**Revista Colombiana de Estadística**
January 2017, Volume 40, Issue 1, pp. 65 to 83
DOI: http://dx.doi.org/10.15446/rce.v40n1.51738

# Bimodal Regression Model

## Modelo de regresión Bimodal

Guillermo Martínez-Flórez[1,a], Hugo S. Salinas[2,b], Heleno Bolfarine[3,c]

[1]Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Córdoba, Colombia

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

[3]Departamento de Estatística, IME, Universidad de Sao Paulo, Sao Paulo, Brasil

**Abstract**

Regression analysis is a technique widely used in different areas of human knowledge, with distinct distributions for the error term. It is the case, however, that regression models with the error term following a bimodal distribution are not common in the literature, perhaps due to the lack of simple to deal with bimodal error distributions. In this paper, we propose a simple to deal with bimodal regression model with a symmetric-asymmetric distribution for the error term for which for some values of the shape parameter it can be bimodal. This new distribution contains the normal and skew-normal as special cases. A real data application reveals that the new model can be extremely useful in such situations.

***Key words*:** Bimodal Distribution, Generalized Gaussian Distribution, Linear Regression, Power Regression Model.

**Resumen**

El análisis de regresión es una técnica muy utilizada en diferentes áreas de conocimiento humano, con diferentes distribuciones para el término de error, sin embargo los modelos de regresión con el termino de error siguiendo una distribución bimodal no son comunes en la literatura, tal vez por la simple razón de no tratar con errores con distribución bimodal. En este trabajo proponemos un camino sencillo para hacer frente a modelos de regresión bimodal con una distribución simétrica - asimétrica para el término de error para la cual para algunos valores del parámetro de forma esta puede ser

[a]PhD. E-mail: gmartinez@correo.unicordoba.edu.co
[b]PhD. E-mail: hugo.salinas@uda.cl
[c]PhD. E-mail: hbolfar@ime.usp.br

bimodal. Esta nueva distribución contiene a la distribución normal y la distribución normal asimétrica como casos especiales. Una aplicación con datos reales muestra que el nuevo modelo puede ser extremadamente útil en algunas situaciones.

***Palabras clave*:** distribución bimodal, distribución gaussiana generalizada, regresión lineal, modelo de regresión exponenciado.

# 1. Introduction

To study the relationship between variables in different areas of human knowledge, linear and nonlinear regression models have been substantially used. It is typically considered that the error term follows a normal distribution although more general symmetric error distributions have also been considered. One of those alternatives is to consider that the errors follow distributions with heavier tails than those of normal distribution, in order to reduce the influence of outlying observations. In this context, Lange, Little and Taylor (1989) proposed the Student-$t$ model with unknown degrees of freedom for parameter $\nu$. Cordeiro, Ferrari, Uribe-Opazo and Vasconcellos (2000), and Galea, Paula and Cysneiros (2005) present results from the study of inferential aspects of symmetrical nonlinear models. For the asymmetric nonlinear model, we use the work of Cancho, Lachos and Ortega (2010). Symmetrical measurement error models have been investigated in Arellano-Valle, Bolfarine and Vilca-Labra (1996).

One situation in which we encounter an anomaly in the error term of the model occurs when it is of interest to explain the fat percentage in the human body as a function of the individual weight. It is the case, however, that given inherent gender peculiarities, the exclusion of the gender variable can lead to a bimodal error distribution model. That is, not taking into account the sex variable , leads to a regression model for which the distribution of the error term is no longer unimodal.

A viable alternative to this situation is to use a mixture of normal distributions for the error term. According to this alternative, there are two models to estimate, one for each component of the normal mixture, namely ($\epsilon_j$ for $j = 1, 2$); they are both normally distributed with mean zeros and variance $\sigma_j$ for $j = 1, 2$. According to De Veaux (1989), for the special case of two explanatory variables $X_1$ and $X_2$, the response variable can be written as

$$y_i = \begin{cases} \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \epsilon_{1i}, & \text{with probability } p, \\ \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \epsilon_{2i}, & \text{with probability } 1 - p \end{cases}$$

where the $\epsilon_{ji} \sim N(0, \sigma_j^2)$ are independent, $j = 1, 2$, $i = 1, 2, \cdots, n$. Consequently, response $y_i$ has a pdf

$$f(y_i) = \frac{p}{\sigma_1}\phi\left(\frac{y_i - \beta_{10} - \beta_{11}x_{1i} - \beta_{12}x_{2i}}{\sigma_1}\right) + \frac{1-p}{\sigma_2}\phi\left(\frac{y_i - \beta_{20} - \beta_{21}x_i - \beta_{22}x_{2i}}{\sigma_2}\right),$$

for $i = 1, 2, \dots, n$. Although widely recommended, this alternative has some important drawbacks. The first results from lack of identifiability of some model parameters, more specifically, $\beta_{10}$ and $\beta_{20}$. Another difficulty is related to convergence problems with the algorithm for parameter estimation, including the proportion of data points for each model. Moreover, the model is not parsimonious at all, since by increasing the number of explanatory variables, the number of parameters in the model jumps to nine. For instance, for one explanatory variable there are seven parameters to be estimated, making algorithm convergence difficult. De Veaux (1989) presents an EM-algorithm for the mixture of regression models. Further results can be found in Quandt (1958), Turner (2000), and Young & Hunter (2010), among others.

In this paper, we suggest using the symmetric-asymmetric bimodal alpha-power model, considered in Bolfarine, Martínez and Salinas (2012), to adjust data with a linear relation. Results from two real data applications are reported the illustrate the usefulness of the models developed. One alternative, clearly, is to undertake data transformation or use mixtures of distributions, as mentioned above.

The paper is organized as follows. Section 2 is devoted to describing the bimodal symmetric-asymmetric alpha-power distribution and some of its main properties. The model considered generalizes both the skew-normal model (Azzalini 1985) and the power-normal model (Pewsey, Gómez and Bolfarine 2012). The extension of the normal multiple regression model to the case in which the error term follows the bimodal symmetric-asymmetric power-normal (ABPN) model is considered in Section 3. Maximum likelihood estimation is discussed in Section 4. In particular, it is shown that the Fisher information matrix is nonsingular and allows for normality to be tested using the likelihood ratio statistics. A real application considered in Section 5 illustrates the fact that the model considered can outperform traditional symmetric models that have been previously considered in the literature, in specifically the mixture of normals.

## 2. The Bimodal Symmetric-Asymmetric Alpha-Power Distribution

The alpha-power distribution was first considered in Durrans (1992), and its pdf is given by

$$g(z; \alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}, \qquad z \in \mathbb{R}, \tag{1}$$

where $\alpha \in \mathbb{R}^+$ is a shape parameter, and $\Phi$ and $\phi$ are the density and distribution functions of the standard normal, respectively. We use the notation $Z \sim PN(\alpha)$. The location-scale extension of $Z$, $Y = \mu + \sigma Z$, where $\xi \in R$ and $\sigma \in R^+$, have a probability density function given by

$$\varphi(y; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\alpha-1}. \tag{2}$$

We use the notation $Y \sim PN(\mu, \sigma, \alpha)$.

Several authors, for example, Gupta and Gupta (2008), Pewsey, Gómez & Bolfarine (2012), Rego, Cintra & Cordeiro (2012), studied properties of this model, but the fact that this model can be seen as an aditive generalized model seems to be unknown. This model can be further extended by considering $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$ replacing $\mu$, where $\boldsymbol{\beta}$ is an unknown vector of regression coefficients and $\mathbf{x}_i$ a vector of known regressors, possibly correlated with the response vector.

Martínez-Flórez, Bolfarine & Gómez (2015) considered the multiple regression model represented by

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{3}$$

where $\boldsymbol{\beta}$ is a vector of unknown constants, $\mathbf{x}_i$ are values of known explanatory variables, and the error terms $\epsilon_i$ are independent random variables with power-normal distribution, $PN(0, \sigma, \alpha)$. This model becomes a viable alternative to the ordinary regression models under normality for the situation of asymmetrically distributed errors with kurtosis above 3 (normal distribution). These authors studied the main properties of this model, obtained equations to estimate model parameters via maximum likelihood, and deduced its information matrices. They found that the Fisher information matrix is nonsingular. Although the new proposal is a viable alternative to model data with low and high asymmetry, this model can only be applied to unimodal situations.

As an extension of the PN model to bimodal data, Bolfarine, Martínez-Flórez & Salinas (2012) introduced the family of bimodal distributions, one symmetric and the other asymmetric. The corresponding density function of the bimodal power-normal distribution is given by

$$\varphi(y; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma}\frac{2^{\alpha-1}}{2^{\alpha}-1}\phi\left(\frac{y-\mu}{\sigma}\right)\left\{\Phi\left(\left|\frac{y-\mu}{\sigma}\right|\right)\right\}^{\alpha-1}, \quad x \in \mathbb{R}, \tag{4}$$

where $\mu$ is the location parameter and $\sigma$ is the scale parameter. We use the notation $BPN(\mu, \sigma, \alpha)$. Note that for $\alpha = 1$ the normal distribution $N(\mu, \sigma^2)$ follows.

The $r$-th moment of the random variable $Y \sim BPN(0, 1, \alpha)$ is given by

$$\mathbb{E}(Z^r) = \begin{cases} 0, & \text{if } r \text{ is odd}, \\ 2\mu_r(0), & \text{if } r \text{ is even}. \end{cases}$$

where

$$\mu_r(0) = \alpha c_\alpha \int_0^\infty z^r \phi(z) \left\{\Phi(z)\right\}^{\alpha-1} dz, \qquad r = 0, 1, 2, \ldots \tag{5}$$

Hence, it follows that $\mathbb{E}(Z) = \mathbb{E}(Z^3) = 0$. The authors show that the pdf is symmetric and, moreover, if $\alpha > 1$, then its density function is bimodal. Furthermore, maximum likelihood estimation is considered for model parameters and the Fisher information matrix is derived and shown to be nonsingular. Under these conditions, and given that it is a regular continuous function, it also follows the $\sqrt{n}$-normal approximation for the maximum likelihood estimators for the parameter vector.

Bolfarine et al. (2012) also studied a distribution for fitting symmetric-asymmetric data with bimodal behaviour and this distribution was termed the *bimodal symmetric-asymmetric power-normal model*.

The density function for the location-scale version of the model can be written as

$$\varphi(y;\mu,\sigma,\alpha,\lambda)=\frac{2\alpha c_\alpha}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\left\{\Phi\left(\left|\frac{y-\mu}{\sigma}\right|\right)\right\}^{\alpha-1}\Phi\left(\lambda\frac{y-\mu}{\sigma}\right),$$

where $y\in\mathbb{R}$, $\mu\in\mathbb{R}$ is a location parameter, $\sigma>0$ is a scale parameter, $\alpha\in\mathbb{R}^+$ is a shape parameter, $\lambda\in\mathbb{R}$ is an asymmetry parameter, and $c_\alpha=2^{\alpha-1}/(2^\alpha-1)$ is the normalizing constant. We use the notation $Y\sim ABPN(\mu,\sigma,\alpha,\lambda)$.

The authors show that for $\alpha>1$ and $\lambda$ satisfying $\left[1-\frac{\lambda}{z}\frac{\phi(\lambda z)}{\Phi(\lambda z)}\right]>0$, this model is bimodal asymmetric; whereas for $\alpha>1$ and $\lambda=0$, it is bimodal symmetric. Conversely, for $\alpha\leq 1$ the resulting model is unimodal. We note that for $\alpha=1$ the skew-normal model follows, for $\alpha=1$ and $\lambda=0$ the normal case follows, and for $\lambda=0$ the bimodal power-normal model follows.

The $r$-th moment of a random variable $Z\sim ABPN(0,1,\alpha,\lambda)$ is given by

$$\mathbb{E}(Z^r)=\begin{cases}2\mu_r(0), & \text{if } r \text{ is even},\\ 2\mu_r(0)+2\mu_r(\beta,\alpha), & \text{if } r \text{ is odd},\end{cases}$$

where

$$\mu_r(\lambda,\alpha)=2\alpha c_\alpha\int_0^\infty z^r\phi(z)\left\{\Phi(z)\right\}^{\alpha-1}\Phi(\lambda z)dz.$$

In addition to these results, these authors have shown that the information matrix is nonsingular at the vicinity of symmetry, that is, $\alpha=1$ and $\lambda=0$. This leads to large sample normal distribution for the maximum likelihood estimators for which the asymptotic covariance matrix is the inverse of the Fisher information matrix.

# 3. The Multiple Regression Model With ABPN Errors

We assume, under the ordinary multiple regression model, that the error term follows a ABPN distribution with parameters $\mu=0,\sigma,\alpha$ and $\lambda$, that is, for $i=1,2,\dots,n$, the $\varepsilon_i$ are independent random variables with $\varepsilon_i\sim ABPN(0,\sigma,\alpha,\lambda)$. Hence, it follows that the density function of $\varepsilon_i$ is given by

$$\varphi(\varepsilon_i;0,\sigma,\alpha,\lambda)=\frac{2\alpha c_\alpha}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)\left\{\Phi\left(\left|\frac{\varepsilon_i}{\sigma}\right|\right)\right\}^{\alpha-1}\Phi\left(\lambda\frac{\varepsilon_i}{\sigma}\right),$$

for $i = 1, \ldots, n$. Therefore, it follows that $y_i$ given $x_i$ (denoted $y_i \mid x_i$) also follows a ABPN distribution, that is, $y_i \mid x_i \sim ABPN(x_i'\boldsymbol{\beta}, \sigma, \alpha, \lambda)$, for $i = 1, 2, \ldots, n$. In this model, $x_i'\boldsymbol{\beta}$ is a location parameter, $\sigma$ is a scale parameter, $\alpha$ is a shape parameter, and $\lambda$ is an asymmetry parameter where $\boldsymbol{\beta}$ is a vector of unknown constants and $x_i$ are values of known explanatory variables.

The interpretation of the systematic part of the model, namely $(\beta_0, \beta_1, \cdots, \beta_p)$, is similar to that of the model under the ordinary normal assumption, and $\sigma$ is a scale parameter related to the error terms.

Under the ABPN model, $E(y_i) \neq x_i'\boldsymbol{\beta}$, and we have to make the following correction to obtain the regression line as the expected value of the response variable $\beta_0^* = \beta_0 + \mu_\varepsilon$, where $\mu_\varepsilon = \mathbb{E}(\varepsilon_i)$. Thus, $E(y_i) = x_i'\boldsymbol{\beta}^*$ where $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1, \ldots, \beta_p)'$.

As special cases this model contains the model with normal errors, that is, $\lambda = 0$ and $\alpha = 1$, as well as the model with skew-normal errors for $\alpha = 1$ and the bimodal symmetric error model for $\lambda = 0$.

# 4. Inference for the Multiple Linear ABPN Model

## 4.1. Likelihood and Score Functions

Considering a matrix notation where $Y$ denotes a $(n \times 1)$-dimensional vector with entries $y_i$ and $\mathbf{X}$ the $(n \times (p+1))$-matrix with rows $x_i'$, the likelihood function for $\boldsymbol{\theta = (\beta', \sigma, \alpha, \lambda)'}$, given a random sample of size $n$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$, can be written as

$$\begin{aligned}\ell(\boldsymbol{\beta}; \mathbf{Y}) &= n[\ln(2\alpha) + \ln(c_\alpha) - \ln(\sigma)] \\ &- \frac{1}{2\sigma^2}(Y - \mathbf{X}\boldsymbol{\beta})'(Y - \mathbf{X}\boldsymbol{\beta}) + \mathbf{1}'\left[(\alpha - 1)\mathbf{U_1} + \mathbf{U_2}\right],\end{aligned}$$

where $\mathbf{1}'$ is a $n$-dimensional vector, $\mathbf{U_1}$ and $\mathbf{U_2}$ are $n$-dimensional vectors with elements $\ln\left\{\Phi\left|\frac{y_i - x_i'\boldsymbol{\beta}}{\sigma}\right|\right\}$ and $\ln\left\{\Phi\left(\lambda\frac{y_i - x_i'\boldsymbol{\beta}}{\sigma}\right)\right\}$, respectively, for $i = 1, 2, \ldots, n$. The score function, $U = (U(\boldsymbol{\beta}), U(\sigma), U(\alpha), U(\lambda))$, has elements that are given by

$$\begin{aligned}U(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta}; \mathbf{Y})}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma}\mathbf{X}'\left[\mathbf{Z} - (\alpha - 1)\mathbf{S}\Lambda_{1\alpha} - \lambda\Lambda_{1\lambda}\right], \\ U(\alpha) = \frac{\partial \ell(\boldsymbol{\beta}; \mathbf{Y})}{\partial \alpha} &= n\left(\frac{1}{\alpha} + \overline{U}_1\right) + n\ln(2)(1 - (1 - 2^{-\alpha})^{-1}), \\ U(\sigma) = \frac{\partial \ell(\boldsymbol{\beta}; \mathbf{Y})}{\partial \sigma} &= \frac{1}{\sigma}\left[-n + \mathbf{Z}'\mathbf{Z} - (\alpha - 1)\left|\mathbf{Z}\right|'\Lambda_{1\alpha} - \lambda\mathbf{Z}'\Lambda_{1\lambda}\right], \\ U(\lambda) = \frac{\partial \ell(\boldsymbol{\beta}; \mathbf{Y})}{\partial \lambda} &= \mathbf{Z}'\Lambda_{1\lambda},\end{aligned}$$

where $\mathbf{S} = \operatorname{diag}\{\operatorname{sgn}(z_1), \ldots, \operatorname{sgn}(z_n)\}$,

$$\mathbf{Z^{k'}} = (z_1^k, \ldots, z_n^k), \qquad \left|\mathbf{Z}^k\right|' = \left(\left|z_1^k\right|, \ldots, \left|z_n^k\right|\right),$$

$\Lambda_{1\alpha} = \left(\frac{\phi(|z_1|)}{\Phi(|z_1|)}, \ldots, \frac{\phi(|z_n|)}{\Phi(|z_n|)}\right)'$, $\Lambda_{1\lambda} = \left(\frac{\phi(\lambda z_1)}{\Phi(\lambda z_1)}, \ldots, \frac{\phi(\lambda z_n)}{\Phi(\lambda z_n)}\right)'$ and $\overline{U}_1 = \frac{1}{n}\sum_{i=1}^{n} U_{1i}$ with $z_i = \frac{y_i - x_i'\boldsymbol{\beta}}{\sigma}$ for $i = 1, \ldots, n$.

After some algebraic manipulations, maximum likelihood estimating equations are given by

$$\boldsymbol{\beta} = \boldsymbol{\beta}_{MQ} + \sigma(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[(\alpha-1)\mathbf{S}\Lambda_{1\alpha} + \lambda\Lambda_{1\lambda}\right], \quad \alpha = -\frac{1}{\overline{U}_1}$$

$$\frac{1}{n}\mathbf{Z}'\mathbf{Z} = 1 + \frac{\alpha-1}{n}\left|\mathbf{Z}\right|'\Lambda_{1\alpha} + \frac{\lambda}{n}\mathbf{Z}'\Lambda_{1\lambda} \quad \text{and} \quad \mathbf{Z}'\Lambda_{1\lambda} = 0$$

where $\boldsymbol{\beta}_{MQ} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$. Hence the MLE of parameter vector $\boldsymbol{\beta}$ is equal to the least squares estimator for $\boldsymbol{\beta}$, plus the asymmetry and bimodal correcting terms. Non analytical solutions are available for the likelihood (score) equations, and, hence, they have to be solved numerically using iterative procedures such as the Newton-Raphson or quase-Newton type algorithms.

Hence, the maximum likelihood estimator for $\theta$ can be obtained by implementing the following iterative procedure:

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + [J(\hat{\theta}^{(k)})]^{-1}U(\hat{\theta}^{(k)}), \tag{6}$$

where $J(\theta) = -\dfrac{\partial^2\ell(\theta)}{\partial\theta\,\partial\theta^\top}$ is the observed information matrix. There are however, other numerical procedures based on the expected (Fisher) information matrix.

These optimization algorithms can be found in the following packages: `nlm`, `optim`, `maxLik` or `optimx` of the R software (R Development Core Team. (2015)). These are procedures that are based on the function score for parameter estimation.

To initialize the estimation process, the following algorithm is considered. Firstly, the ordinary normal linear regression model is fitted and model errors are estimated. Using these estimates, the ABPN model is fitted, from which estimates for $\lambda$ and $\alpha$ are computed. Then, $\mu_\varepsilon$ can be estimated. For $\varepsilon_i^* = \varepsilon_i - \mu_\varepsilon$, we have $\mathbb{E}(\varepsilon^*) = 0$ and $Var(\varepsilon^*) = \sigma^2\Phi_2(\alpha, \lambda)$, where $\Phi_2$ is the variance of the random variable $ABPN(0, 1, \alpha, \lambda)$.

Hence, the errors sum of squares are minimized, namely,

$$\sum_{i=1}^{n}\varepsilon_i^{*2} = \sum_{i=1}^{n}\left(y_i - x_i'\boldsymbol{\beta}^*\right)^2.$$

We the obtain the least squares estimators of $\boldsymbol{\beta}^*$ and $\sigma$, which are given by:

$$\widehat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \text{and} \quad \widehat{\sigma}^2 = \frac{\Phi_2^{-1}(\widehat{\alpha}, \widehat{\lambda})}{n-2}\sum_{i=1}^{n}\left(y_i - x_i'\widehat{\boldsymbol{\beta}}^*\right)^2.$$

Moreover, $Var(\widehat{\boldsymbol{\beta}}^*) = \sigma^2\Phi_2(\alpha, \lambda)(\mathbf{X}'\mathbf{X})^{-1}$.

## 4.2. The Observed and Expected Information Matrices

Before computing the information matrix (Fisher), we present the elements of the observed information matrix which, after extensive algebraic manipulations, are given by

$$
\begin{aligned}
j_{\boldsymbol{\beta}'\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\mathbf{X}'\left[\mathbf{I}_n + (\alpha-1)\Lambda_{2\alpha} + \lambda^2\Lambda_{2\lambda}\right]\mathbf{X},\\
j_{\sigma\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\mathbf{X}'\left[2\mathbf{Z} + (\alpha-1)\Lambda_{3\alpha} + \lambda\Lambda_{3\lambda}\right],\\
j_{\sigma\sigma} &= \frac{1}{\sigma^2}\left[-n + 3\mathbf{Z}'\mathbf{Z} + (\alpha-1)\left[\left(-2\left|\mathbf{Z}\right|' + \left|\mathbf{Z^3}\right|'\right)\Lambda_{1\alpha} + \mathbf{Z}'\Lambda_{1\alpha}\Lambda_{1\alpha}'\mathbf{Z}\right]\right]\\
&+ \frac{\lambda}{\sigma^2}\left[\left(-2\mathbf{Z}' + \lambda^2\mathbf{Z^3}'\right)\Lambda_{1\lambda} + \lambda\mathbf{Z}'\Lambda_{1\lambda}\Lambda_{1\lambda}'\mathbf{Z}\right],\\
j_{\alpha\alpha} &= n[\alpha^{-2} - 2^{\alpha}(2^{\alpha}-1)^{-2}\ln^2(2)],\\
j_{\lambda\sigma} &= \frac{1}{\sigma}\left[\mathbf{Z}'\Lambda_{1\lambda} - \lambda^2\mathbf{Z^3}'\Lambda_{1\lambda} - \lambda\mathbf{Z}'\Lambda_{1\lambda}\Lambda_{1\lambda}'\mathbf{Z}\right],\\
j_{\lambda\lambda} &= \lambda\mathbf{Z^3}'\Lambda_{1\lambda} + \mathbf{Z}'\Lambda_{1\lambda}\Lambda_{1\lambda}'\mathbf{Z},\\
j_{\alpha\boldsymbol{\beta}} &= \frac{1}{\sigma}\mathbf{X}'\mathbf{S}\Lambda_{1\alpha},\quad j_{\alpha\sigma} = \frac{1}{\sigma}\left|\mathbf{Z}\right|'\Lambda_{1\alpha},\quad j_{\lambda\boldsymbol{\beta}} = -\frac{1}{\sigma}\mathbf{X}'\Lambda_{3\lambda}\ \text{ and }\ j_{\alpha\lambda} = 0.
\end{aligned}
$$

where

$$\Lambda_{2\alpha} = \mathrm{diag}\left\{\left(\frac{\phi(|z_1|)}{\Phi(|z_1|)}\right)^2 + |z_i|\frac{\phi(|z_i|)}{\Phi(|z_i|)}\right\},\ \Lambda_{2\lambda} = \mathrm{diag}\left\{\lambda z_i\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} + \left(\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}\right)^2\right\},$$

$\Lambda_{3\alpha} = (a_1, a_2, \dots, a_n)'$, with

$$a_i = \left\{\mathrm{sgn}(z_i)z_i^2\frac{\phi(|z_i|)}{\Phi(|z_i|)} + z_i\left(\frac{\phi(|z_i|)}{\Phi(|z_i|)}\right)^2 - \mathrm{sgn}(z_i)\frac{\phi(|z_i|)}{\Phi(|z_i|)}\right\},$$

$\Lambda_{3\lambda} = (b_1, b_2, \dots, b_n)'$, with

$$b_i = \left\{\lambda^2 z_i^2\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} + \lambda z_i\left(\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}\right)^2 - \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}\right\},$$

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

The above expressions can be computed numerically. The observed information matrix is obtained after replacing unknown parameters with the corresponding maximum likelihood estimators. The expected information matrix then follows by taking expectations of the above components (multiplied by $n^{-1}$).

Considering: as in Bolfarine et al. (2012):

$$a_{kj} = \mathbb{E}\{z^k(\phi(z)/\Phi(|z|))^j\},\quad a_{kj}^{*} = \mathbb{E}\{|z|^k(\phi(z)/\Phi(|z|))^j\},$$

$$a_{kj}^{**} = \mathbb{E}\{\mathrm{sgn}(z)z^k(\phi(z)/\Phi(|z|))^j\},\quad a_{1kj} = \mathbb{E}\{z^k(\phi(\lambda z)/\Phi(\lambda z))^j\},$$

the elements of the expected information matrix are given by

$$i_{\boldsymbol{\beta}'\boldsymbol{\beta}} = \frac{1+(\alpha-1)(a_{02}-a_{11}^{**})+\lambda^2(\beta a_{111}+a_{102})}{\sigma^2}\mathbf{X}'\mathbf{X}, \quad i_{\lambda\lambda} = \lambda a_{131} + a_{122}$$

$$i_{\sigma\boldsymbol{\beta}} = \frac{2a_{10}+(\alpha-1)\left(a_{01}^{**}-a_{21}^{**}-a_{12}\right)+\lambda(\lambda^2 a_{121}+\lambda a_{112}-a_{101})}{\sigma^2}\mathbf{X}'\mathbf{1},$$

$$i_{\lambda\boldsymbol{\beta}} = \frac{a_{101}-\lambda(\lambda a_{121}+a_{112})}{\sigma}\mathbf{X}'\mathbf{1}, \quad i_{\lambda\sigma} = \frac{a_{111}}{\sigma} - \frac{\lambda(\lambda a_{131}+\lambda a_{122})}{\sigma^2}$$

$$i_{\sigma\sigma} = -\frac{1}{\sigma^2} + \frac{3}{\sigma^2}a_{20} + \frac{\alpha-1}{\sigma^2}\left[a_{31}^{*}+a_{22}-2a_{11}^{*}\right] + \frac{\lambda}{\sigma^2}[\lambda^2 a_{131}+\lambda a_{122}-2a_{111}]$$

$$i_{\alpha\boldsymbol{\beta}} = -\frac{1}{\sigma}a_{01}^{**}\mathbf{X}'\mathbf{1}, \quad i_{\alpha\sigma} = \frac{1}{\sigma}a_{11}^{*}, \quad i_{\alpha\lambda} = 0, \quad i_{\alpha\alpha} = \alpha^{-2} - 2^{\alpha}(2^{\alpha}-1)^{-2}(\ln 2)^2.$$

## 4.3. Bimodal Symmetric Case

The bimodal symmetric regression model is the one where the errors follow the probability density function and is given by

$$\varphi(\varepsilon_i;0,\sigma,\alpha,0) = \frac{\alpha c_\alpha}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)\left\{\Phi\left(\left|\frac{\varepsilon_i}{\sigma}\right|\right)\right\}^{\alpha-1},$$

for $i = 1, 2, \ldots, n$. We use the notation $\epsilon_i \sim BPN(0,\sigma,\alpha)$. Bolfarine et al. (2012) demonstrate that this density is symmetric bimodal for $\alpha > 1$ and unimodal otherwise. We note that this model is a particular case of the ABPN model and take the value $\lambda = 0$. Therefore, the score functions and information matrix for the parameter vector $\theta_1 = (\boldsymbol{\beta}', \sigma, \alpha)'$ can be obtained from the previously obtained ones for the asymmetric regression model, which is followed by making $\lambda = 0$ in the first derivatives with respect to the parameter $\boldsymbol{\beta}$, $\sigma$ and $\alpha$ and similarly for the second derivatives with respect to the vector $\theta_1$.

For $\alpha = 1$, we have $\varphi(\varepsilon_i;0,\sigma,1) = \phi\left(\varepsilon_i/\sigma\right)/\sigma$, the density function of the location-scale normal density, the information matrix is reduced to

$$I(\theta_1) = \begin{pmatrix} \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} & \mathbf{0}_{p+1} & \mathbf{0}_{p+1} \\ \mathbf{0}'_{p+1} & \frac{2}{\sigma^2} & \frac{0.2063}{\sigma} \\ \mathbf{0}'_{p+1} & \frac{0.2063}{\sigma} & 1-2(\ln 2)^2 \end{pmatrix}$$

with determinant $|I(\theta_1)| > 0$, so that the information matrix is nonsingular for the special case of the symmetric normal distribution.

The upper $2 \times 2$ submatrix of $I(\theta_1)^{-1}$ corresponds to the information matrix for the normal distribution. In the next section, we discuss consistency and asymptotic normality for the maximum likelihood estimators. As is well-known, the asymptotic variance is the inverse of the Fisher information above.

### 4.4. Large Sample Distribution of the MLE for the ABPN Model

As mentioned above, the information matrix for the parameter vector

$$\theta = (\boldsymbol{\beta}', \sigma, \alpha, \lambda')' = (\theta_1', \lambda)'$$

for the bimodal regression model is obtained by finding the expectations for the observed information matrix. These expectations are not available in closed form and have to be obtained numerically.

In the particular case where $\alpha = 1$, $\lambda = 0$ so that $\varphi(\varepsilon_i; 0, \sigma, 1, 0) = \phi\left(\varepsilon_i/\sigma\right)/\sigma$, the location-scale normal density function, the information matrix becomes

$$I(\theta) = \begin{pmatrix} I(\theta_1) & I(\theta_1, \lambda) \\ I'(\theta_1, \lambda) & \frac{2}{\pi} \end{pmatrix}.$$

The determinant is given by $|I(\theta)| \neq 0$, which is then nonsingular at the vicinity of symmetry, that is, for the normal case, so that the usual $\sqrt{n}$-asymptotic behaviour holds for the MLEs. Moreover, The upper $2 \times 2$ submatrix of $I(\theta_1)^{-1}$ corresponds to the information matrix for the normal distribution. For large $n$,

$$\widehat{\theta} \xrightarrow{A} N_{p+4}(\theta, I(\theta)^{-1}),$$

and hence, $\widehat{\theta}$ is consistent and asymptotically normal with asymptotic covariance matrix $I(\theta)^{-1}$. For this to be the case, regularity conditions must be satisfied.

We have shown that the Fisher information matrix is not singular, and, moreover, since second derivatives exist and are continuous with respect to each one of the parameters $\sigma, \lambda, \alpha$, and $\beta_j$ for $j = 1, 2, \cdots, p+1$ it is possible to differentiate under the integral sign. This shows that part of the regularity conditions for large sample normality of the maximum likelihood estimators are satisfied. To verify the remaining conditions, following Lin & Stoyanov (2009), for $y > 0$ and $\lambda > 0$, $\liminf_{y\to\infty} \Phi(\lambda y) \geq 1/2$ so that $\frac{1}{2}\frac{\phi(\lambda y)}{\Phi(\lambda y)} \leq \phi(\lambda y) \to 0$ as $y \to \infty$, and for $\lambda < 0$, $\log(\Phi(\lambda y)) \approx -\frac{1}{2}(\lambda y)^2$ for $y \to \infty$. From here it follows that $\Phi(\lambda y) \approx e^{-\frac{1}{2}(\lambda y)^2}$, leading to $\frac{\phi(\lambda y)}{\Phi(\lambda y)} \approx (2\pi)^{-1/2}$ as $y \to \infty$. On the other hand, it is well known that the failure rate of the standard normal distribution $h(y)$ satisfies $\frac{\phi(y)}{1-\Phi(y)} > y$, $\forall y$. Therefore, the third derivatives with respect to the model parameters are bounded by an integrable function. Finally, since the distribution support is independent of model parameters, we have shown that the regularity conditions (see regularity conditions in Lehmann & Casella., (1998) and Casella & Berger., 2002) are satisfied. Thus, we have the following

**Proposition 1.** *If $\hat{\theta}$ is the MLE of $\theta$, then*

$$\hat{\theta} \xrightarrow{A} N_{p+4}(\theta, I(\theta)^{-1}),$$

*resulting that the asymptotic variance of the MLE $\hat{\theta}$ is the inverse of the Fisher information matrix $I(\theta)$, which can be denoted by $\Sigma(\theta) = I(\theta)^{-1}$.*

# 5. Numerical Results

## 5.1. Simulation Study

Results of two simulation studies, one for the BPN model and the other for the ABPN model, report properties such as empirical bias and mean squared error for the maximum likelihood estimators. We consider regression models with errors $BPN(0, \sigma, \alpha)$ and $ABPN(0, \sigma, \alpha, \lambda)$. For each model, 5000 samples of sizes $n = 50$, 150, 300 and 1000 were generated from the BPN and ABPN models, with parameter values given by $\beta_0 = 0.75$, $\beta_1 = 2.25$, $\sigma = 1$ and $\alpha = 0.25,\ 0.75,\ 1.75\ \text{ and }\ 3.5$. For the ABPN model we took $\lambda = 1.0$ and 3.0

Estimators performances were evaluated by computing the absolute value of empirical bias ($|Bias|$ = empirical $|bias|$ value) and the square root of the empirical mean squared error ($\sqrt{MSE}$). Results are presented in Tables 1, 2, and 3. Results show that the absolute value of the bias and the root of the mean square error of the maximum verisimilitude estimators of the model parameters decrease as the sample sizes are increased, ie estimators are approximately unbiased, (see Tables 1, 2, and 3). This indicates a good performance of the MLE for moderate sample sizes. Small bias for large samples is expected given the asymptotic convergence of the MLE discussed above. For small and moderate sample sizes, it can also be depicted from the simulations that the bias $\beta_0$ is small under the BPN model.

Table 1: Simulations for the BPN regression model with 5000 iterations to $\alpha = 0.25$, 0.75, 1.75, 3.5; $\sigma = 1.0$; $\beta_0 = 0.75$ and $\beta_1 = 2.25$, with sample sizes $n = 50$, 150, 300, and 1000 respectively.

| | | $\hat{\beta}_0$ | | $\hat{\beta}_1$ | | $\hat{\sigma}$ | | $\hat{\alpha}$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | n | $\lvert Bias\rvert$ | $\sqrt{MSE}$ | $\lvert Bias\rvert$ | $\sqrt{MSE}$ | $\lvert Bias\rvert$ | $\sqrt{MSE}$ | $\lvert Bias\rvert$ | $\sqrt{MSE}$ |
| 0.25 | 50 | 0.0074 | 0.3302 | 0.0134 | 0.5195 | 0.1294 | 0.1323 | 1.1796 | 1.0744 |
| | 150 | 0.0043 | 0.1583 | 0.0073 | 0.2673 | 0.0557 | 0.0831 | 0.4570 | 0.5542 |
| | 300 | 0.0015 | 0.1018 | 0.0023 | 0.1754 | 0.0315 | 0.0595 | 0.2629 | 0.3859 |
| | 1000 | 0.0013 | 0.0598 | 0.0019 | 0.1010 | 0.0095 | 0.0351 | 0.0758 | 0.2257 |
| 0.75 | 50 | 0.0016 | 0.3600 | 0.0099 | 0.6388 | 0.1053 | 0.1263 | 1.0000 | 1.0886 |
| | 150 | 0.0010 | 0.1590 | 0.0014 | 0.2784 | 0.0374 | 0.0807 | 0.3200 | 0.5743 |
| | 300 | 0.0006 | 0.1141 | 0.0003 | 0.2055 | 0.0191 | 0.0593 | 0.1589 | 0.4130 |
| | 1000 | 0.0003 | 0.0608 | 0.0003 | 0.1045 | 0.0055 | 0.0347 | 0.0430 | 0.2346 |
| 1.75 | 50 | 0.0067 | 0.3324 | 0.0128 | 0.5201 | 0.0669 | 0.1232 | 0.7070 | 1.1659 |
| | 150 | 0.0014 | 0.1738 | 0.0021 | 0.2891 | 0.0162 | 0.0753 | 0.1475 | 0.6376 |
| | 300 | 0.0012 | 0.1228 | 0.0019 | 0.2106 | 0.0073 | 0.0537 | 0.0617 | 0.4495 |
| | 1000 | 0.0003 | 0.0662 | 0.0007 | 0.116 | 0.002 | 0.0300 | 0.0161 | 0.2400 |
| 3.5 | 50 | 0.0041 | 0.2574 | 0.0135 | 0.4795 | 0.0384 | 0.1041 | 0.5292 | 1.2747 |
| | 150 | 0.0027 | 0.1482 | 0.0075 | 0.2454 | 0.0124 | 0.0605 | 0.1737 | 0.6816 |
| | 300 | 0.0025 | 0.1020 | 0.0034 | 0.1795 | 0.0057 | 0.0430 | 0.0729 | 0.4601 |
| | 1000 | 0.0004 | 0.0563 | 0.0013 | 0.0987 | 0.0020 | 0.0238 | 0.0211 | 0.2473 |

Table 2: Simulations for the ABPN regression model with 5000 iterations for $\alpha = 0.25$, 0.75, 1.75, 3.5; $\sigma = 1.0$; $\beta_0 = 0.75$, $\beta_1 = 2.25$ and $\lambda = 1.0$, with sample sizes $n = 50$, 150, 300, and 1000 respectively.

| | | $\alpha = 0.25$ | | $\alpha = 0.75$ | | $\alpha = 1.75$ | | $\alpha = 3.5$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | n | $\lvert Bias\rvert$ | $\sqrt{MSE}$ | $\lvert Bias\rvert$ | $\sqrt{MSE}$ | $\lvert Bias\rvert$ | $\sqrt{MSE}$ | $\lvert Bias\rvert$ | $\sqrt{MSE}$ |
| | 50 | 0.0982 | 0.3843 | 0.0474 | 0.3932 | 0.0174 | 0.3567 | 0.0125 | 0.2776 |
| $\hat{\beta}_0$ | 150 | 0.0941 | 0.3048 | 0.0459 | 0.3121 | 0.0123 | 0.2532 | 0.0020 | 0.1657 |
| | 300 | 0.0442 | 0.234 | 0.0477 | 0.2587 | 0.0074 | 0.2027 | 0.0012 | 0.1236 |
| | 1000 | 0.0394 | 0.1348 | 0.0436 | 0.1729 | 0.0074 | 0.2027 | 0.0011 | 0.0655 |
| | 50 | 0.0053 | 0.3807 | 0.0186 | 0.4251 | 0.0088 | 0.4308 | 0.0026 | 0.3776 |
| $\hat{\beta}_1$ | 150 | 0.0042 | 0.2289 | 0.0021 | 0.2403 | 0.0049 | 0.2299 | 0.0015 | 0.2149 |
| | 300 | 0.0027 | 0.1631 | 0.0019 | 0.1670 | 0.0004 | 0.1691 | 0.0004 | 0.1570 |
| | 1000 | 0.0002 | 0.0815 | 0.0014 | 0.0894 | 0.0001 | 0.0929 | 0.0002 | 0.0848 |
| | 50 | 0.0862 | 0.2314 | 0.0906 | 0.2161 | 0.0629 | 0.1787 | 0.0492 | 0.1331 |
| $\hat{\sigma}$ | 150 | 0.0149 | 0.1504 | 0.0236 | 0.1533 | 0.0209 | 0.1203 | 0.0165 | 0.0813 |
| | 300 | 0.0096 | 0.1038 | 0.0026 | 0.1175 | 0.0159 | 0.0915 | 0.0073 | 0.0576 |
| | 1000 | 0.0073 | 0.0538 | 0.0107 | 0.0700 | 0.0057 | 0.0555 | 0.0022 | 0.0308 |
| | 50 | 0.8782 | 3.3055 | 0.4588 | 2.4702 | 0.1659 | 1.3517 | 0.0139 | 0.4590 |
| $\hat{\lambda}$ | 150 | 0.4562 | 1.2525 | 0.2456 | 1.1182 | 0.0451 | 0.4810 | 0.0047 | 0.2400 |
| | 300 | 0.3315 | 0.7103 | 0.1639 | 0.6036 | 0.0003 | 0.3563 | 0.0025 | 0.1680 |
| | 1000 | 0.1152 | 0.3535 | 0.1199 | 0.3844 | 0.0050 | 0.2320 | 0.0003 | 0.0893 |
| | 50 | 1.4899 | 1.1151 | 1.3681 | 1.0916 | 1.0502 | 1.1270 | 0.7574 | 1.3624 |
| $\hat{\alpha}$ | 150 | 0.8874 | 0.7949 | 0.7259 | 0.7137 | 0.4250 | 0.6799 | 0.2532 | 0.7353 |
| | 300 | 0.5649 | 0.6921 | 0.4705 | 0.6037 | 0.2213 | 0.5051 | 0.1205 | 0.5119 |
| | 1000 | 0.1762 | 0.4440 | 0.2092 | 0.4476 | 0.0702 | 0.3069 | 0.0353 | 0.2707 |

## 5.2. Numerical Illustration

The following illustration is based on a data set including 202 Australian athletes, which can be downloaded at the following directory `http://azzalini.stat.unipd.it/SN/`. The data set is related to certain body features such as height, weight, and body mass index, among others, for all 202 athletes.

The linear model considered is

$$\texttt{Bfat}_k = \beta_0 + \beta_1 \texttt{bmi}_k + \beta_2 \texttt{lbm}_k + \varepsilon_k,$$

for $k = 1, 2, \dots, 202$, where $\texttt{Bfat}_k$ is the body fat percentage for the $k$-th athlete, and the covariates $\texttt{bmi}_k$ and $\texttt{lbm}_k$ are the body mass index and lean body mass, respectively, for the $k$-th athlete.

We start by fitting the regression model under the assumption that the error term follows the ordinary normal model. Summary statistics are shown seen in Table 4, in which quantities $\sqrt{b_1}$ and $b_2$ represent sample asymmetry and kurtosis coefficients. As well as in Figure 1(a), there is in indication that the normal

Table 3: Simulations for the ABPN regression model with 5000 iterations for $\alpha = 0.25$, 0.75, 1.75, 3.5; $\sigma = 1.0$; $\beta_0 = 0.75$, $\beta_1 = 2.25$ and $\lambda = 3.0$, with sample sizes $n = 50$, 150, 300, and 1000 respectively.

| | | $\alpha = 0.25$ | | $\alpha = 0.75$ | | $\alpha = 1.75$ | | $\alpha = 3.5$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $n$ | $\|Bias\|$ | $\sqrt{MSE}$ | $\|Bias\|$ | $\sqrt{MSE}$ | $\|Bias\|$ | $\sqrt{MSE}$ | $\|Bias\|$ | $\sqrt{MSE}$ |
| | 50 | 0.1563 | 0.2428 | 0.0291 | 0.2262 | 0.0047 | 0.2321 | 0.0415 | 0.2289 |
| $\hat{\beta}_0$ | 150 | 0.1131 | 0.1439 | 0.0848 | 0.1519 | 0.0293 | 0.1583 | 0.0061 | 0.1564 |
| | 300 | 0.1024 | 0.123 | 0.0861 | 0.1307 | 0.0279 | 0.1306 | 0.0056 | 0.1329 |
| | 1000 | 0.0533 | 0.0822 | 0.054 | 0.0959 | 0.0102 | 0.0964 | 0.0011 | 0.0839 |
| | 50 | 0.0014 | 0.3422 | 0.0037 | 0.2916 | 0.0035 | 0.3106 | 0.0016 | 0.3266 |
| $\hat{\beta}_1$ | 150 | 0.0001 | 0.1604 | 0.0005 | 0.1527 | 0.0015 | 0.1790 | 0.0019 | 0.1871 |
| | 300 | 0.0002 | 0.1089 | 0.0039 | 0.1160 | 0.0038 | 0.1287 | 0.0010 | 0.1323 |
| | 1000 | 0.0001 | 0.0578 | 0.0001 | 0.0614 | 0.0003 | 0.0683 | 0.0012 | 0.0707 |
| | 50 | 0.0633 | 0.1732 | 0.0696 | 0.1632 | 0.0546 | 0.1471 | 0.0417 | 0.1263 |
| $\hat{\sigma}$ | 150 | 0.0189 | 0.0890 | 0.0166 | 0.0891 | 0.0132 | 0.0845 | 0.0123 | 0.0722 |
| | 300 | 0.0127 | 0.0634 | 0.0089 | 0.0609 | 0.0071 | 0.0593 | 0.0079 | 0.0526 |
| | 1000 | 0.0062 | 0.0354 | 0.0034 | 0.0328 | 0.0013 | 0.0319 | 0.0029 | 0.0284 |
| | 50 | 1.9560 | 3.9919 | 1.8591 | 3.1235 | 0.5809 | 2.2944 | 0.3756 | 1.8389 |
| $\hat{\lambda}$ | 150 | 2.0091 | 3.4448 | 1.5292 | 3.0733 | 0.6038 | 1.9650 | 0.2244 | 1.1408 |
| | 300 | 1.3391 | 2.2723 | 1.0631 | 2.1654 | 0.3976 | 1.5633 | 0.0730 | 0.7056 |
| | 1000 | 0.4955 | 0.7335 | 0.4657 | 0.7776 | 0.1183 | 0.6815 | 0.0132 | 0.3613 |
| | 50 | 1.0601 | 1.0039 | 0.9117 | 1.0110 | 0.6094 | 1.1338 | 0.2134 | 1.3831 |
| $\hat{\alpha}$ | 150 | 1.0673 | 0.9535 | 0.8377 | 0.8987 | 0.4103 | 0.8430 | 0.1443 | 0.9856 |
| | 300 | 0.9233 | 0.9153 | 0.7451 | 0.8720 | 0.3270 | 0.7984 | 0.0795 | 0.8308 |
| | 1000 | 0.4613 | 0.6465 | 0.4533 | 0.7030 | 0.1156 | 0.6352 | 0.0356 | 0.5782 |

symmetric model may not be the most adequate and that an asymmetric model such as the PN or its asymmetrical bimodal extension, namely the model ABPN, can present a better fit.

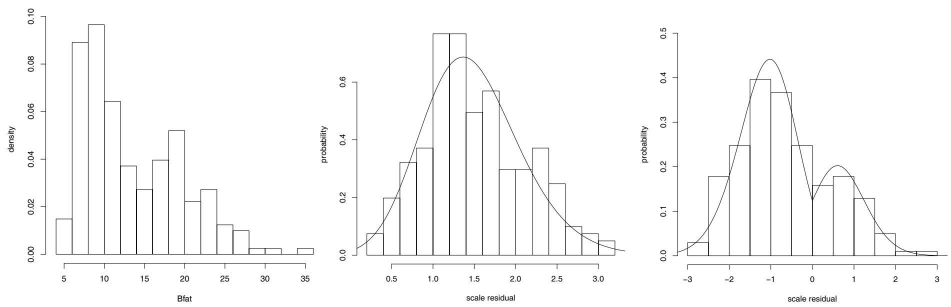
Table 4: Descriptive statistics for the data set.

| n | $\bar{e}$ | $s_e$ | $\sqrt{b_1}$ | $b_2$ |
|---|---|---|---|---|
| 202 | 0.0000 | 0.8559 | $-0.5920$ | 2.5484 |

Additionally, the Shapiro-Wilk test for normality, with p-values given in parenthesis, is given by $SW = 0.9811(0.008)$, giving then indication that model errors are not normally distributed. Thus, we fitted the regressions based on the PN and ABPN models. In order to future investigate the model fit, we computed the scaled residuals $e_k = (y_k - x_k'\widehat{\beta})/\widehat{\sigma}$ for PN and ABPN linear Models.

Figures 1(b) and (c) reveal the fit of regression models PN and ABPN. We found that the PN model fits the data better than the ordinary normal regression model. Moreover the ABPN model presents a better fit than the PN model. The main idea is that the ABPN model is able to capture asymmetry and bimodality and the others are not.

The models above were fit using software R nonlinear function `nlm` program from (see R Development Core Team 2015).




(a) (b) (c)

FIGURE 1: (a) Histogram for the body fat percentage. Histogram and model fitted residuals under: (b) PN model and (c) ABPN model.

Moreover, the results in Table 6 present estimates, with standard errors in parenthesis, for the model parameters. It also reveals that, according to the ABPN regression model, the percentage (%) of body fat depends on the `bmi` and `lbm` quantities the athletes in the sample.

The model selection approaches considered are the BIC, written as $BIC = -2\ell(\cdot) + k\ln(n)$ and CAIC, written as $CAIC = -2\ell(\cdot) + k(1 + \ln(n))$, where $k$ is the number of unknown parameters for the model under study. According to either the BIC or CAIC scores (see Table 6), the ABPN linear model presents the best fit when compared to the normal and PN linear models. For the sake of comparison, we also fitted the regression model with the error terms distributed as a mixture of two normal distributions. Parameter estimates were obtained using function `mixreg` in R. These are shown in Table 5.

TABLE 5: Descriptive statistics for the data set.

| $j$ | $\hat{\beta}_{j0}$ | $\hat{\beta}_{j1}$ | $\hat{\beta}_{j2}$ | $\hat{\sigma}_j$ | $\hat{p}_j$ |
|---|---|---|---|---|---|
| 1 | −0.825 | −0.194 | 1.011 | 2.923 | 0.344 |
| 2 | 5.280 | −0.509 | 1.862 | 15.961 | 0.656 |

TABLE 6: Parameter estimates (SD) for Normal, Power Normal, and ABPN models.

| Linear model | Normal | PN | ABPN |
|---|---|---|---|
| $\widehat{\beta}_0$ | –0.546(2.417) | –10.745(3.279) | 0.920(1.620) |
| $\widehat{\beta}_0^*$ | | | -1.586(2.468) |
| $\widehat{\beta}_1$ | 1.965(0.148) | 1.902(0.143) | 1.789(0.096) |
| $\widehat{\beta}_2$ | –0.479(0.032) | –0.459(0.033) | –0.400(0.021) |
| $\widehat{\sigma}$ | 4.205(0.209) | 6.907(0.681) | 3.953(0.223) |
| $\widehat{\alpha}$ | | 8.929(3.859) | 3.698(0.588) |
| $\widehat{\lambda}$ | | | –0.598(0.111) |
| BIC | 1174.69 | 1173.55 | 1170.05 |
| CAIC | 1178.69 | 1178.55 | 1176.05 |

We obtained $BIC = 1170.17$ and $CAIC = 1179.17$, meaning that, according to the BIC and CAIC criteria, the regression model ABPN fits the data better than the mixture of two normal distributions.

Following Therneau, Grambsch & Fleming (1990), we can adapt the deviance component residual for the ABPN model with no censored data by considering

$$r_{M_i} = 1 + \ln\left(\hat{S}_{ABPN}(y)\right), \quad i = 1, 2, \ldots, n, \tag{7}$$

where $S_{ABPN}(y)$ is the ML estimate of the reliability function of the ABPN model.

Therneau et al. (1990) proposed the deviance component residual as a transformation of the martingale type residual so that the deviance component residual for noncensored data can be taken as

$$r_{MT_i} = sgn(r_{M_i})\left\{-2\left[r_{M_i} + \ln(1 - r_{M_i})\right]\right\}^{1/2}, \ i = 1, 2, \ldots, n. \tag{8}$$

We use the residual $r_{MT_i}$ as a residual type martingale, given that they are symmetrically distributed around zero. Furthermore, to evaluate the global influence of each observation on the parameter estimation, Cook's distance was computed by removing one observation at a certain time and evaluating estimation changes for parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}, \sigma, \alpha, \lambda)$. This distance is computed as

$$GDC_i(\boldsymbol{\beta}) = \frac{1}{(p+1)+3}\left[\left(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)}\right)' \hat{\Sigma}^{-1}_{\hat{\boldsymbol{\theta}}}\left(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)}\right)\right], \ i = 1, \ldots, n \tag{9}$$

where $p+1$ is the number of regression coefficients in the regression model, $\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}$ is an estimator for the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_i$ is the maximum likelihood estimator $\boldsymbol{\beta}$ after removing the $i$-th observation.

Figure 2(a) and (b) presents Cook's distances values of the residuals versus fitted values from which it can be depicted that (a) there are two influential observations, namely #56 and #133, and, moreover, (b) for some observations Cook's distances fall a little outside the bands $-2$ and 2. This shows that no influential observations are present in the data.

The impact is measured by the relative changes in the estimates, represented as

$$RC_{\theta_j} = \left|\frac{\hat{\theta}_j - \hat{\theta}_{j(I)}}{\hat{\theta}_j}\right| * 100, \tag{10}$$

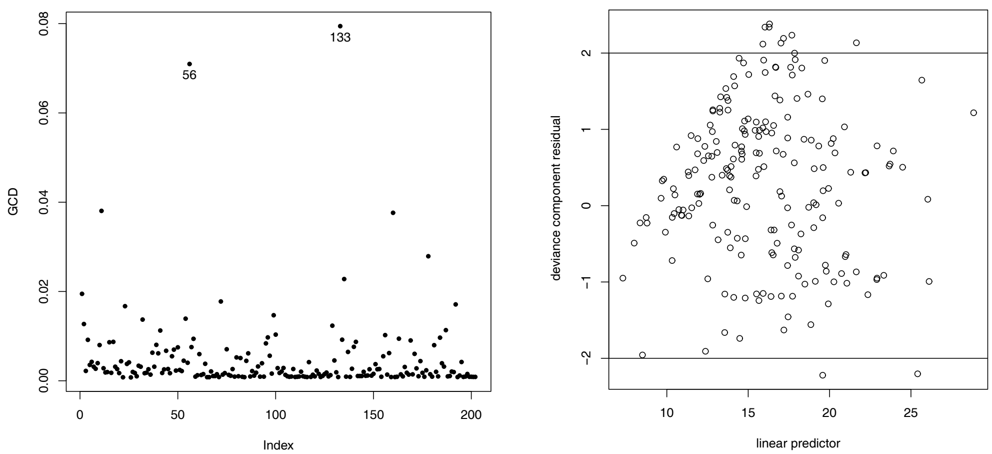
where $\hat{\theta}_j$ denotes the maximum likelihood estimate for the parameter $\theta_j$ including all observations, and $\hat{\theta}_{j(I)}$ denotes the estimate of the same parameter while deleting the influential observations.

Table 7 depicts the relative changes on the parameter estimates.

Table 7: Relative change, $RC\%$ for model parameters.

| Index | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}$ | $\hat{\lambda}$ | $\hat{\alpha}$ |
|---|---|---|---|---|---|---|
| $\{56\}$ | 0.3560 | 0.0184 | 0.0204 | 0.0154 | 0.0238 | 0.0331 |
| $\{133\}$ | 0.1484 | 0.0292 | 0.0129 | 0.0032 | 0.0448 | 0.0075 |
| $\{56, 133\}$ | 0.5097 | 0.0112 | 0.0050 | 0.0177 | 0.0162 | 0.0618 |

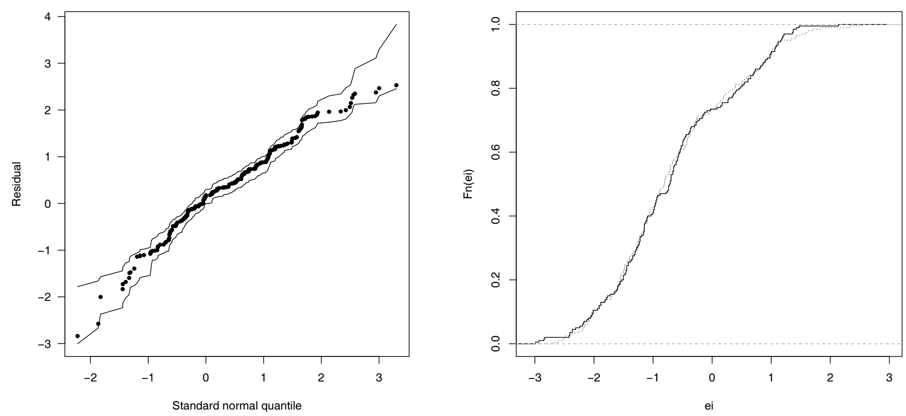
Table 7 indicates that no observation exerts a great influence on the maximum likelihood estimates. This corroborates the results that are presented above.




(a) (b)

Figure 2: Plot for the ABPN model (a) Index versus GCD and (b) deviance component residual versus predictor.

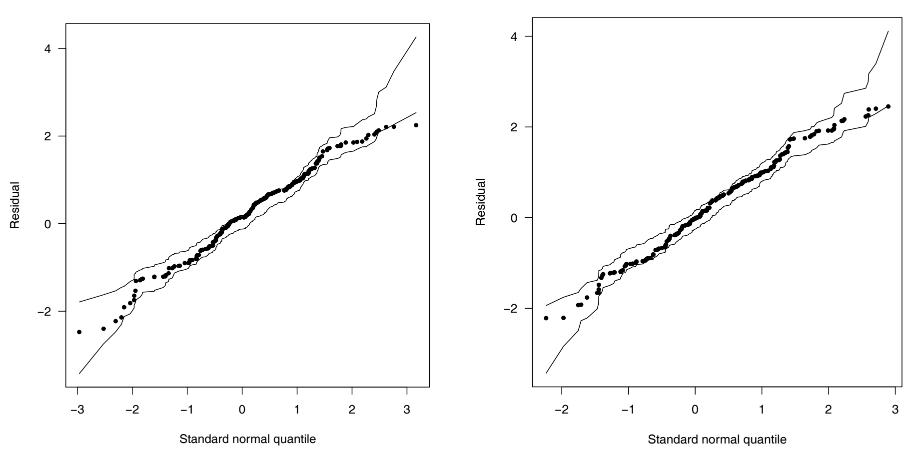
Figures 3(a) and (b) present the QQ-plot with envelops for the deviance component residual for the ABPN models. This also indicates a good fit for the ABPN linear model and the empirical cdf for the scaled residuals under the ABPN model (solid line). The dotted line corresponds to the cfd for the ABPN model.




(a) (b)

Figure 3: (a) QQ-plot with envelope under the ABPN model and (c) cdf for the ABPN model.

It can be noticed that envelops for the ABPN model also indicates the presence of outlying (influential) observations under the ABPN regression model. Figures 4(a) and (b) presents the QQ-plot with envelops for the deviance component residual for the normal and PN regression models.




(a) (b)

Figure 4: QQ-plot with envelopes (a) normal model and (b) PN model.

# 6. Final Discussion

In this paper we extended the model in Bolfarine et al. (2012) for the case of regression models. Emphasis was placed on the asymmetric bimodal power-normal (ABPN) distribution. The skew-normal model (Azzalini 1985) is a special case. Estimations were made by implementing the maximum likelihood approach. Large sample estimates were obtained by using the observed information matrix (minus the inverse of the estimated Hessian matrix). Results from a real data application for the linear model situation illustrates the usefulness of the model developed.

# Acknowledgments

The authors acknowledge helpful comments and suggestions from the referee which substantially improved the presentation.

[Received: April 2015 — Accepted: February 2016]

# References

Arellano-Valle, R. B., Bolfarine, H. & Vilca-Labra, F. (1996), 'Ultrastructural elliptical models', *The Canadian Journal of Statistics* **24**(2), 207–216.

Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**(2), 171–178.

Bolfarine, H., Martínez-Flórez, G. & Salinas, H. S. (2012), 'Bimodal symmetric-asymmetric power-normal families', *Communications in Statistics-Theory and Methods* . DOI:10.1080/03610926.2013.765475.

Cancho, V. G., Lachos, V. H. & Ortega, E. M. M. (2010), 'A nonlinear regression model with skew-normal errors', *Statistical Papers* **51**(3), 547–558.

Casella, G. & Berger, R. (2002), *Statistical Inference*, 2 edn, Thomson, Singapore.

Cordeiro, G. M., Ferrari, S. L. P., Uribe-Opazo, M. A. & Vasconcellos, K. L. P. (2000), 'Corrected maximum likelihood estimation in a class of symmetric nonlinear regression models', *Statistics & Probability Letters* **46**(4), 317–328.

De Veaux, R. (1989), 'Mixtures of linear regressions', *Computational Statistics and Data Analysis* **8**, 227–245.

Durrans, S. R. (1992), 'Distributions of fractional order statistics in hydrology', *Water Resources Research* **28**(6), 1649–1655.

Galea, M., Paula, G. A. & Cysneiros, J. A. (2005), 'On diagnostic in symmetrical nonlinear models', *Statistics & Probability Letters* **73**(4), 459–467.

Gupta, R. D. & Gupta, R. C. (2008), 'Analyzing skewed data by power normal model', *TEST* **17**(1), 197–210.

Lange, K. L., Little, R. J. A. & Taylor, J. M. G. (1989), 'Robust Statistical Modeling Using the t Distribution', *Journal of the American Statistical Association* **84**(408), 881–896.

Lehman, E. L. & Casella, G. (1998), *Theory of Point Estimation*, 2 edn, Springer, New York.

Lin, G. & Stoyanov, J. (2009), 'The logarithmic Skew-Normal Distributions are Moment-Indeterminate', *Journal of Applied Probability Trust* **46**, 909–916.

Martínez-Flórez, G., Bolfarine, H. & Gómez, H. W. (2015), 'Likelihood-based inference for the power regression model', *SORT* **39**(2), 187–208.

Pewsey, A., Gómez, H. W. & Bolfarine, H. (2012), 'Likelihood based inference for distributions of fractional order statistics', *TEST* **21**(4), 775–789.

Quandt, R. (1958), 'The estimation of the parameters of a linear regression system obeying two separate regimes', *Journal of the American Statistical Association* **53**(284), 873–880.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org/

Rego, L. C., Cintra, R. J. & Cordeiro, G. M. (2012), 'On some properties of the beta normal distribution', *Communications in Statistics-Theory and Methods* **41**, 3722–3738.

Therneau, T., Grambsch, P. & Fleming, T. (1990), 'Martingale-based residuals for survival models', *Biometrika* **77**, 147–160.

Turner, T. (2000), 'Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions', *Applied Statistics* **49**(3), 371–384.

Young, D. S. & Hunter, D. R. (2010), 'Mixtures of regressions with predictor-dependent mixing proportions', *Computational Statistics and Data Analysis* **54**(10), 2253–2266.