model is the RG model or the RSHN model.

6. DATA ANALYSIS

In this section, the regression models formulated in Section 2 are applied in the analysis of two datasets.

6.1. Hunter-gatherer group dataset

In this section, the regression models formulated in Section 2 are applied in the analysis of a dataset described in Kelly [16]. The dataset is related to the macroecological relationship between the size of the homerange (measured in km^2) of a hunter-gatherer group (response variable) and the contribution (in percentage) of hunted foods to the diet. The dataset comprises 39 groups. The sample mean, median and standard deviation of the size of the homerange are 4004.4, 906.0 and 10728.1 km^2, respectively, while the sample skewness and kurtosis coefficients are $\sqrt{\widehat{\nu}_3} = 4.46$ and $\widehat{\nu}_4 = 23.43$.

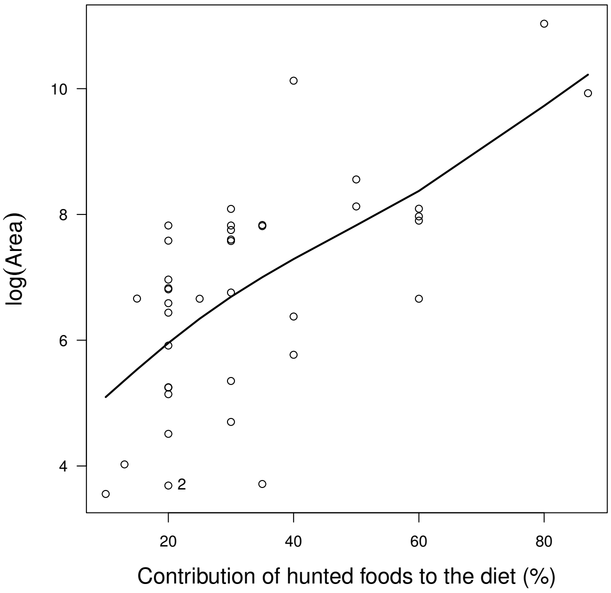


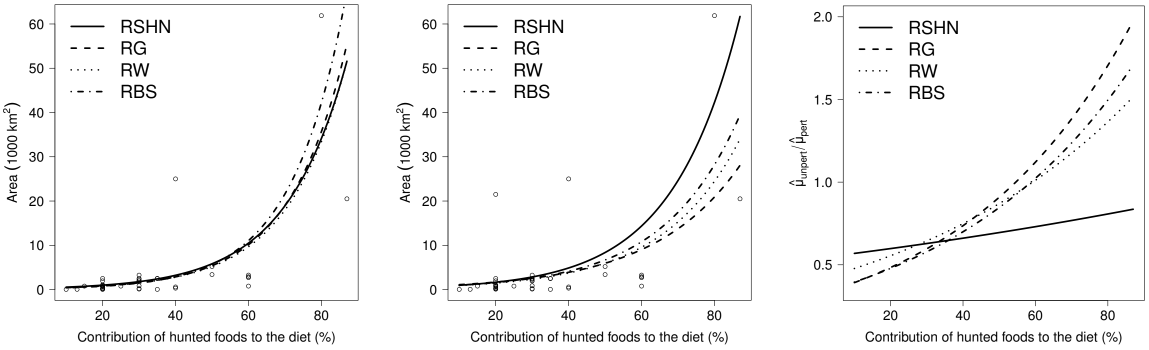
Figure 2: Scatterplot and smoothing spline of the homerange, in 1000 km^2, and the contribution of hunted foods to the diet (observation 2 was perturbed in the analysis).

Figure 2 shows the scatterplot of the data and a smoothing spline, which indicates that the logarithmic link function is adequate. We fit the RG, RW, RBS and RSHN models, with results presented in Table 6. The deviance and quantile residuals plots with envelopes are presented in the upper panels in Figures 4 and 5. The lines in these plots represent the 2.5%, 50% and 97.5% quantile values of the residuals computed from 100 bootstrap samples generated from the models in Table 6. Note that, based on both residuals, all models seem appropriate for this dataset. Furthermore, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) values are similar for all models.

Table 6: Parameter estimates (standard errors) and information criteria for the RG, RW, RBS and RSHN regression models fitted to the hunter-gatherer group dataset.

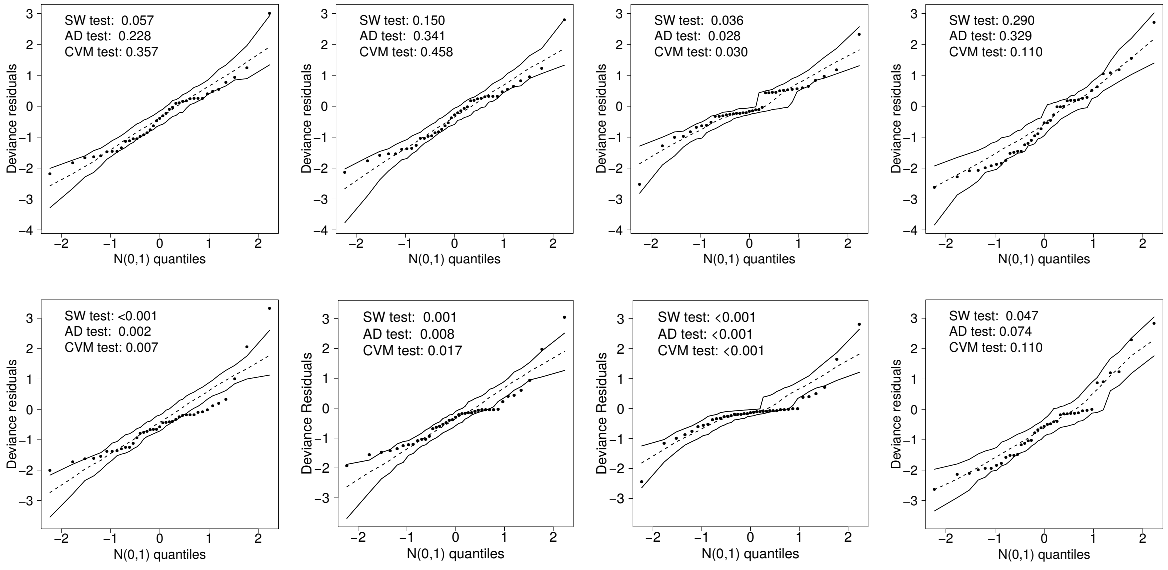
Dataset	Parameter	Model			
		RG	RW	RBS	RSHN
Unperturbed	β_0	5.442 (0.504)	5.456 (0.436)	5.290 (0.478)	5.718 (0.136)
	β_1	0.063 (0.013)	0.062 (0.012)	0.067 (0.013)	0.059 (0.010)
	α	—	—	—	2.225 (1.541)
	ϕ	0.811 (0.159)	—	—	—
	δ	—	0.845 (0.100)	—	—
	ξ	—	—	0.805 (0.227)	—
	AIC	670.26	669.25	668.02	670.11
	BIC	675.25	674.24	673.01	675.10
Perturbed	β_0	6.588 (0.759)	6.345 (0.482)	6.407 (0.491)	6.332 (0.340)
	β_1	0.042 (0.020)	0.047 (0.013)	0.048 (0.014)	0.054 (0.013)
	α	—	—	—	1.517 (1.301)
	ϕ	0.602 (0.115)	—	—	—
	δ	—	0.695 (0.078)	—	—
	ξ	—	—	0.587 (0.227)	—
	AIC	698.80	693.57	691.66	688.20
	BIC	703.79	698.56	696.64	693.19

In order to illustrate the robustness of the RSHN model, we perturb the response variable of observation 2 in Figure 2 by adding two standard deviations (originally with an area of 4,000 km^2). The lower panels in Figures 4 and 5 show the deviance and the quantile residuals plots for the models fitted to the perturbed data. Note that for both residuals, the SW, AD and CVM tests support that the residuals of the RSHN model come from the standard normal distribution for datasets without and with perturbation. This fact suggests that the RG, RW and RBS models do not yield a good fit for the perturbed dataset, differently from the RSHN model, which yields a good fit in both scenarios. Information criteria for the perturbed dataset in Table 6 also suggest that the best fit is achieved with the RSHN model. Due to the perturbation, estimates of the coefficient of the contribution of hunted foods to the diet (β_1) decrease 33.3%, 24.2% and 28.4% under the RG, RW and RBS models, respectively, whereas for the RSHN model the reduction amounts to 8.5%. Estimated means of the homerange for unperturbed and perturbed data are displayed in Figure 3. We stress that the ratio of the estimated area for unperturbed data to perturbed data is much more stable for the RSHN model, especially for large values of the contribution of hunted foods to the diet, as can be seen in Figure 3(c).



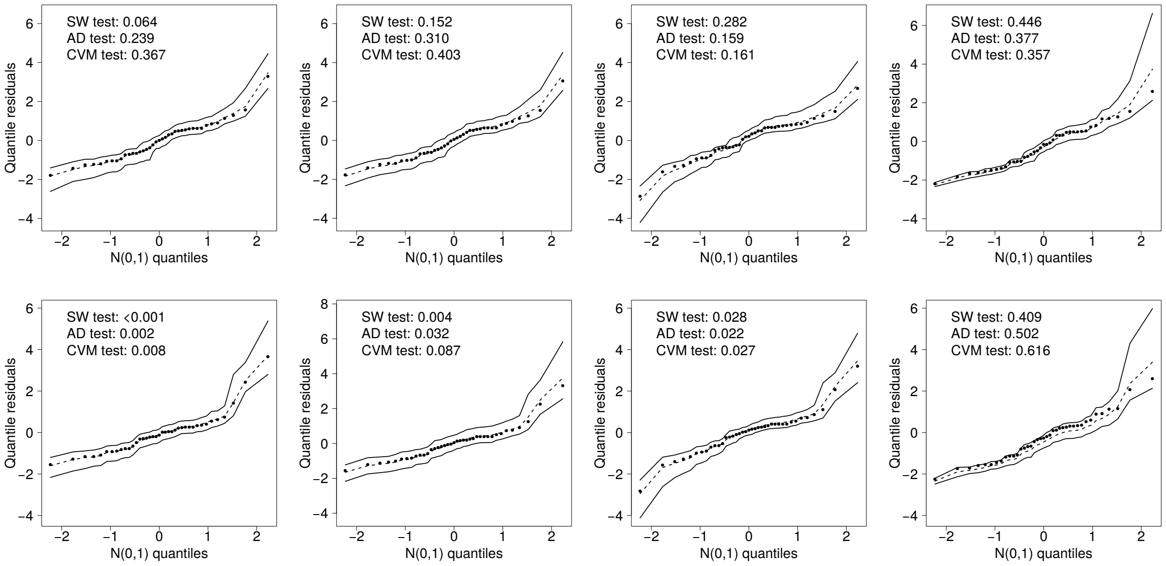
(a) **(b)** **(c)**

Figure 3: Scatterplot of the homerange and the contribution of hunted foods to the diet together with estimated means under different models for data (a) without perturbation and (b) with perturbation, and (c) ratio of the estimated area for unperturbed data ($\widehat{\mu}_{\text{unpert}}$) to perturbed data ($\widehat{\mu}_{\text{pert}}$).



(a) **(b)** **(c)** **(d)**

Figure 4: Deviance residual plots with simulated envelopes for the (a) RG, (b) RW, (c) RBS and (d) RSHN regression models fitted to the hunter-gatherer group dataset without perturbation (upper panel) and with perturbation (lower panel).



(a) **(b)** **(c)** **(d)**

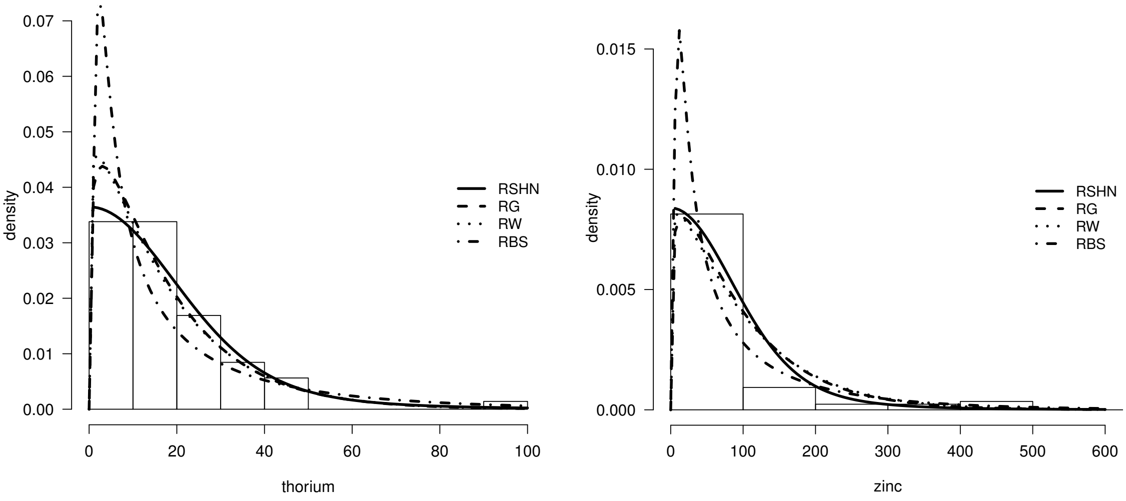
Figure 5: Quantile residual plots with simulated envelopes for the (a) RG, (b) RW, (c) RBS and (d) RSHN regression models fitted to the hunter-gatherer group dataset without perturbation (upper panel) and with perturbation (lower panel).

6.2. Minerals concentration dataset

This dataset is related to the concentration of some minerals in soil samples obtained at the Mining Department, University of Atacama, Chile. This dataset was previously analyzed in Gómez *et al.* [14] and Olmos *et al.* [21] . The measurements are related to nickel (Ni) and zinc (Zn) respectively. In our application, we consider to model jointly the positive measurements related to thorium (Th, $n = 71$), uranium (U, $n = 57$), vanadium (V, $n = 86$) and zinc (Zn, $n = 86$). The unit of measurement of the concentrations (response variable) is parts-per million (ppm). The dataset comprises 300 observations. The sample mean, median and standard deviation of the concentrations are 72.43, 29.00 and 110.06, respectively, while the sample skewness and kurtosis coefficients are $\sqrt{\widehat{\nu}_3} = 4.37$ and $\widehat{\nu}_4 = 35.87$. Note that the kurtosis is unusually greater than the kurtosis of the normal distribution. Given the high value of kurtosis, we consider appropriate to model this dataset with the RSHN model in Section 2, linking the covariates to the mean as $\mu_i = \exp\bigl(\beta_{\mathrm{Th}} x_{i\mathrm{Th}} + \beta_{\mathrm{U}} x_{i\mathrm{U}} + \beta_{\mathrm{V}} x_{i\mathrm{V}} + \beta_{\mathrm{Zn}} x_{i\mathrm{Zn}}\bigr)$, $i = 1, \ldots, 300$, where $x_{i\mathrm{Th}}$, $x_{i\mathrm{U}}$, $x_{i\mathrm{V}}$ and $x_{i\mathrm{Zn}}$ are indicator variables assuming the value 1 when the i-th observation corresponds to the referred mineral. We also compare the results with the RG, RW and RBS regression models. Results are presented in Table 7. Note that AIC and BIC attain the smallest values for the RSHN model. Figure 6 shows the histogram of thorium and zinc concentrations compared with the fitted density functions. Table 8 also presents the p-value for the univariate Kolmogorov–Smirnov (KS) test for comparison of empirical and fitted cdf's from each mineral. Note that all p-values are greater than 5% for the RSHN model, suggesting a better fit for this model over the RG, RW and RBS models.

Table 7: Parameter estimates (standard errors) and information criteria for the RG, RW, RBS and RSHN regression models fitted to the minerals dataset.

Parameter	Model			
	RG	RW	RBS	RSHN
β_{Th}	2.871 (0.119)	2.866 (0.110)	3.005 (0.146)	2.989 (0.127)
β_{U}	2.436 (0.133)	2.434 (0.123)	2.508 (0.155)	2.581 (0.136)
β_{V}	4.896 (0.108)	4.892 (0.100)	4.646 (0.107)	5.071 (0.124)
β_{Zn}	4.572 (0.108)	4.589 (0.101)	4.555 (0.122)	4.458 (0.114)
α	—	—	—	2.871 (2.541)
ϕ	1.206 (0.088)	—	—	—
δ	—	1.080 (0.046)	—	—
ξ	—	—	1.147 (0.082)	—
AIC	2917.55	2920.67	2980.57	2906.97
BIC	2936.07	2939.18	2999.09	2925.49



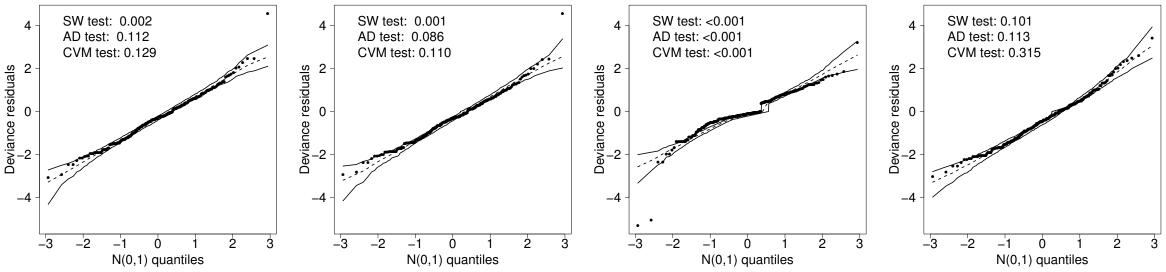
(a) **(b)**

Figure 6: Histogram and fitted density functions for RSHN, RG, RW and RBS models in minerals dataset: (a) thorium and (b) zinc.

Table 8: p-values for the Kolmogorov–Smirnov goodness-of-fit test.

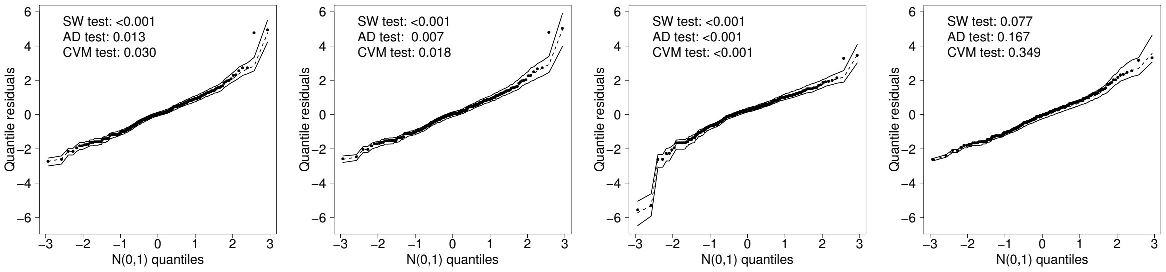
Mineral	RG	RW	RBS	RSHN
Th	0.269	0.195	0.009	0.580
U	0.947	0.955	0.119	0.535
V	0.105	0.112	<0.001	0.348
Zn	0.003	0.002	0.040	0.065

Besides the information criteria in Table 7, Figures 7 and 8 show the deviance and the quantile residuals plots for the fitted models. Note that for both residuals, the SW, AD and CVM tests support (at a 5% significance level) that only the residuals of the RSHN model come from the standard normal distribution. This fact suggests that the RG, RW and RBS models do not yield a good fit for this dataset.



(a) **(b)** **(c)** **(d)**

Figure 7: Deviance residual plots with simulated envelopes for the (a) RG, (b) RW, (c) RBS and (d) RSHN regression models fitted to the minerals dataset.



(a) **(b)** **(c)** **(d)**

Figure 8: Quantile residual plots with simulated envelopes for the (a) RG, (b) RW, (c) RBS and (d) RSHN regression models fitted to the minerals dataset.

7. CONCLUSION

In this work, a reparameterization of the distribution proposed by Olmos *et al.* [21] based on the mean motivated us to propose a regression model for positive data. The proposed model is an alternative to some well-known models for positive response variables. Maximum likelihood estimates are computed with the EM algorithm. A simulation study was carried out to assess some properties of the proposed estimator. The analysis of two datasets illustrates the robustness of the model. Extensions of this work might include Bayesian inference, influence assessment and mixed models.

ACKNOWLEDGMENTS

We would like to thank the Editor, an AE and two anonymous reviewers for their time and valuable remarks. The work of the first author is partially supported by Proyecto DIUDA Programa de Inserción N.º 22367, Chile. The work of the second author is partially supported by grant FONDECYT 11160670, Chile. The work of the third author is partially supported by CNPq, Brazil.

REFERENCES

[1] Asgharzadeh, A.; Nadarajah, S. and Sharafi, F. (2018). Weibull Lindley distribution, *REVSTAT – Statistical Journal*, **16**, 87–113.

[2] Astorga, J.M.; Gómez, H.W. and Bolfarine, H. (2017). Slashed generalized exponential distribution, *Communications in Statistics – Theory and Methods*, **46**, 2091–2102.

[3] Atkinson, A.C. (1985). *Plots, Transformations, and Regression*, Clarendon, Oxford.

[4] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[5] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistics*, **12**, 199–208.

[6] Bourguignon, M.; Leão, J.; Leiva, V. and Santos-Neto, M. (2017). The transmuted Birnbaum–Saunders distribution, *REVSTAT – Statistical Journal*, **15**, 601–628.

[7] Chou, C.Y. and Liu, H.R. (1998). Properties of the half-normal distribution and its application to quality control, *Journal of Industrial Technology*, **14**, 4–7.

[8] Cooray, K. and Ananda, M.M.A. (2008). A generalization of the half-normal distribution with applications to lifetime data, *Communications in Statistics – Theory and Methods*, **10**, 195–224.

[9] Cordeiro, G.M.; Pescim, R.R. and Ortega, E.M.M. (2012). The Kumaraswamy generalized half-normal distribution for skewed positive data, *Journal of Data Science*, **10**, 195–224.

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

[11] Dunn, P.K. and Smyth, G.K. (1996). Randomized quantile residuals, *Journal of Computational and Graphical Statistics*, **5**, 236–244.

[12] Gilbert, P. and Varadhan, R. (2016). numDeriv: Accurate Numerical Derivatives, *R package version 2016.8-1*, `http://CRAN.R-project.org/package=numDeriv`.

[13] Gómez, Y.M. and Bolfarine, H. (2015). Likelihood-based inference for power half-normal distribution, *Journal of Statistical Theory and Applications*, **14**, 383–398.

[14] Gómez, H.W.; Venegas, O. and Bolfarine, H. (2006). Skew-symmetric distributions generated by the distribution function of the normal distribution, *Envirometrics*, **18**, 395–407.

[15] Henze, N. (1986). A probabilistic representation of the skew-normal distribution, *Scandinavian Journal of Statistics*, **13**, 271–275.

[16] Kelly, R.L. (2013). *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*, 2nd ed., Cambridge, New York.

[17] Khan, M.A. and Islam, H.N. (2012). Bayesian analysis of system availability with half-normal life time, *The Mathematical Scientist*, **9**, 203–209.

[18] Leiva, V.; Ferreira, M.; Gomes, M.I. and Lillo, C. (2016). Extreme value Birnbaum–Saunders regression models applied to environmental data, *Stochastic Environmental Research and Risk Assessment*, **30**, 1045–1058.

[19] Lin, T.T.; Lee, J.C. and Hsieh, W.J. (2007). Robust mixture modeling using the skew t distribution, *Statistics and Computing*, **17**, 81–92.

[20] Moral, R.A.; Hinde, J. and Demétrio, C.G.B. (2017). Half-normal plots and overdispersed models in R: the hnp package, *Journal of Statistical Software*, **81**, 1–23.

[21] Olmos, N.M.; Varela, H.; Gómez, H.W. and Bolfarine, H. (2012). An extension of the half-normal distribution, *Statistical Papers*, **53**, 875–886.

[22] Pewsey, A. (2002). Large-sample inference for the general half-normal distribution, *Communications in Statistics – Theory and Methods*, **31**, 1045–1054.

[23] Pewsey, A. (2004). Improved likelihood based inference for the general half-normal distribution, *Communications in Statistics – Theory and Methods*, **33**, 197–204.

[24] R Core Team (2018). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org/`.

[25] Reyes, J.; Iriarte, Y.; Jodrá, P. and Gómez, H.W. (2019). The slash Lindley–Weibull distribution, *Methodology and Computing in Applied Probability*, **21**, 235–251.

[26] Reyes, J.; Venegas, O. and Gómez, H.W. (2017a). Modified slash Lindley distribution, *Journal of Probability and Statistics*, Article ID 6303462, 1–9.

[27] Reyes, J.; Vilca, F.; Gallardo, D.I. and Gómez, H.W. (2017b). Modified slash Birnbaum–Saunders distribution, *Hacettepe Journal of Mathematics and Statistics*, **46**, 969–984.

[28] Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with discussion), *Applied Statistics*, **54**, 507–554.

[29] Rigby, R.A. and Stasinopoulos, D.M. (2019). gamlss.dist: distributions for generalized additive models for location scale and shape, *R package version 5.1-3*, `https://CRAN.R-project.org/package=gamlss.dist`.

[30] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2014). On a reparameterized Birnbaum–Saunders distribution and its moments, estimation and applications, *REVSTAT – Statistical Journal*, **12**, 247–272.

[31] Santos-Neto, M.; Cysneiros, F.J.; Leiva, V. and Barros, M. (2016). Reparameterized Birnbaum–Saunders regression models with varying precision, *Electronic Journal of Statistics*, **10**, 2825–2855.

[32] Wiper, M.P.; Girón, F.J. and Pewsey, A. (2008). Objective Bayesian inference for the half-normal and half-t distributions, *Communications in Statistics – Theory and Methods*, **37**, 3165–3185.

[33] Yazici, B. and Yocalan, S. (2007). A comparison of various tests of normality, *Journal of Statistical Computation and Simulation*, **77**, 175–183.