Received: 19 July 2021 | Revised: 17 April 2022 | Accepted: 20 April 2022

DOI: 10.1002/asmb.2687

RESEARCH ARTICLE

WILEY

A new look at the Birnbaum–Saunders regression model

Marcelo Bourguignon[1] | **Diego I. Gallardo**[2]

[1]Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil

[2]Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile

Correspondence
Marcelo Bourguignon, Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil.
Email: m.p.bourguignon@gmail.com

Abstract

Motivated by problems of vibration in commercial aircraft that caused fatigue in materials, Birnbaum and Saunders (1969a,b) proposed a family of two-parameter distributions to model failure time due to fatigue under cyclic loading and the assumption that failure follows from the development and growth of a dominant crack. This article is cited in at least 930 scientific articles. After the work of Birnbaum and Saunders (1969a,b), the Birnbaum–Saunders model is one of the most important and used models for analyzing lifetime data. Regression analysis is commonly used when covariates are involved in the life test. In this article, we provide a comprehensive review of Birnbaum–Saunders regression models based on parameterizations of the Birnbaum–Saunders distribution in terms of central tendency measures. In particular, we propose a new and straightforward reparameterization of the Birnbaum–Saunders distribution based on an approximation to its mode. In contrast to the original parameterization, the proposed parameterization leads to regression coefficients directly associated (approximately) with the mode of the response variable. We discuss the estimation of model parameters by maximum likelihood, hypothesis testing, and residual analysis. Furthermore, we show that the proposed regression model and other parameterized Birnbaum–Saunders models have the same log-likelihood function estimate when no covariates are used to model the shape parameter. Additionally, we propose a simple script in the R software to solve numerical problems in the Fisher information matrix for some values of the shape parameter. Finally, we present and explore two empirical applications to compare the investigated regression models.

KEYWORDS

Birnbaum–Saunders distribution, data analysis, maximum likelihood, modal regression, reparameterization

1 | INTRODUCTION

Birnbaum and Saunders[1,2] introduced a family of lifetime distributions (named the Birnbaum–Saunders (BS) distribution) to model failure time subjected to cyclic patterns of stresses and strains, and the ultimate failure of the specimen is assumed to be due to the growth of a dominant crack in the material. At each increment of load, this dominant crack extends by a random, non-negative amount. In particular, the BS distribution possesses important features and can be

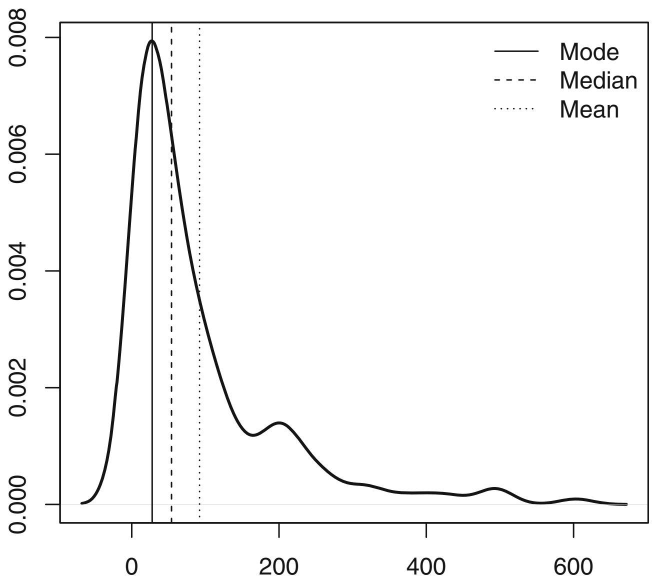

FIGURE 1 Empirical density of stress carbon fibers

viewed as an alternative to the normal distribution when non-negative data following a positively skewed distribution are studied. The precise derivation of the BS distribution is given in Birnbaum and Saunders.[1]

In probability theory, the BS distribution is a family of continuous probability distributions on the positive real line. The BS model corresponds to a unimodal, positively skewed, two-parameter distribution with positive support, characteristics usually present in most of the life distributions, being a good alternative to the gamma, Weibull, and inverse Gaussian distributions. Furthermore, the BS distribution has received attention in several other areas, including finances,[3] earth sciences,[4] and medicine.[5] For more details about the BS distribution, see Johnson et al.,[6] Leiva,[7] Balakrishnan and Kundu,[8] and references therein.

Recently, Dasilva et al.[9] studied three regression models based on the BS distribution: (i) the first is obtained directly (based on the median) from the Birnbaum–Saunders distribution[10]; (ii) the second is obtained via a logarithmic transformation in the response variable[11]; and (iii) the third model employs a mean parameterization of the BS distribution.[12] In particular, Dasilva et al.[9] performed a comparative analysis of these three BS regression models, and we note that the models based on the mean and median present the same fit. Sánchez et al.[13] considered the BS quantile regression model. The BS distribution is skewed, so it can be more intuitive to think in terms of its mode instead of its mean, median, or quantile. However, to the best of our knowledge, a specific parametric modal regression model based on the BS distribution has never been considered in the literature, especially considering that the mode of the BS model has no closed form.

In lifetime data analysis, the data can follow a skew distribution. In this case, it is well known that the widely popular mean regression model could be inadequate if the probability distribution of the observed responses does not follow a symmetric or multimodal distribution. On the other hand, the mode can be more intuitive than the mean, especially for skewed distributions, because the distribution reaches its greatest height when we use the mode.[14] Furthermore, the mode is not affected by outliers. For example, Figure 1 plots the empirical density for the breaking stress of carbon fibers in Gba[15] (mode = 27.76, median = 54, and mean = 92.07). From Figure 1, we can see that the dataset is asymmetrical. Here, the mean is pulled to the tail, making it a less representative measure of central tendency. As the BS distribution is skewed, so it can be more intuitive to think in terms of its mode instead of its mean because the distribution reaches its greatest height when we use the mode. Furthermore, according to Chen et al.,[16] the conditional mean both fails to capture the major trends present in the response, and produces unnecessarily wide prediction bands. Thus, modal regression is an improvement in both of these regards (better trend estimation and narrower prediction bands).

The chief goal of this article is to study several parameterizations of the BS regression model. We provide a comprehensive review of BS regression models based on parameterizations of the BS distribution in terms of central tendency measures. We also provide a detailed study of a general parameterization of the BS regression models. In the general framework, we establish expressions for important characteristics of the BS regression models and discuss the estimation of the parameters via the maximum likelihood (ML) method. Additionally, we propose a novel BS parameterization based on a mode approximation and a parameter ϕ, which is closely related to the distribution variability. Under this parameterization, we propose a regression model and allow a regression structure for the mode parameter, that is, in this new regression model, the modal response is related to a linear predictor through a link function and the linear predictor involves covariates and unknown regression parameters.

The rest of the article proceeds as follows. In Section 2, we present the parameterizations proposed for the BS model until now and introduce a new parameterization of the BS distribution indexed by a mode and dispersion parameter. In Section 3, we describe the BS regression models based on the distributions described in Section 2. In particular, we present the score function and the expected Fisher information matrix for general parameterization of the BS regression models. We also perform hypothesis testing, estimation of the model parameters, and residual analysis. In Section 4, we present and discuss two empirical applications to illustrate and compare the adjustments of the regression models. Finally, some concluding remarks are given in Section 5.

2 | A REVIEW OF REPARAMETERIZATIONS FOR THE BS REGRESSION MODEL

In this section, we shall provide a background of the BS distribution and discuss different parameterizations for the BS model. We also propose a new parameterization based on an approximation of the mode.

2.1 | The original parameterization of the BS model (based on the median)

A random variable Y follows a BS distribution with parameters $\alpha > 0$ (shape) and $\beta > 0$ (scale), denoted by BS(α, β), if it can be written as

$$Y = \beta\left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right]^2,$$

where Z is a standard normal random variable. Its cumulative distribution function (cdf) is given by

$$F(y;\alpha,\beta) = \Phi\left(\frac{1}{\alpha}\zeta\left(\frac{y}{\beta}\right)\right), \quad y > 0, \tag{1}$$

where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution and $\zeta(z) = z - z^{-1}$. As α decreases toward zero, the BS distribution approaches the normal distribution with mean β and variance ξ, where $\xi \to 0$ when $\alpha \to 0$. On the other hand, besides being a scale parameter, β is also the median of the distribution: $F(\beta) = \Phi(0) = 0.5$. We can note that for any $k > 0$, $kY \sim$ BS$(\alpha, k\beta)$. Additionally, the reciprocal property holds, that is, $Y^{-1} \sim$ BS(α, β^{-1}), see Saunders.[17]

The corresponding probability density function (pdf) from (1) is given by

$$f(y;\alpha,\beta) = \kappa(\alpha,\beta)\, y^{-3/2}\,(y+\beta)\,\exp\left[-\frac{\zeta(y/\beta)}{2\alpha^2}\right], \quad y > 0,$$

where $\kappa(\alpha,\beta) = \exp(\alpha^{-2})/(2\alpha\sqrt{2\pi\beta})$. We can verify that this distribution has an upside-down bathtub hazard rate function.[18] For all values of α and β, the pdf is unimodal.

The inverse function of the cdf of a random variable, also known as the quantile function, is defined by $F^{-1}(\tau) = \inf_{x\in\mathbb{R}}\{F(x) \geq \tau\}$, for $\tau \in [0,1]$. Then, the quantile function of Y is

$$F^{-1}(\tau;\alpha,\beta) = \beta\left(\alpha z_\tau/2 + ((\alpha z_\tau/2)^2+1)^{1/2}\right)^2, \quad 0 < \tau < 1,$$

where $z_\tau = \Phi^{-1}(\tau)$, with $\Phi^{-1}(\cdot)$ being the inverse function of the standard normal cdf and F^{-1} is the inverse function of F.

The rth moment of $Y \sim$ BS(α, β) is

$$\mathrm{E}(Y^r) = \beta^r\,(K_{r+1/2}(1/\alpha^2) + K_{r-1/2}(1/\alpha^2))/2K_{1/2}(1/\alpha^2),$$

with $K_\nu(u)$ denoting the modified Bessel function of the third kind of order ν and argument u given by

$$K_\nu(u) = \frac{1}{2}\left(\frac{u}{2}\right)^\nu \int_0^\infty w^{-\nu-1} \exp\left(-w - \frac{u^2}{4w}\right) \mathrm{d}w;$$

see Gradshteyn and Randzhik.[19(p. 907)] In particular, the expected value and variance are $\mathrm{E}(Y) = \beta(1 + \alpha^2/2)$ and $\mathrm{Var}(Y) = (\alpha\beta)^2(1 + 5\alpha^2/4)$, respectively.

2.2 | A mean-parameterized version of the BS model

Santos-Neto et al.[20,21] introduced a mean-parameterized version of the BS model considering $\mu = \beta(1 + \alpha^2)$ and $\phi = 2/\alpha^2$, that is, $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \varpi(\mu)\varsigma(\phi)$, where $\varpi(\mu) = \mu^2$ acts as a "variance function" and $\varsigma(\phi) = \sqrt{2\phi + 5}/(\phi + 1)$ is the coefficient of variation (CV) of Y ($0 < \mathrm{CV} < \sqrt{5}$), which only depends on ϕ. This parameterization is denoted as $\mathrm{RBS}(\mu, \phi)$. Leiva et al.[12] studied in details the case where a regression structure is considered only in μ and Santos-Neto et al.[22] discussed the case where μ and ϕ are modeled using covariates.

2.3 | A quantile-parameterized version of the BS model

Sánchez et al.[13] discussed a version of the BS model parameterized in terms of the τth quantile, where $0 < \tau < 1$ is a (fixed) quantile of interest. This parameterization considers $\phi = \alpha$ and $\mu = (\beta/4)\left(\alpha z_\tau + \sqrt{\alpha^2 z_\tau^2 + 4}\right)^2$, where μ represents the τth quantile of the distribution. We denote this parameterization by $\mathrm{RBSQ}_\tau(\mu, \phi)$. As $\tau = 0.5$ represents the median, it follows directly that $\mathrm{RBSQ}_{\tau=0.5}(\mu, \phi) = \mathrm{BS}(\beta = \mu, \alpha = \phi)$. The authors discussed the case where a regression structure is used only in μ.

2.4 | A mode-parameterized BS model (new)

Here, we propose a new parameterization of the BS model based on an approximation of its mode. The mode of the BS distribution (denoted by $m_{\alpha,\beta}$) is obtained as the solution of the following non-linear equation:[7]

$$(\beta - m_{\alpha,\beta})(y + \beta)^2 = \alpha^2\beta\, m_{\alpha,\beta}(m_{\alpha,\beta} + 3\beta). \quad (2)$$

Thus, the mode of $\mathrm{BS}(\alpha, \beta)$ cannot be obtained in an explicit form. In the case of a moderately skewed distribution (for the BS distribution this implies $\alpha < 1$), the difference between the mean and mode is almost three times the difference between the mean and median (by the Karl Pearson's formula). Hence, the empirical mean, median, and mode relation for the BS distribution is given by $\beta(1 - \alpha^2/2) - m_{\alpha,\beta} \approx 3(\beta(1 - \alpha^2/2) - \beta)$. Then, $m_{\alpha,\beta} = \beta(1 - \alpha^2)$ acts as an approximation to the mode, for $\alpha < 1$. The major limitation of this approximation is the restriction of the parameter α. In order to assess the accuracy of mode approximations, Table 1 shows the comparison for different values of the $\mathrm{BS}(\alpha, \beta)$ model. Note that the approximation works well for α small (independently of the value for β) and not well for α close to 1. In the last case, the error in the approximation is greater when β is also increased.

Now, consider the parameterization $\mu = \beta(1 - \alpha^2)$ and $\phi = \alpha^2$, that is, $\beta = \mu/(1 - \phi)$ and $\alpha = \sqrt{\phi}$. We note that this parameterization has not been proposed in the statistical literature. Under this new parameterization, we have that

$$\mathrm{Mode}(Y) = \mu \quad \text{and} \quad \mathrm{Var}(Y) = \mu^2 \frac{\phi\,(1 + 5\phi/4)}{1 - \phi},$$

where $0 < \phi < 1$ modifies the shape of the distribution. Furthermore, ϕ can be interpreted as a dispersion parameter, that is, for fixed values of μ and $\phi \to 1$, the variance of Y tends to ∞. We denote by $\mathrm{RBSM}(\mu, \phi)$ the random variable Y following this specific parameterization of the BS distribution. Figure 2A,B show the pdf of $Y \sim \mathrm{RBSM}(\mu, \phi)$ for different values of ϕ and fixed μ. Figure 2 shows the densities of the BS distribution for which the mode falls at $y \approx 2.0$. We can confirm from this figure that the parameter ϕ controls the shape of the pdf.

TABLE 1 Comparison between the true mode and its approximation ($m_{\alpha,\beta}$) for different BS models.

	$\beta = 0.2$		$\beta = 0.5$		$\beta = 1.0$		$\beta = 5.0$		$\beta = 10.0$		$\beta = 100.0$	
α	$m_{\alpha,\beta}$	True	$m_{\alpha,\beta}$	True	$m_{\alpha,\beta}$	True	$m_{\alpha,\beta}$	True	$m_{\alpha,\beta}$	True	$m_{\alpha,\beta}$	True
0.05	0.200	0.199	0.499	0.499	0.998	0.997	4.988	4.988	9.975	9.975	99.750	99.750
0.10	0.198	0.198	0.495	0.495	0.990	0.990	4.950	4.950	9.900	9.900	99.000	99.002
0.25	0.188	0.188	0.469	0.469	0.938	0.939	4.688	4.693	9.375	9.385	93.750	93.852
0.50	0.150	0.154	0.375	0.384	0.750	0.769	3.750	3.843	7.500	7.685	75.000	76.850
0.75	0.087	0.109	0.219	0.272	0.437	0.545	2.187	2.724	4.375	5.448	43.750	54.480
0.95	0.020	0.077	0.049	0.193	0.098	0.386	0.488	1.930	0.975	3.860	9.750	38.598

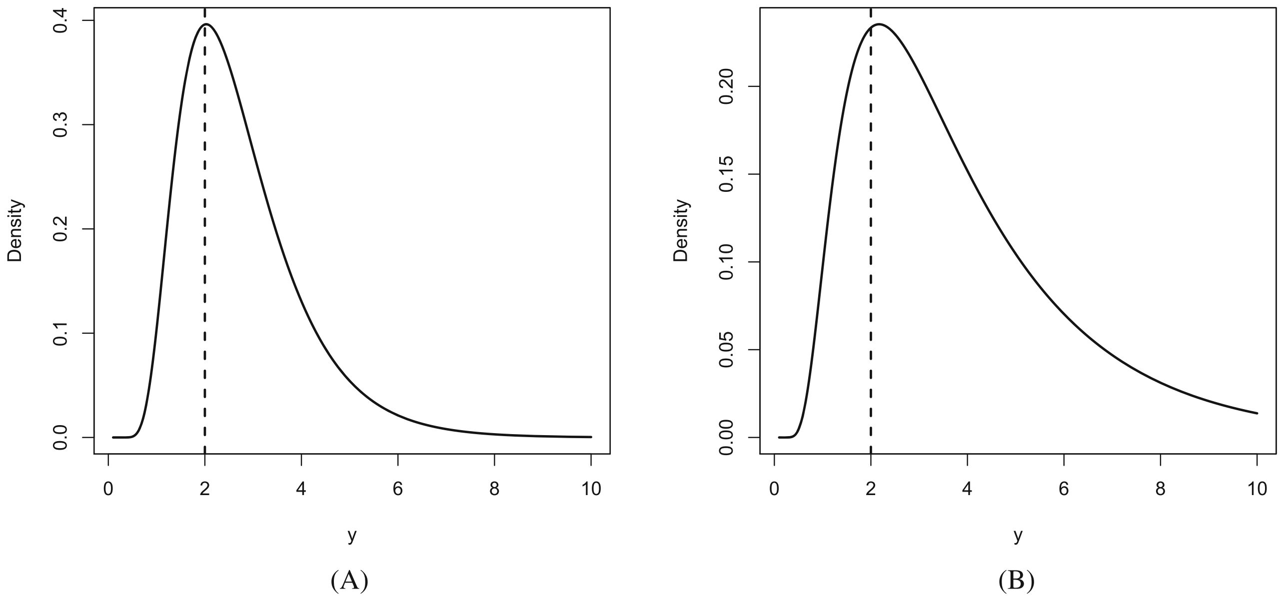

FIGURE 2 RBSM($\mu = 2, \phi$) density function for different values of ϕ. (A) $\phi = 0.2$; (B) $\phi = 0.4$

Remark 1. From Equation (2),

$$\alpha = \frac{\sqrt{\beta - \mu}(\mu + \beta)}{\sqrt{\beta\ \mu(\mu + 3\beta)}},$$

can be expressed in terms of μ and β, and consequently the pdf of the BS distribution can be re-parameterized in term of μ and β. However, in this case, the BS distribution is parameterized by two positive scale parameters, so the distribution is less flexible. Moreover, there is a strong functionally dependency ($\mu < \beta$) between β and μ, implying that the process of obtaining the parameter estimates of the model requires particular care because of the constraints imposed on its parameters.

2.5 | Other parameterizations of the BS model

Santos-Neto et al.[20] discussed nine alternative parameterizations of the model based on physical aspects, generalized linear models, variance, Tweedie models, among others. However, in this article, we consider the four parameterization previously discussed in this section as they are more relevant in terms of interpretability and, in practice, the mean, median, mode, and quantile are the most common used statistics.

3 | REGRESSION ANALYSIS AND INFERENCE ON BS MODELS

In this section, we shall describe the BS regression models based on the parameterizations described in the previous section. We also establish expressions for important characteristics of the BS regression models, discuss estimation and inference of the parameters, and residual analysis via maximum likelihood estimation.

3.1 | Modeling the mean/median/quantile/mode

Let $Y_1, \dots, Y_n$ be n independent random variables, where each Y_i, $i = 1, \dots, n$, follows a BS, RBS, RBSQ$_\tau$, or RBSM model. For the BS(α, β) model, suppose that the median and shape parameters of Y_i satisfy the following functional relations:

$$g(\beta_i) = \mathbf{x}_i^\top \lambda \quad \text{and} \quad h(\alpha_i) = \mathbf{z}_i^\top \boldsymbol{\eta}, \tag{3}$$

where $\lambda = (\lambda_1, \dots, \lambda_q)^\top \in \mathbb{R}^q$ and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_r)^\top \in \mathbb{R}^r$ are vectors of unknown regression coefficients, with $q + r < n$, and $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})^\top$ and $\mathbf{z}_i = (z_{i1}, \dots, z_{ir})^\top$ are observations of q and r known regressors related to the median and shape parameters, respectively, for $i = 1, \dots, n$. Here, we assume that the covariate matrices $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ and $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^\top$ have rank q and r, respectively. The link functions $g, h : \mathbb{R} \to \mathbb{R}^+$ in (3) must be strictly monotone, positive, and at least twice differentiable, such that $\beta_i = g^{-1}(\mathbf{x}_i^\top\, \boldsymbol{\nu})$ and $\alpha_i = h^{-1}(\mathbf{z}_i^\top\, \xi)$, with $g^{-1}(\cdot)$ and $h^{-1}(\cdot)$ being the inverse functions of $g(\cdot)$ and $h(\cdot)$, respectively. There are several possible choices for the link functions $g(\cdot)$ and $h(\cdot)$. For instance, one can use the logarithmic specification $g(\cdot) = \log(\cdot)$, square root $g(\cdot) = \sqrt{\cdot}$, or identity $g(\cdot) = \cdot$ (with special attention to the positivity of the estimates).

Analogously, suppose that μ and ϕ for the RBS(μ, ϕ), RBSQ$_\tau$(μ, ϕ), or RBSM(μ, ϕ) models satisfy

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\nu} \quad \text{and} \quad h(\phi_i) = \mathbf{z}_i^\top \boldsymbol{\xi}, \tag{4}$$

where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_q)^\top$ and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_r)^\top$ are newly vectors of unknown regression coefficients. The link functions $g, h : \mathbb{R} \to \mathbb{R}^+$ (except $h : \mathbb{R} \to (0, 1)$ for the RBSM model) in (4) must be strictly monotone, positive, and at least twice differentiable, such that $\mu_i = g^{-1}(\mathbf{x}_i^\top\, \boldsymbol{\nu})$ and $\phi_i = h^{-1}(\mathbf{z}_i^\top\, \xi)$, with $g^{-1}(\cdot)$ and $h^{-1}(\cdot)$ being the inverse functions of $g(\cdot)$ and $h(\cdot)$, respectively. Note that the RBS, RBSQ$_\tau$, and RBSM models satisfy the relation $\mu_i = \beta_i \cdot \rho(\alpha_i)$, with $\rho(\alpha) = 1 + \alpha^2/2$ for the RBS model, $\rho(\alpha) = \frac{1}{4}(\alpha\, z_\tau + \sqrt{\alpha\, z_\tau^2 + 4})^2$ for the RBSQ$_\tau$ model, and $\rho(\alpha) = 1 - \alpha^2$ for the RBSM model. We establish the following Theorem related to the regression coefficients of the BS, RBS, RBSQ$_\tau$, and RBSM models.

Theorem 1. *If $Y_1, \dots, Y_n$ are independent such as $Y_i \sim BS(\beta_i, \alpha_i)$, where $g(\beta_i) = \mathbf{x}_i^\top \lambda$, $i = 1, \dots, n$ and suppose that $h(\alpha_i) = \eta_0$ (i.e., we are modeling only the median parameter with covariates). Suppose that we assume an alternative parameterization for the BS model, $Y_i \sim RBS(\mu_i, \phi_i)$, $Y_i \sim RBSQ_\tau(\mu_i, \phi_i)$, or $Y_i \sim RBSM(\mu_i, \phi_i)$, with $g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\nu}$ (the same g is used) and similarly, $h(\phi_i) = \xi_0$. If intercept terms are considered in $\mathbf{x}_i^\top$ and $g(\cdot)$ is the (i) log function, (ii) identity function, or (iii) square root, then the elements of the vector $(\boldsymbol{\nu}^\top, \xi_0)$ can be obtained uniquely from $(\lambda^\top, \eta_0)$.*

Proof. Note that the RBS, RBSQ$_\tau$, and RBSM models satisfy $\mu_i = \beta_i \cdot \rho(\alpha)$. Considering the partition $\mathbf{x}^\top = (1, \mathbf{x}^{\star\top})^\top$, $\lambda^\top = (\lambda_0, \lambda^{\star\top})$, and $\boldsymbol{\nu}^\top = (\nu_0, \boldsymbol{\nu}^{\star\top})$ (note that $\mathbf{x}^{\star\top}$, $\lambda^{\star\top}$, and $\boldsymbol{\nu}^{\star\top}$ are related to the non-intercept terms), we have

$$g^{-1}(\lambda_0 + \mathbf{x}_i^{\star\top} \lambda^\star) = g^{-1}(\nu_0 + \mathbf{x}_i^{\star\top} \boldsymbol{\nu}^\star) \times \rho(\alpha). \tag{5}$$

As g is invertible, we have the following cases:

i) For $g(\cdot) = \log(\cdot)$, Equation (5) implies

$$\lambda_0 + \mathbf{x}_i^{\star\top} \lambda^\star = \nu_0 + \log \rho(\alpha) + \mathbf{x}_i^{\star\top} \boldsymbol{\nu}^\star,$$

where we conclude that $\lambda_0 = \nu_0 + \log \rho(\alpha)$ and $\lambda^\star = \boldsymbol{\nu}^\star$.

ii) For $g(\cdot) = \cdot$, Equation (5) implies

$$\lambda_0 + \mathbf{x}_i^{\star\top} \lambda^\star = \rho(\alpha) \times \nu_0 + \mathbf{x}_i^{\star\top} \left[\rho(\alpha) \times \boldsymbol{\nu}^\star\right],$$

concluding that $\lambda_0 = \rho(\alpha) \times \nu_0$ and $\lambda^\star = \rho(\alpha) \times \boldsymbol{\nu}^\star$.

iii) For $g(\cdot) = \sqrt{\cdot}$, Equation (5) implies

$$\lambda_0 + \mathbf{x}_i^{\star\top}\boldsymbol{\lambda}^\star = \sqrt{\rho(\alpha)} \times \nu_0 + \mathbf{x}_i^{\star\top}\left[\sqrt{\rho(\alpha)} \times \boldsymbol{\nu}^\star\right],$$

then $\lambda_0 = \sqrt{\rho(\alpha)} \times \nu_0$ and $\lambda^\star = \sqrt{\rho(\alpha)} \times \boldsymbol{\nu}^\star$.

On the other hand, $\phi = \varphi(\alpha)$, with $\varphi(\alpha) = 2/\alpha^2$, $\varphi(\alpha) = \alpha$, and $\varphi(\alpha) = \alpha^2$ for the RBS, RBSQ$_\tau$, and RBSM models, respectively. In all cases, α can be determined in a unique form given ϕ and h invertible, we then conclude that the vector $(\boldsymbol{\lambda}^\top, \eta_0)$ is uniquely determined from $(\boldsymbol{\nu}^\top, \xi_0)$. ■

Remark 2. Theorem 1 implies that, when no covariates are used to model the shape parameter α in the BS regression model and parameter ϕ in the RBS, RBSQ$_\tau$, or RBSM regression models, all the models define the same pdf. Consequently, model selection criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) provide the same value. Note that Leiva et al.[12] and Sánchez et al.[13] introduced covariates only in the mean and quantile parameters, respectively.

Remark 3. When covariates (additional to the intercept term) are used to model both parameters, β and α in the BS regression model (or μ and ϕ in the RBS, RBSQ$_\tau$, and RBSM models), the models define different pdf's. This is simple to visualize because in this case the relation $\mu_i = \beta_i \cdot \rho(\alpha_i)$ involves only the covariates $\mathbf{x}_i$ on the left side of the equality but both sets of covariates on the right side. For instance, Santos-Neto et al.[22] introduced covariates in both the mean and dispersion parameters (μ and ϕ, respectively).

3.2 | Score vector and the Fisher information matrix

In order to unify the notation, observe that the BS regression model can be seen as a special case with $\rho(\alpha) = 1$ and $\varphi(\alpha) = \alpha$, and let $\boldsymbol{\theta} = (\boldsymbol{\nu}^\top, \boldsymbol{\xi}^\top)^\top$ be the parameter vector. Under this notation, the score vector for the BS, RBS, RBSQ$_\tau$, and RBSM regression models is given by

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\nu} \\ \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\xi} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{A}_\mu\left(\mathbf{C}_\mu^\beta\mathbf{D}_\beta + \mathbf{C}_\mu^\alpha\mathbf{D}_\alpha\right) \\ \mathbf{Z}^\top\mathbf{A}_\phi\left(\mathbf{C}_\phi^\beta\mathbf{D}_\beta + \mathbf{C}_\phi^\alpha\mathbf{D}_\alpha\right) \end{bmatrix}, \tag{6}$$

where $\mathbf{A}_u = \mathrm{diag}(a_u^{(1)}, \dots, a_u^{(n)})$, $\mathbf{D}_u = \mathrm{diag}(d_u^{(1)}, \dots, d_u^{(n)})$, for $u \in \{\alpha, \beta\}$, $a_\mu^{(i)} = 1/g(\mu_i)$, $a_\phi^{(i)} = 1/h(\phi_i)$,

$$d_\beta^{(i)} = \frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\beta_i} = -\frac{1}{2\beta_i} + \frac{1}{y_i + \beta_i} - \frac{1}{2\alpha_i^2}\left[\frac{-y_i}{\beta_i^2} + \frac{1}{y_i}\right],$$
$$d_\alpha^{(i)} = \frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\alpha_i} = -\frac{2}{\alpha_i^3} - \frac{1}{\alpha_i} + \frac{1}{\alpha_i^3}\left[\frac{y_i}{\beta_i} + \frac{\beta_i}{y_i}\right],$$

and $\mathbf{C}_u^v = \mathrm{diag}(C_u^{v(1)}, \dots, C_u^{v(n)})$, for $u \in \{\mu, \phi\}$, $v \in \{\beta, \alpha\}$, and $C_u^{v(i)} = \partial u_i/\partial v_i$. Note that $C_\mu^{\beta(i)} = \rho(\alpha_i)$, $C_\mu^{\alpha(i)} = \beta_i\,\rho'(\alpha_i)$, $C_\phi^{\alpha(i)} = \varphi'(\alpha_i)$, and $C_\phi^{\beta(i)} = 0$. Table 2 displays some useful quantities for the RBS, RBSQ$_\tau$, and RBSM models.

The Fisher information matrix for the BS, RBS, RBSQ$_\tau$, and RBSM regression models can be written as

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E}\begin{bmatrix} -\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}^\top & -\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\nu}\partial\boldsymbol{\xi}^\top \\ \cdot & -\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^\top \end{bmatrix} = \widetilde{\mathbf{X}}^\top\mathbf{W}(\boldsymbol{\theta})\widetilde{\mathbf{X}},$$

where

$$\widetilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X} & \mathbf{0}_{p\times1} \\ \mathbf{0}_{q\times1} & \mathbf{Z} \end{pmatrix} \quad \text{and} \quad \mathbf{W}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^\top\mathbf{A}_\mu\mathbf{V}_{\beta\beta}\mathbf{A}_\mu\mathbf{X} & \mathbf{X}^\top\mathbf{A}_\mu\mathbf{V}_{\beta\alpha}\mathbf{A}_\phi\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{A}_\phi\mathbf{V}_{\alpha\beta}\mathbf{A}_\mu\mathbf{X} & \mathbf{Z}^\top\mathbf{A}_\phi\mathbf{V}_{\alpha\alpha}\mathbf{A}_\phi\mathbf{Z} \end{bmatrix},$$

TABLE 2 Partial derivatives of different parameterizations for the BS models

Model	$\rho(\alpha_i)$	$\rho'(\alpha_i)$	$\varphi(\alpha_i)$	$\varphi'(\alpha_i)$
BS	1	0	α_i	1
RBS	$1+\alpha_i^2/2$	α_i	$2/\alpha_i^2$	$-4/\alpha_i^3$
RBSQ$_\tau$	$\frac{1}{4}(\alpha_i z_q+\sqrt{\alpha_i z_q^2+4})^2$	$\frac{z_q}{2}\frac{\alpha_i z_q+\sqrt{\alpha_i z_q^2+4}}{\left(1+\frac{z_q}{2\sqrt{\alpha_i z_q^2+4}}\right)}$	α_i	1
RBSM	$1-\alpha_i^2$	$-2\alpha_i$	α_i^2	$2\alpha_i$

with $\mathbf{V}_{uv} = \mathrm{diag}(V_{uv}^{(1)}, \ldots, V_{uv}^{(n)})$, for $u, v \in \{\alpha, \beta\}$, $V_{\beta\beta}^{(i)} = I_{\beta\beta}^{(i)}\left[C_\mu^{\beta(i)}\right]^2 + I_{\alpha\alpha}^{(i)}\left[C_\mu^{\alpha(i)}\right]^2$, $V_{\beta\alpha}^{(i)} = V_{\alpha\beta}^{(i)} = I_{\beta\beta}^{(i)} C_\mu^{\beta(i)} C_\phi^{\beta(i)} + I_{\alpha\alpha}^{(i)} C_\mu^{\alpha(i)} C_\phi^{\alpha(i)}$, and $V_{\alpha\alpha}^{(i)} = I_{\beta\beta}^{(i)}\left[C_\phi^{\beta(i)}\right]^2 + I_{\alpha\alpha}^{(i)}\left[C_\phi^{\alpha(i)}\right]^2$, with

$$I_{\beta\beta}^{(i)} = \mathrm{E}\left(\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \beta_i^2}\right) = \mathrm{E}\left(\frac{1}{2\beta_i^2} - \frac{1}{(y_i+\beta_i)^2} - \frac{y_i}{\alpha_i^2\beta_i^3}\right) = \frac{1+\alpha_i(2\pi)^{-1}\psi(\alpha_i)}{\alpha_i^2\beta_i^2},$$
$$I_{\alpha\alpha}^{(i)} = \mathrm{E}\left(\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \alpha_i^2}\right) = \mathrm{E}\left(\frac{6}{\alpha_i^4} + \frac{1}{\alpha_i^2} - \frac{3}{\alpha_i^4}\left[\frac{y_i}{\beta_i} + \frac{\beta_i}{y_i}\right]\right) = \frac{2}{\alpha_i},$$

where $\psi(\alpha) = \alpha(\pi/2)^{1/2} - (\pi/2)e^{2/\alpha^2}\left[1-\Phi(2/\alpha)\right]$. Note that

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{F} \\ -\mathbf{F}^\top\mathbf{E}^{-1} & \mathbf{G}^{-1}+\mathbf{F}^\top\mathbf{E}^{-1}\mathbf{F} \end{bmatrix},$$

with $\mathbf{E} = \mathbf{X}^\top\mathbf{A}_\mu\mathbf{V}_{\beta\beta}\mathbf{A}_\mu\mathbf{X} - \mathbf{F}\mathbf{Z}^\top\mathbf{A}_\phi\mathbf{V}_{\alpha\beta}\mathbf{A}_\mu\mathbf{X}$, $\mathbf{F} = \mathbf{X}^\top\mathbf{A}_\mu\mathbf{V}_{\beta\alpha}\mathbf{A}_\phi\mathbf{Z}\left(\mathbf{Z}^\top\mathbf{A}_\phi\mathbf{V}_{\alpha\alpha}\mathbf{A}_\phi\mathbf{Z}\right)^{-1}$, and $\mathbf{G} = \mathbf{Z}^\top\mathbf{A}_\phi\mathbf{V}_{\alpha\alpha}\mathbf{A}_\phi\mathbf{Z}$.

Remark 4. $\mathbf{I}^{-1}(\boldsymbol{\theta})$ is reduced to a diagonal block matrix for the BS regression model whereas for the rest of the models it is not. This implies that the ML estimators for the BS regression models are asymptotically independent whereas for the RBS, RBSQ$_\tau$, and RBSM models they are not. Moreover, it is immediate that the asymptotic covariance matrix for the ML estimators of $\boldsymbol{\nu}$ and $\boldsymbol{\xi}$ is $\mathrm{Cov}(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\xi}}) = -\mathbf{E}^{-1}\mathbf{F}$.

Remark 5. Lemonte[23] discussed computational problems for the function $\psi(\alpha)$ in the R software[24] when α is small (see table 1 in Lemonte[23]) and proposed an approximated function for it. The authors presented an R program to compute $\psi(\alpha)$ that have the simplified form

```
fh<-function(alpha)
return(alpha*sqrt(pi/2)-pi*exp(2/(alpha^2))*(1 - pnorm(2/alpha)))
```

We tested this function in R version 4.0.2. and it effectively produced `NaN` (not a number) for $\alpha <\approx 0.053$ or inclusive a wrong value. This is a very critical point for the RBSM model because the approximation to the mode works well to small values of α. However, a very little modification in the code (without any approximation) can avoid the problem, such as

```
fh.alt<-function(alpha)
return(alpha*sqrt(pi/2)-pi*exp(2/(alpha^2)
+pnorm(2/alpha,log=TRUE,lower.tail=FALSE)))
```

Figure 3 displays the functions `fh` and `fh.alt` when $\alpha \in (0, 0.3)$. From Figure 3, it is evident that the function `fh` presents numerical problems for small values of α, whereas `fh.alt` does not. Thus, we recommend using the `fh.alt` function to make inference on BS model parameters.

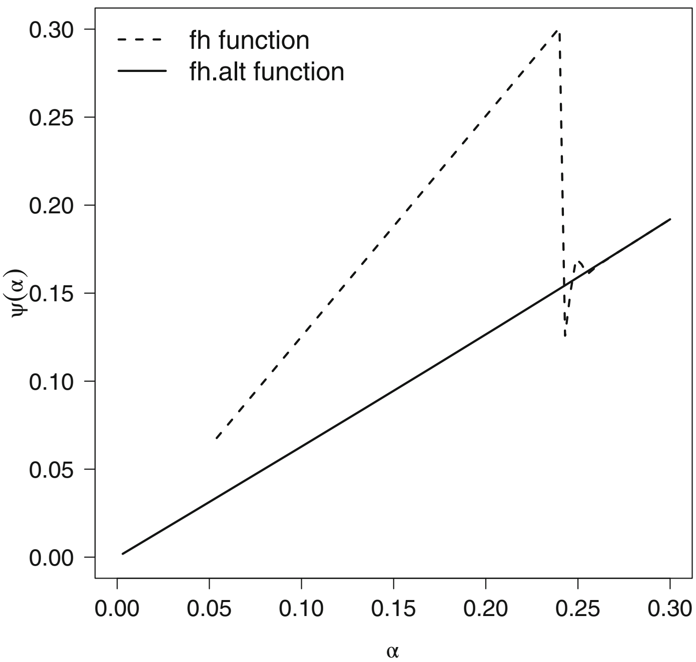

FIGURE 3 `fn` and `fn.alt` functions

3.3 | Estimation

Let $Y_1, \dots, Y_n$ be n independent random variables, where each $Y_i, i = 1, \dots, n$, follows a BS, RBS, RBSQ$_\tau$, or RBSM model. The log-likelihood function for $\boldsymbol{\theta}$ has the form

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\nu}, \boldsymbol{\xi}) = \sum_{i=1}^{n} \ell_i(\alpha_i, \beta_i; y_i),$$

where

$$\ell_i(\alpha_i, \beta_i; y_i) = \frac{1}{\alpha_i^2} - \log(\alpha_i) - \frac{1}{2}\log(\beta_i) + \log(y_i + \beta_i) - \frac{\zeta(y_i/\beta_i)}{2\alpha_i^2},$$

with $\alpha_i = \alpha_i(\phi_i) = \varphi^{-1}(\phi_i)$ and $\beta_i = \beta_i(\mu_i, \phi_i) = \mu_i/\rho(\varphi^{-1}(\phi_i))$. The ML estimates of $\boldsymbol{\nu}$ and $\boldsymbol{\xi}$ are computed as the solution of the non-linear system $\boldsymbol{U}(\boldsymbol{\theta}) = \mathbf{0}_{q+r}$, where $\boldsymbol{U}(\boldsymbol{\theta})$ is defined in Equation (6) and $\mathbf{0}_{q+r}$ denotes a $(q+r) \times 1$ vector of zeros. An alternative way to obtain the ML estimates of $\boldsymbol{\theta}$ is by using the Fisher scoring iterative procedure, providing the following estimation algorithm:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{(k+1)} &= \widehat{\boldsymbol{\theta}}^{(k)} + \left[\mathbf{I}\left(\widehat{\boldsymbol{\theta}}^{(k)}\right)\right]^{-1} \times \mathbf{U}\left(\widehat{\boldsymbol{\theta}}^{(k)}\right)\\ &= \widehat{\boldsymbol{\theta}}^{(k)}\left[\widetilde{\mathbf{X}}^\top \mathbf{W}\left(\widehat{\boldsymbol{\theta}}^{(k)}\right)\widetilde{\mathbf{X}}\right]^{-1}\widetilde{\mathbf{X}}^\top \mathbf{W}_1\left(\widehat{\boldsymbol{\theta}}^{(k)}\right), \quad k = 0, 1, 2, \dots,\end{aligned}$$

where $\mathbf{W}_1(\boldsymbol{\theta}) = \left(\left(\mathbf{C}_\mu^\beta \mathbf{D}_\beta + \mathbf{C}_\mu^\alpha \mathbf{D}_\alpha\right)\mathbf{A}_\mu, \left(\mathbf{C}_\phi^\beta \mathbf{D}_\beta + \mathbf{C}_\phi^\alpha \mathbf{D}_\alpha\right)\mathbf{A}_\phi\right)^\top$. We suggest initializing the algorithm using as an initial guess for $\boldsymbol{\nu}$ the ordinary least squares estimates obtained from the linear regression of the transformed responses $g(y_1), \dots, g(y_n)$ on $\mathbf{X}$, that is, $\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Z}$, where $\mathbf{Z} = (g(y_1), \dots, g(y_n))^\top$.

In order to approach interval estimation and hypothesis testing on the model parameters $\boldsymbol{\nu}$ and $\boldsymbol{\xi}$, normal approximation for the ML estimators can be applied.[10] Note that under certain conditions for the parameters, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is multivariate normal $N_{q+r}(\mathbf{0}_{q+r}, \mathbf{I}^{-1}(\boldsymbol{\theta}))$.

Let θ_j be the jth component of $\boldsymbol{\theta}$. The asymptotic $100(1-\gamma)\%$ confidence interval for θ_j is given by

$$\widehat{\theta}_j \pm z_{1-\gamma/2}\ \mathrm{se}\left(\widehat{\theta}_j\right), \qquad j = 1, \ldots, q+r,$$

where $\mathrm{se}\left(\widehat{\theta}_r\right)$ is the asymptotic standard error of $\widehat{\theta}_j$, which is the square root of the jth diagonal element of the inverse of $\mathbf{I}(\widehat{\boldsymbol{\theta}})$.

3.4 | Hypothesis testing

In the BS regression model, hypothesis testing inference is usually performed using the gradient (GR), likelihood ratio (LR), score (SC), and Wald (WA) tests.[22] Here, three hypothesis tests are of interest. In all cases, we include an intercept in both structures (μ and ϕ). For the BS regression models, such hypotheses are:

1) $H_0^{(1)} : (\boldsymbol{\nu}, \boldsymbol{\xi}) \in \boldsymbol{\Omega}_0^{(1)}$ versus $H_A^{(1)} : \boldsymbol{\nu} \notin \boldsymbol{\Omega}_0^{(1)}$, where $\boldsymbol{\Omega}_0^{(1)} = \{(\boldsymbol{\nu}, \boldsymbol{\xi}) : \nu_2 = \ldots = \nu_q = 0\}$. (i.e., $\alpha_i = \alpha_0$, $i = 1, \ldots, n$, are constants).
2) $H_0^{(2)} : (\boldsymbol{\nu}, \boldsymbol{\xi}) \in \boldsymbol{\Omega}_0^{(2)}$ versus $H_A^{(2)} : (\boldsymbol{\nu}, \boldsymbol{\xi}) \notin \boldsymbol{\Omega}_0^{(2)}$, where $\boldsymbol{\Omega}_0^{(2)} = \{(\boldsymbol{\nu}, \boldsymbol{\xi}) : \xi_2 = \ldots = \xi_r = 0\}$. (i.e., $\beta_i = \beta_0$, $i = 1, \ldots, n$, are constants).
3) $H_0^{(3)} : (\boldsymbol{\nu}, \boldsymbol{\xi}) \in \boldsymbol{\Omega}_0^{(3)}$ versus $H_A^{(3)} : (\boldsymbol{\nu}, \boldsymbol{\xi}) \notin \boldsymbol{\Omega}_0^{(3)}$, where $\boldsymbol{\Omega}_0^{(3)} = \{(\boldsymbol{\nu}, \boldsymbol{\xi}) : \nu_2 = \ldots = \nu_q = \xi_2 = \ldots \xi_r = 0\}$ (i.e., $\beta_i = \beta_0$ and $\alpha_i = \alpha_0$, $i = 1, \ldots, n$, are both constants).

Such hypotheses can be tested based on the GR, LR, SC, and WA test statistics, which under $H_0^{(j)}$, $j = 1, 2, 3$ follow asymptotically a $\chi^2_{\#\boldsymbol{\Omega}_0^{(j)}}$ distribution, where $\#\boldsymbol{\Omega}_0^{(j)}$ is the number of parameters being tested in $\boldsymbol{\Omega}_0^{(j)}$, that is, $\#\boldsymbol{\Omega}_0^{(1)} = q - 1$, $\#\boldsymbol{\Omega}_0^{(2)} = r - 1$, and $\#\boldsymbol{\Omega}_0^{(3)} = q + r - 2$. Similar hypotheses can be proposed for the RBS, RBSQ$_\tau$, and RBSM regression models in terms of $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$. However, for $H_0^{(1)}$ and $H_0^{(2)}$, they cannot be expressed in closed form. Then, in the following subsection we define expressions for the GR, LR, SC, and WA tests for $H_0^{(3)}$. We denote $\widetilde{\boldsymbol{\theta}}$ as the ML estimator of $\boldsymbol{\theta}$ under $H_0^{(3)}$ (or any function or element of this vector) and $\widehat{\boldsymbol{\theta}}$ the ML estimator under $H_A^{(3)}$.

3.4.1 | LR test

The LR test statistic for the RBS model is given by $\mathrm{LR} = 2\sum_{i=1}^{n} \Lambda_i$, where

$$\Lambda_i = \left(\widehat{\alpha}_i\right)^{-2} - \tilde{\alpha}^{-2} - \log\left(\frac{\widehat{\alpha}_i}{\tilde{\alpha}}\right) - \frac{1}{2}\log\left(\frac{\widehat{\beta}_i}{\tilde{\beta}}\right) + \log\left(\frac{y_i + \widehat{\beta}_i}{y_i + \tilde{\beta}}\right) - \frac{1}{2}\left(\frac{\zeta(y_i/\widehat{\beta}_i)}{\widehat{\alpha}_i} - \frac{\zeta(y_i/\tilde{\beta})}{\tilde{\alpha}}\right),$$

for $i = 1, \ldots, n$, $\widehat{\alpha}_i$ and $\widehat{\beta}_i$ are the ML estimators of α_i and β_i under $H_A^{(3)}$, respectively, and $\tilde{\alpha}$ and $\tilde{\beta}$ are the ML estimators of α_i and β_i under $H_0^{(3)}$, respectively.

3.4.2 | SC test

The matrices $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{G}$ are simplified to $E_0 = b_{\tilde{\mu}\tilde{\mu}}\mathbf{X}^\top\mathbf{X} - b_{\tilde{\mu}}b_{\tilde{\phi}}F_0\mathbf{Z}^\top\mathbf{X}$, $F_0 = b_{\tilde{\mu}}a_{\tilde{\phi}}\mathbf{X}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}$, and $\mathbf{G}_0 = b_{\tilde{\phi}\tilde{\phi}}\mathbf{Z}^\top\mathbf{Z}$, where $b_{\tilde{\mu}\tilde{\mu}} = a_{\tilde{\mu}}^2 V_{\tilde{\beta}\tilde{\beta}}$, $b_{\tilde{\mu}\tilde{\phi}} = a_{\tilde{\mu}}a_{\tilde{\phi}}V_{\tilde{\beta}\tilde{\alpha}}$, and $b_{\tilde{\phi}\tilde{\phi}} = a_{\tilde{\phi}}^2 V_{\tilde{\alpha}\tilde{\alpha}}$. Therefore, the SC statistic can be written as

$$\begin{aligned}\mathrm{SC} = {} & a_{\tilde{\mu}}^2\left(C_{\tilde{\mu}}^{\tilde{\beta}}d_{\tilde{\beta}} + C_{\tilde{\mu}}^{\tilde{\alpha}}d_{\tilde{\alpha}}\right)^2 \mathrm{X}^\top\mathbf{E}_0^{-1}\mathbf{X} + a_{\tilde{\phi}}^2\left(C_{\tilde{\phi}}^{\tilde{\beta}}d_{\tilde{\beta}} + C_{\tilde{\phi}}^{\tilde{\alpha}}d_{\tilde{\alpha}}\right)^2 \mathrm{Z}^\top\left[\mathbf{G}_0^{-1} + \mathbf{F}_0^\top\mathbf{E}_0^{-1}\mathbf{F}_0\right]\mathbf{Z} \\ & - 2a_{\tilde{\mu}}a_{\tilde{\phi}}\left(C_{\tilde{\mu}}^{\tilde{\beta}}d_{\tilde{\beta}} + C_{\tilde{\mu}}^{\tilde{\alpha}}d_{\tilde{\alpha}}\right)\left(C_{\tilde{\phi}}^{\tilde{\beta}}d_{\tilde{\beta}} + C_{\tilde{\phi}}^{\tilde{\alpha}}d_{\tilde{\alpha}}\right)\mathbf{X}^\top\mathbf{E}_0^{-1}\mathbf{F}_0\mathbf{Z}.\end{aligned}$$

Given the simplicity of the Fisher information matrix for the BS regression model, the SC test statistic is reduced to

$$\mathrm{SC} = \left(d_{\tilde{\beta}}^2 / I_{\tilde{\beta}\tilde{\beta}}\right) \mathbf{X}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} + \left(d_{\tilde{\alpha}}^2 / I_{\tilde{\alpha}\tilde{\alpha}}\right) \mathbf{Z}^\top (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}.$$

3.4.3 | GR test

The GR test statistic is given by

$$\mathrm{GR} = a_{\tilde{\mu}} \left(C_{\tilde{\mu}}^{\tilde{\beta}} d_{\tilde{\beta}} + C_{\tilde{\mu}}^{\tilde{\alpha}} d_{\tilde{\alpha}} \right) \mathrm{X}^\top \left(\hat{\boldsymbol{\xi}} - \widetilde{\mathbf{1}}_0 \right) + a_{\tilde{\phi}} \left(C_{\tilde{\phi}}^{\tilde{\alpha}} d_{\tilde{\alpha}} + C_{\tilde{\phi}}^{\tilde{\alpha}} d_{\tilde{\alpha}} \right) \mathrm{Z}^\top \left(\hat{\boldsymbol{\nu}} - \mathbf{1}_0 \right),$$

where $\mathbf{1}_0^\top = (1, 0, \dots, 0)$. For the BS regression model, the GR test statistic is reduced to $\mathrm{GR} = d_{\tilde{\beta}} \mathrm{X}^\top \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right) + d_{\tilde{\alpha}} \mathrm{Z}^\top \left(\hat{\boldsymbol{\nu}} - \mathbf{1}_0 \right)$.

3.4.4 | WA test

The WA test statistic is given by

$$\begin{aligned}\mathrm{WA} = {} & \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right)^\top \mathbf{E}_0^{-1} \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right) + \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right)^\top \left[\mathbf{G}_0^{-1} + \mathbf{F}_0^\top \mathbf{E}_0^{-1} \mathbf{F}_0 \right] \left(\hat{\boldsymbol{\nu}} - \mathbf{1}_0 \right) \\ & - 2 \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right)^\top \mathbf{E}_0^{-1} \mathbf{F}_0 \left(\hat{\boldsymbol{\nu}} - \mathbf{1}_0 \right).\end{aligned}$$

with the corresponding simplification for the BS regression model expressed as

$$\mathrm{WA} = I_{\tilde{\beta}\tilde{\beta}}^{-1} \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right)^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \left(\hat{\boldsymbol{\xi}} - \mathbf{1}_0 \right) + I_{\tilde{\alpha}\tilde{\alpha}}^{-1} \left(\hat{\boldsymbol{\nu}} - \mathbf{1}_0 \right)^\top \left(\mathbf{Z}^\top \mathbf{Z} \right)^{-1} \left(\hat{\boldsymbol{\nu}} - \mathbf{1}_0 \right).$$

3.5 | Randomized quantile residuals

Balakrishnan and Zhu[10] suggested using the following transformation to perform a residual analysis for the BS regression model:

$$e_i = \frac{1}{\hat{\alpha}_i} \left(\sqrt{\frac{t_i}{\hat{\beta}_i}} - \sqrt{\frac{\hat{\beta}_i}{t_i}} \right), \quad i = 1, \dots, n,$$

where $\hat{\alpha}_i = \varphi^{-1}(\hat{\phi}_i)$ and $\hat{\beta}_i = \hat{\mu}_i / \rho(\varphi^{-1}(\hat{\phi}_i))$. We can check that $e_1, \dots, e_n$ is a random sample from a $N(0, 1)$ model by using the normal quantile-quantile (QQ) plot. The authors did not mention, but e_i also corresponds to the definition of randomized quantile residuals (rQR) presented by Dunn and Smyth.[25] For the BS model, such residuals are defined as

$$e_i = \Phi^{-1} \left(F(y_i; \hat{\alpha}_i, \hat{\beta}_i) \right) = \Phi^{-1} \left(\Phi \left(\frac{1}{\hat{\alpha}_i} \zeta \left(\frac{y}{\hat{\beta}_i} \right) \right) \right) = \frac{1}{\hat{\alpha}_i} \left(\sqrt{\frac{t_i}{\hat{\beta}_i}} - \sqrt{\frac{\hat{\beta}_i}{t_i}} \right), \quad i = 1, \dots, n.$$

Note that for the case where ϕ is not modeled and by Theorem 1 and by the invariance property of the ML estimators, $e_1, \dots, e_n$ are identical for the BS, RBS, RBSQ$_\tau$, and RBSM regression models. In practice and based on our experience with many real data sets tested for this work, the performance of rQR in the four parametrizations is similar.

4 | ILLUSTRATIVE EXAMPLES

In this section, we shall present and discuss two empirical applications to compare the regression models adjustments presented in this article (BS, RBS, RBSQ$_\tau$, and RBSM). We perform the parameter estimation under the maximum likelihood paradigm, as discussed in Section 3. The required numerical evaluations for data analysis were implemented using the R software.[24]

4.1 | Land rent data

To illustrate the BS, RBS, RBSQ$_\tau$, and RBSM regression models, we revisited the land rent data presented in Weisberg[26] and analyzed by the RBS regression model in Santos-Neto et al.[22] The response variable is the ratio between the average rent per acre planted with alfalfa and the corresponding average rent for other agricultural uses (Y), and the density of dairy cows (X, number per square mile) is the explanatory variable. The dataset contains 67 osbervations. Figure 4 shows the plot for these two variables.

Note that it is reasonable to consider the following models:

(M1) $y_i \sim \text{BS}(\mu_i, \phi_i)$;
(M2) $y_i \sim \text{RBS}(\mu_i, \phi_i)$;
(M3) $y_i \sim \text{RBSQ}_{\tau=0.25}(\mu_i, \phi_i)$, that is, we are interested in modeling the 25th quantile;
(M4) $y_i \sim \text{RBSM}(\mu_i, \phi_i)$,

for $i = 1, \ldots, 67$, where $\mu_i = \nu_0 + \nu_1 \cdot X_i$ and $\phi_i = \xi_0 + \xi_1 \cdot X_i$. Table 3 presents the log-likelihood function estimates for the four models with different regression structures for μ and ϕ. As discussed in Remarks 2 and 3, when no covariates are included in ϕ (independently if μ is modeled or not), the BS, RBS, RBSQ$_{\tau=0.25}$, and RBSM models define the same pdf, and consequently the same log-likelihood function estimate. Only when ϕ is modeled we obtain different log-likelihood function estimates for the four models. However, as illustrated in Table 4, ξ_1 is not significant, independently of the parameterization used for the model.

Table 5 shows the estimated parameters, standard errors (S.E.), and log-likelihood function estimates for the considered models. From Table 5, observe that the log-likelihood estimates are equal for the BS, RBS, RBSQ$_{\tau=0.25}$, and RBSM models, as expected. Furthermore, we observe that the estimates of ν_1 are the same for all models. Figure 5 shows the

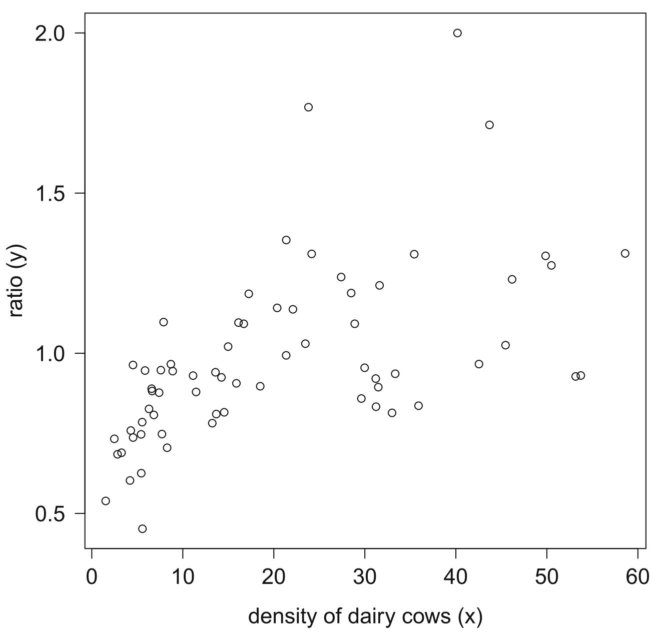

FIGURE 4 Plot for ratio and the density of dairy cows; land rent dataset.

TABLE 3 Log-likelihood function estimates for the BS, RBS, RBSQ$_{\tau=0.25}$, and RBSM models under different regression structures; land rent dataset.

Covariates		**Model**			
μ	ϕ	**BS**	**RBS**	**RBSQ$_{\tau=0.25}$**	**RBSM**
Only intercept	Only intercept	−1.17			
X	Only intercept	13.15			
X	X	14.15	14.35	14.02	13.98

TABLE 4 p-values for the indicated hypothesis tests for the BS, RBS, RBSQ$_{\tau=0.25}$, and RBSM models under different regression structures.

Model	**BS**				**RBS**			
Test	**LR**	**SC**	**GR**	**WA**	**LR**	**SC**	**GR**	**WA**
$H_0 : \xi_1 = 0$	0.158	0.164	0.154	0.166	0.121	0.129	0.117	0.130
$H_0 : \nu_1 = \xi_1 = 0$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Model	**RBSQ$_{\tau=0.25}$**				**RBSM**			
Test	**LR**	**SC**	**GR**	**WA**	**LR**	**SC**	**GR**	**WA**
$H_0 : \xi_1 = 0$	0.187	0.193	0.184	0.195	0.198	0.202	0.195	0.204
$H_0 : \nu_1 = \xi_1 = 0$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

TABLE 5 Parameter estimates and corresponding S.E. for the BS, RBS, RBSQ$_{\tau=0.25}$ and RBSM models modeling μ while leaving ϕ constant; land rent dataset.

	BS		**RBS**		**RBSQ$_{\tau=0.25}$**		**RBSM**	
Parameter	**Estimate**	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**
ν_0	−0.2598	0.0431	−0.2163	0.0437	−0.4019	0.0448	−0.3053	0.0438
ν_1	0.0101	0.0017	0.0101	0.0017	0.0101	0.0017	0.0101	0.0017
ξ_0	−1.5567	0.0864	3.8066	0.1728	−1.5566	0.0864	−3.0680	0.1808
Log-likelihood estimate	13.15							

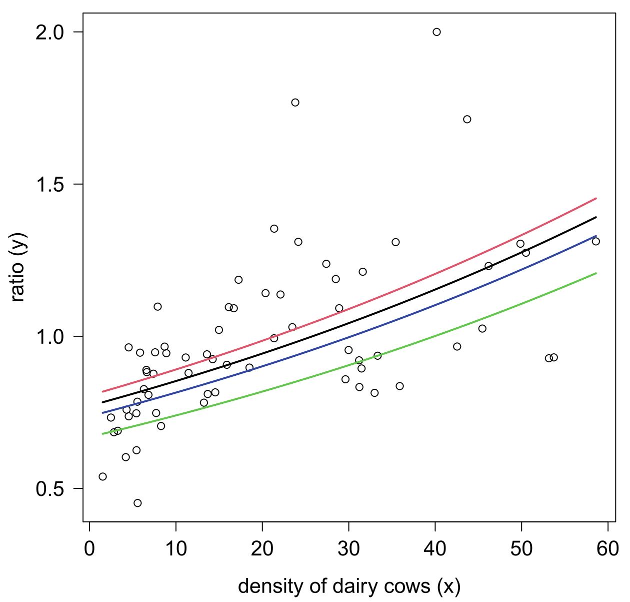

FIGURE 5 Estimated median (black), mean (red), 25th quantile (green), and mode (blue) for ratio against the density of dairy cows. The same estimates are obtained for the BS, RBS, RBSQ$_{\tau=0.25}$, and RBSM models because ϕ is constant.

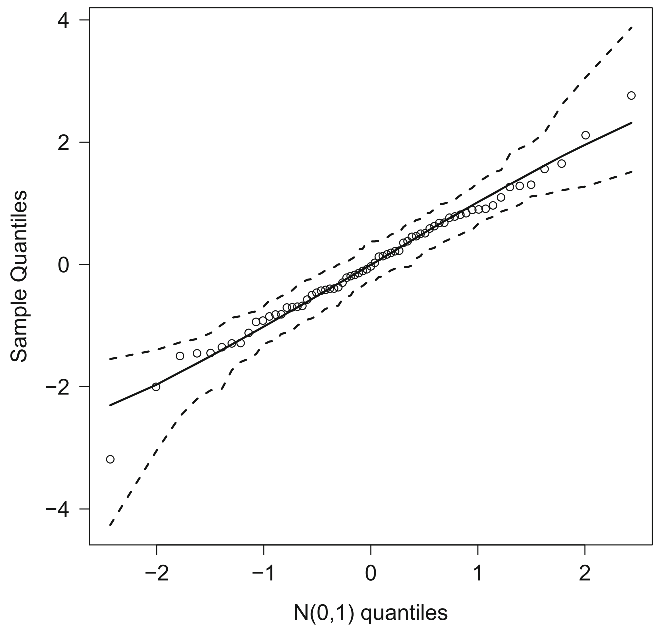

FIGURE 6 QQ plot with envelope for rQR modeling μ while leaving ϕ constant. The same residuals correspond to the BS, RBS, $\text{RBSQ}_{\tau=0.25}$, and RBSM models; land rent dataset.

TABLE 6 Log-likelihood function estimates for the BS, RBS, $\text{RBSQ}_{\tau=0.25}$ and RBSM models under different regression structures; AIS dataset.

Covariates		**Model**			
μ	ϕ	**BS**	**RBS**	$\textbf{RBSQ}_{\tau=0.25}$	**RBSM**
Only intercept	Only intercept	−629.2681			
`ssf+lbm`	Only intercept	−414.7954			
`ssf+lbm`	`ssf+lbm`	−397.4842	−383.9829	−393.3199	−394.9068

estimates for the median, mean, mode, and 25th quantile, which are the same for the four regression models as proved in Theorem 1. Finally, Figure 6 shows the QQ plot for rQR, which are the same for the four regression models when ϕ is constant, as discussed in Section 3.5.

4.2 | Australian Institute of Sport data

In this application, we consider the traditional Australian Institute of Sport (AIS) data presented in Cook and Weisberg.[27] The data correspond to 202 athletes and their health measures, such as body fat percentage (`Bfat`), sum of skin folds (`ssf`), lean body mass (`lbm`), among others. We model the `Bfat` in terms of `ssf` and `lbm`. Similarly to the previous application, we consider the following models:

(M1) $\texttt{Bfat}_i \sim \text{BS}(\mu_i, \phi_i)$;
(M2) $\texttt{Bfat}_i \sim \text{RBS}(\mu_i, \phi_i)$;
(M3) $\texttt{Bfat}_i \sim \text{RBSQ}_\tau(\mu_i, \phi_i)$, with $\tau = 0.25$ (i.e., we are interested in modeling the 25th quantile);
(M4) $\texttt{Bfat}_i \sim \text{RBSM}(\mu_i, \phi_i)$,

for $i = 1, \dots, 202$, where $\mu_i = \nu_0 + \nu_1 \cdot \texttt{ssf}_i + \nu_2 \cdot \texttt{lbm}_i$ and $\phi_i = \xi_0 + \xi_1 \cdot \texttt{ssf}_i + \xi_2 \cdot \texttt{lbm}_i$. Table 6 shows the log-likelihood function estimates for the four models with different regression structures for μ and ϕ. In this case, the

TABLE 7 Parameter estimates and corresponding S.E. for the BS, RBS, $\text{RBSQ}_{\tau=0.25}$, and RBSM models modeling μ and ϕ; AIS dataset.

	BS		RBS		$\text{RBSQ}_{\tau=0.25}$		RBSM	
Parameter	**Estimate**	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**
ν_0	1.9078	0.0671	1.7588	0.0587	1.6141	0.0831	1.7373	0.0746
ν_1	0.0151	0.0006	0.0167	0.0006	0.0145	0.0005	0.0147	0.0005
ν_2	−0.0066	0.0008	−0.0054	0.0007	−0.0030	0.0011	−0.0040	0.0009
ξ_0	−3.1135	0.3728	4.8868	0.7184	−1.3831	0.3957	−2.7553	0.8312
ξ_1	0.0097	0.0017	−0.0247	0.0034	0.0088	0.0016	0.0159	0.0031
ξ_2	0.0076	0.0055	0.0231	0.0097	−0.0185	0.0056	−0.0352	0.0117
Log-likelihood estimate	−397.4842		−383.9829		−393.3199		−394.9068	

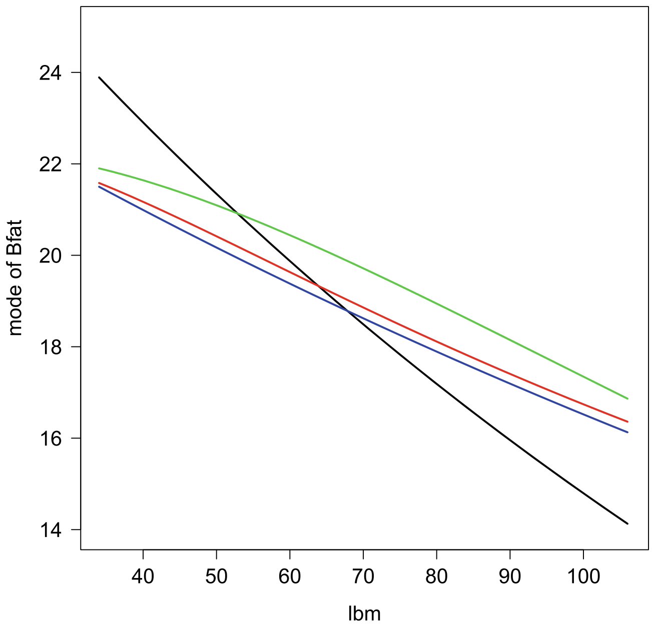

FIGURE 7 Estimated mode for the BS (black), RBS (red), $\text{RBSQ}_{\tau=0.25}$ (green), and RBSM (blue) regression models in terms of `lbm` with `ssf` fixed at 100.

difference between the models with covariates in ϕ and models with ϕ constant is considerable. Moreover, for the tests $H_0 : \nu_1 = \nu_2 = 0$ versus H_1 : otherwise and $H_0 : \xi_1 = \xi_2 = \nu_1 = \nu_2 = 0$ versus H_1 : otherwise, the LR, SC, GR, and WA tests reject the null hypothesis with any usual level of significance.

The estimated parameters for the four models are presented in Table 7. For this particular problem, the RBS provided a higher log-likelihood function estimate. As all models have the same number of parameters, traditional criteria such as AIC or BIC will choose such a model. However, in practice, it depends on which we are interested in modeling (median, mean, quantile, or mode). For instance, suppose that we are particularly interested in the mode of the distribution. The mode of the BS model has not a closed form and need to be computed numerically solving Equation (2). For the BS, RBS and $\text{RBSQ}_{\tau=0.25}$ models, such a mood depends on $\boldsymbol{\nu}$ and $\boldsymbol{\xi}$, whereas for the RBSM model depends only on $\boldsymbol{\nu}$. Of course, in this context this provides the advantage for the RBSM model over the other models, since it allows interpreting the regression coefficients. Thus, for a fixed `lbm` the mode of `Bfat` is increased in 1.48% (because $\exp(\hat{\nu}_1) \approx 1.0148$) when the `ssf` is increased by one unit and for a fixed `ssf` the mode of `Bfat` is decreased in 0.40% ($\exp(\hat{\nu}_2) \approx 0.9960$) when the `lbm` is increased by one unit. This kind of interpretation to the coefficients in relation to the mode of the distribution is not possible for the rest of models. In order to illustrate the differences among the estimation for the mode when ϕ is modeled, Figure 7 presents the estimated mode of `Bfat` for the four regression models in terms of `lbm` with `ssf` fixed at 100. Differently from the previous application, the mode varies in terms of the covariate for each model. We highlight

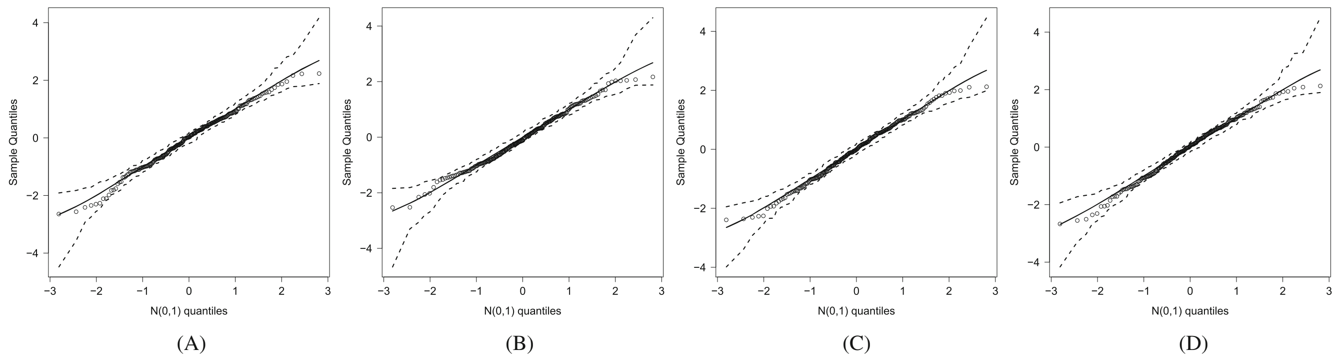

FIGURE 8 QQ plots with envelope for rQR modeling both μ and ϕ; AIS dataset. (A) BS; (B) RBS; (C) RBSQ$_{\tau=0.25}$; (D) RBSM

the great difference among the BS and the other models, especially when `lbm` assumes values at the observed sample extremes.

Finally, Figure 8 presents the QQ plot for rQR of the four models with regression structure in μ and ϕ, suggesting that the four models are appropriate for these data.

5 | CLOSING REMARKS

In this article, we studied a general parameterization of BS regression models in terms of central tendency measures and precision/dispersion parameters. We also established some properties, discussed estimation of the parameters, and conducted hypothesis testing on the model parameters. In the recent literature concerning BS regression models, many papers have assumed a regression structure only in the measures of central tendency parameters but, in these cases, we showed (see Theorem 1) that the models define the same pdf. Additionally, we developed a new parameterized BS distribution in terms of the mode and precision parameters. The reparameterization was based on a simple approximation (in case of moderate skewness) for the mode of the BS distribution. Based on this parameterization, we proposed a new parametric regression model to study the association between an asymmetric positive real response and covariates via inferring the conditional mode of the response. The results from the real data analysis discussed in Section 4 confirmed the potential of using the new approach based on a reparameterized BS modal regression. In general, the results presented by the proposed model were satisfactory and comparable to conventional approaches. However, as discussed in Section 2.4, the RBSM regression model is only appropriated in situations where the response variable is not very skewed. Otherwise, the model is likely to be rejected based on the rQR QQ plot and is expected to provide a much larger AIC.

ACKNOWLEDGMENTS

The authors thank the Editor, Associate Editor and two anonymous reviewers for the constructive comments in a early version of this work. Marcelo Bourguignon gratefully acknowledge partial financial support of the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: Grant 304140/2021-0).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Marcelo Bourguignon https://orcid.org/0000-0002-1182-5193
Diego I. Gallardo https://orcid.org/0000-0001-8184-7403

REFERENCES

1. Birnbaum Z, Saunders S. A new family of life distributions. *J Appl Probab*. 1969a;6:319-327.
2. Birnbaum Z, Saunders S. Estimation for a family of life distributions with applications to fatigue. *J Appl Probab*. 1969b;6:328-347.

3. Saulo H, Leão J, Leiva V, Aykroyd RG. Birnbaum-Saunders autoregressive conditional duration models applied to high-frequency financial data. *Stat Pap*. 2019;60:1605-1629.
4. Marchant C, Leiva V, Christakos G, Cavieres MF. Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile. *Environmetrics*. 2019;30(5):e2551.
5. Leão J, Leiva V, Saulo H, Tomazella V. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Stat Med*. 2018;37(29):4421-4440.
6. Johnson N, Kotz S, Balakrishnan N. *Continuous Univariate Distributions*. Vol 2. Wiley; 1995.
7. Leiva V. *The Birnbaum-Saunders Distribution*. Academic Press; 2016.
8. Balakrishnan N, Kundu D. Birnbaum-Saunders distribution: a review of models, analysis, and applications. *Appl Stoch Model Bus Ind*. 2019;35(1):4-49.
9. Dasilva A, Dias R, Leiva V, Marchant C, Saulo H. [invited tutorial] Birnbaum-Saunders regression models: a comparative evaluation of three approaches. *J Stat Comput Simul*. 2020;90(14):2552-2570.
10. Balakrishnan N, Zhu X. Inference for the Birnbaum-Saunders lifetime regression model with applications. *Commun Stat Simul Comput*. 2015;44(8):2073-2100.
11. Rieck J, Nedelman J. A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*. 1991;3:51-60.
12. Leiva V, Santos-Neto M, Cysneiros F, Barros M. Birnbaum-Saunders regression model: a new approach. *Stat Model*. 2014;14:21-48.
13. Sánchez L, Leiva V, Galea M, Saulo H. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. *Appl Stoch Models Bus Ind*. 2021;37(1):53-73.
14. Bourguignon M, Leão J, Gallardo DI. Parametric modal regression with varying precision. *Biom J*. 2020;62(1):202-220.
15. Nichols MD, Padgett WJ. A bootstrap control chart for Weibull percentiles. *Qual Reliab Eng Int*. 2006;22:141-151.
16. Chen Y-C, Genovese CR, Tibshirani RJ, Wasserman L. Nonparametric modal regression. *Ann Stat*. 2016;44(2):489-514.
17. Saunders SC. A family of random variables closed under reciprocation. *J Am Stat Assoc*. 1974;69:533-539.
18. Kundu D, Kannan N, Balakrishnan N. On the hazard function of Birnbaum-Saunders distribution and associated inference. *Comput Stat Data Anal*. 2008;52:2692-2702.
19. Gradshteyn I, Randzhik I. *Table of Integrals, Series, and Products*. Academic Press; 2000.
20. Santos-Neto M, Cysneiros F, Leiva V, Ahmed S. On new parameterizations of the Birnbaum-Saunders distribution. *Pak J Stat*. 2012;28:1-26.
21. Santos-Neto M, Cysneiros F, Leiva V, Barros M. On new parameterizations of the Birnbaum-Saunders distribution and its moments, estimation and application. *Revstat Stat J*. 2014;12:247-272.
22. Santos-Neto M, Cysneiros F, Leiva V, Barros M. Reparameterized Birnbaum-Saunders regression models with varying precision. *Electron J Stat*. 2016;10:2825-2855.
23. Lemonte A. A note on the Fisher information matrix of the Birnbaum-Saunders distribution. *J Stat Theory Appl*. 2016;15:196-205.
24. R Core Team. *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020 https://www.R-project.org/
25. Dunn P, Smyth G. Randomized quantile residuals. *J Comput Graph Stat*. 1996;5:236-244.
26. Weisberg S. *Applied Linear Regression*. Wiley; 2005.
27. Cook R, Weisberg S. *An Introduction to Regression Graphics*. Vol 1. Wiley; 1994.

How to cite this article: Bourguignon M, Gallardo DI. A new look at the Birnbaum–Saunders regression model. *Appl Stochastic Models Bus Ind*. 2022;38(6):935-951. doi: 10.1002/asmb.2687