SORT 39 (2) July-December 2015, 187-208

Likelihood-based inference for the power regression model

Guillermo Martínez-Flórez[1], Heleno Bolfarine[2] and Héctor W. Gómez[3]

Abstract

In this paper we investigate an extension of the power-normal model, called the alpha-power model and specialize it to linear and nonlinear regression models, with and without correlated errors. Maximum likelihood estimation is considered with explicit derivation of the observed and expected Fisher information matrices. Applications are considered for the Australian athletes data set and also to a data set studied in Xie et al. (2009). The main conclusion is that the proposed model can be a viable alternative in situations were the normal distribution is not the most adequate model.

MSC: 60E05.

Keywords: Correlation, maximum likelihood, power-normal distribution, regression.

1. Introduction

Linear and nonlinear regression models are statistical techniques typically used for modeling and studying relationships between variables in several areas of human knowledge such as biomedical and agricultural sciences, engineering, and many others, being extremely useful for data analysis. One important step in regression analysis is parameter estimation, usually made under the assumption of normality. However, there are situations were the normal assumption is not realistic and several distributions have been suggested as alternatives to the normal model. Among such models we have the Student-t, logistic and exponential power distributions (Cordeiro et al., 2000 and Galea et al., 2005), whereas for the asymmetric nonlinear model we have only the work of Cancho

[1] Departamento de Matemáticas y Estadística, Facultad de Ciencias, Universidad de Córdoba, Córdoba, Colombia, gmartinez@correo.unicordoba.edu.co

[2] Departamento de Estatística, IME, Universidad de Sao Paulo, Sao Paulo, Brasil, hbolfar@ime.usp.br

[3] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile, hector.gomez@uantof.cl
Received: February 2013
Accepted: April 2015

et al. (2008). In this paper we suggest an alternative asymmetric model, the alpha-power model, for fitting linear and nonlinear regression models. The maximum likelihood approach is used for parameter estimation and the normality assumption can be tested using the likelihood ratio statistics since large sample properties are satisfied for the maximum likelihood estimator (Pewsey et al., 2012). Real data applications reveal that the model considered can be a viable alternative to existing asymmetric models in the literature.

The paper is organized as follows. In Section 2 asymmetric models are reviewed and some of their main properties discussed. Emphasis is placed on the alpha-power model, a special case of which is the power-normal model (Gupta and Gupta, 2008). In Section 3 a general definition of asymmetric regression models is presented and previous works on linear and nonlinear versions are listed. Section 4 is devoted to the study of the linear multiple regression model with power-normal errors. Inference via maximum likelihood for this model is also considered. The nonlinear power-normal model is considered in Section 5. Estimation is considered via maximum likelihood. The autoregressive model is studied in Section 6, with inference via maximum likelihood. A score type statistic is developed for testing null correlation. A small-scale Monte Carlo study is conducted in Section 7, including a study on model robustness. The main conclusion is that estimators under the regression model studied are fairly robust against data contamination. Results of two real data applications are reported illustrating the usefulness of the models considered in Section 8. In Section 9 (Appendix), we present the elements for the observed information matrices for the models considered in the previous Sections.

2. Skew distributions

Lehmann (1953) studied the family of distributions with a general distribution function given by

$$\mathscr{F}_F(z;\alpha) = \{F(z)\}^{\alpha}, \quad z \in \mathbb{R}, \tag{1}$$

where F is a distribution function and α is a rational number.

Durrans (1992), in a hydrological context, extended Lehmann's model by considering α real (and positive) for the special case $F = \Phi$, the distribution function of the normal distribution. We consider in this paper an extension of Lehmann's model, which we call the *alpha-power model*, with density function given by

$$\varphi_f(z;\alpha) = \alpha f(z)\{F(z)\}^{\alpha-1}, \qquad z \in \mathbb{R}, \;\; \alpha \in \mathbb{R}^+, \tag{2}$$

where F is an absolutely continuous distribution function with density function $f = dF$. Properties for a particular case of this distribution (with $F = \Phi$, the distribution function of the normal distribution), were studied in Gupta and Gupta (2008). We use the notation $Z \sim P_F(\alpha)$. We refer to this model as the standard alpha-power distribution (see also Pewsey et al., 2012). This is an alternative to asymmetric models with higher amounts of asymmetry and kurtosis as is the case with the skew-normal distribution (Azzalini, 1985), see also Mudholkar and Hutson (2000) for some special cases. Parameter α is a shape parameter that controls the amount of asymmetry in the distribution. Extensions of the power-normal model are also considered in Rego et al. (2012).

In the particular case that $F = \Phi$, the distribution function of the normal distribution, Z is said to follow a power-normal distribution (denoted PN(α)) with density function given by

$$\varphi(z;\alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}, \qquad z \in \mathbb{R}. \tag{3}$$

If Z is a random variable from a standard $P_F(\alpha)$ distribution then the location-scale extension of Z, $X = \xi + \eta Z$, where $\xi \in \mathbb{R}$ and $\eta \in \mathbb{R}^+$, has probability density function given by

$$\varphi_F(x;\xi,\eta,\alpha) = \frac{\alpha}{\eta} f\left(\frac{x-\xi}{\eta}\right)\left\{F\left(\frac{x-\xi}{\eta}\right)\right\}^{\alpha-1}. \tag{4}$$

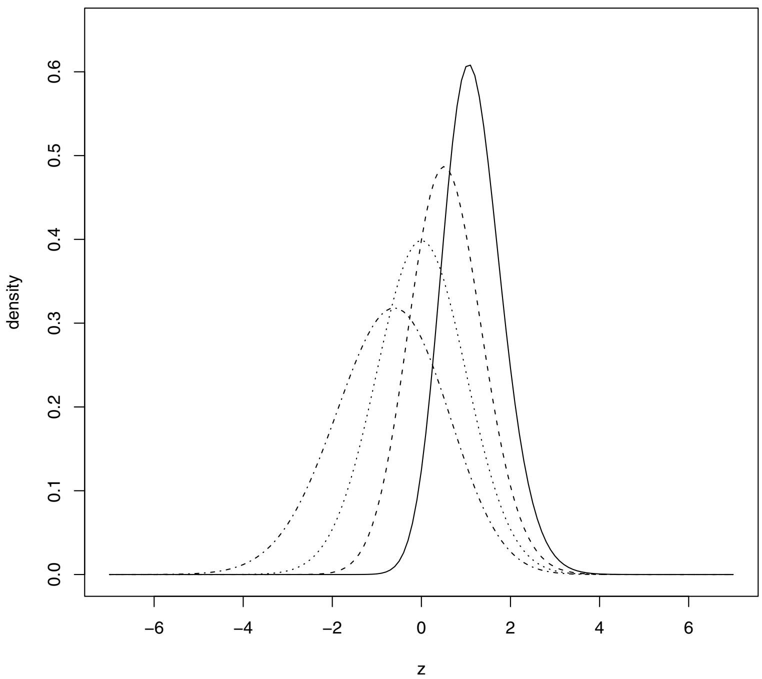


Figure 1: *Density $\varphi_Z(z;\alpha)$ for α equals to* 5 *(solid line),* 2 *(dashed line),* 1 *(dotted line) and* 0.5 *(dashed and dotted line).*

We will denote this extension by using the notation $X \sim P_F(\xi,\eta,\alpha)$. Notice that this model can be further extended by considering $\xi_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ replacing ξ, where $\boldsymbol{\beta}$ is an unknown vector of regression coefficients and x_i a vector of known regressors possibly correlated with the response vector.

As can be deduced from Figure 1, parameter α controls also the distribution kurtosis. Moreover, it can be noticed that for $\alpha > 1$, the kurtosis is greater than that of the normal distribution and, for $0 < \alpha < 1$, the opposite is observed.

Pewsey et al. (2012) derived the Fisher information matrix for the location-scale version of the power-normal model and have shown that it is not singular for $\alpha = 1$. We recall that the Fisher information matrix for the skew-normal distribution (Azzalini, 1985) is singular under the symmetry hypothesis. Hence, with the power-normal model, normality can be tested using ordinary large sample properties of the likelihood ratio statistics. They also found the asymmetry and kurtosis ranges to be $[-0.6115, 0.9007]$ and $[1.7170, 4.3556]$, respectively. This illustrates the fact that the model is more flexible, respective to kurtosis, than the model skew-normal (Azzalini, 1985), for which the kurtosis range is given by $[3, 3.8692)$.

A generalization for the PN(α) model is given in Eugene et al. (2002), by introducing the beta-normal distribution, denoted BN(α,β), with BN(α, 1)=PN(α). Therefore, model BN(α,β) is more flexible than model PN(α). However, model BN(α,β) contains two parameters to be estimated and the asymmetry and kurtosis ranges for both models are the same, namely $[-0.6115, 0.9007]$ and $[1.7170, 4.3556]$, respectively. General properties of the model BN where studied by Gupta and Nadarajah (2004) and Rego et al. (2012).

3. The asymmetric regression model

The multiple regression model is typically represented by

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{5}$$

where $\boldsymbol{\beta}$ is a vector of unknown constants and x_i are values of known explanatory variables. The error terms ϵ_i are independent random variables with N($0,\sigma^2$) distribution. It may occur that the symmetry assumption is not an adequate assumption for the error term so that an asymmetric model may present a better fit for the data set under study. As seen in the literature, some asymmetric distributions that can be considered are the epsilon-skew-normal (ESN, Mudholkar and Hutson, 2000) distribution, the skew-exponential power (SEP, see Azzalini, 1986) distribution and the Beta-Normal (BN) distribution, among others. Hutson (2004) replaces in (5) the normal distribution by the ESN distribution, DiCiccio and Monti (2004) consider that the error terms follow model SEP while Razzaghi (2009) consider the BN distribution for fitting a quadratic dose-response modeling. Asymmetric nonlinear regression is studied in Cancho et al. (2008) by considering that the error terms follow a skew-normal model distribution. Xie et al. (2009) studied the case where the error term follows the skew-t-normal model (see Gómez et al., 2007).

4. The multiple regression model with PN errors

In this section, we assume under the ordinary multiple regression model that the error term follows a PN (denoted PNR) distribution with parameters 0, η_e and α, that is,

$$\epsilon_i \sim PN(0,\eta_e,\alpha) \quad \text{for} \quad i=1,2,\ldots,n.$$

Hence, it follows that the density function of ϵ_i is given by

$$\varphi(\epsilon_i;\boldsymbol{\beta},\eta_e,\alpha) = \frac{\alpha}{\eta_e}\phi\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)\left\{\Phi\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)\right\}^{\alpha-1}, \quad i=1,2,\ldots,n, \tag{6}$$

Therefore, it follows that y_i given $\mathbf{x}_i$, $(y_i|\mathbf{x}_i)$, also follows a PN distribution, that is,

$$y_i|\mathbf{x}_i \sim PN(\mathbf{x}_i^\top\boldsymbol{\beta},\eta_e,\alpha), \quad i=1,2,\ldots,n, \tag{7}$$

with location parameter $\mathbf{x}_i^\top\boldsymbol{\beta}$, $i=1,2,\ldots,n$, scale parameter η_e and shape parameter α. Under the PN model,

$$E(\epsilon_i) = \alpha\eta_e\int_0^1 \Phi^{-1}(z)z^{\alpha-1}dz \neq 0$$

so that the expected value of the error term is not null as is the case under normality. Therefore, $E(y_i) \neq \mathbf{x}_i^\top\boldsymbol{\beta}$ and we have to make the following correction to obtain the regression line as the expected value of the response variable: $\beta_0^* = \beta_0 + \mu_\epsilon$, where $\mu_\epsilon = E(\epsilon_i)$. Thus,

$$E(y_i) = \mathbf{x}_i^\top\boldsymbol{\beta}^* \ \text{ where } \ \boldsymbol{\beta}^* = (\beta_0^*,\beta_1,\ldots,\beta_p)^\top.$$

The next section discusses maximum likelihood estimation for the corrected model.

4.1. Inference for the multiple PNR model

We discuss in the following maximum likelihood estimation for the multiple power-normal regression model. A detailed derivation of the Fisher information matrix is considered, resulting that it is nonsingular at the vicinity of symmetry.

4.2. Likelihood and score functions

Considering now a matrix notation where $\mathbf{y}$ denotes the vector with entries y_i and dimension $(n\times 1)$ and $\mathbf{X}$ the $(n\times(p+1))$-matrix with rows $\mathbf{x}_i^\top$, the likelihood function for $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\eta_e,\alpha)^\top$ given a random sample of size n, $\mathbf{y}=(y_1,y_2,\ldots,y_n)^\top$, can be written as

$$\ell(\boldsymbol{\theta};\mathbf{y})=n\log\left(\frac{\alpha}{\eta_e}\right)-\frac{1}{2\eta_e^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+(\alpha-1)\sum_{i=1}^{n}\log\left\{\Phi\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)\right\},$$

with score function:

$$U(\boldsymbol{\beta})=\frac{1}{\eta_e^2}\mathbf{X}^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})-\frac{\alpha-1}{\eta_e}\mathbf{X}^\top\Lambda_1,\qquad U(\alpha)=n\left(\frac{1}{\alpha}+\overline{u}\right),\tag{8}$$

$$U(\eta_e)=-\frac{n}{\eta_e}+\frac{1}{\eta_e^3}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})-\frac{\alpha-1}{\eta_e^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\Lambda_1\tag{9}$$

where

$$\Lambda_1=(w_1,\ldots,w_n)^\top\qquad\text{and}\qquad u_i=\log\left\{\Phi\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)\right\},$$

with $w_i=\phi\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)/\Phi\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)$, for $i=1,2,\ldots,n$. After some algebraic manipulations, maximum likelihood estimating equations are given by

$$\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{MQ}-(\alpha-1)\eta(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\Lambda_1,\qquad\alpha=-\frac{1}{\overline{u}},\tag{10}$$

$$\eta=\frac{(1-\alpha)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\Lambda_1}{2n}+\frac{\sqrt{(1-\alpha)^2(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\Lambda_1\Lambda_1^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+4n(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}}{2n},\tag{11}$$

where $\hat{\boldsymbol{\beta}}_{MQ}=(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$.

Hence, the maximum likelihood estimator for the parameter vector $\boldsymbol{\beta}$ is equal to the least squares estimator for $\boldsymbol{\beta}$ plus the symmetry correcting term. No analytical solutions are available for the likelihood equations and hence they have to be solved numerically.

For the simple linear regression model, namely $p=1$, the following system of equations results

$$\beta_1=\eta_e(\alpha-1)\frac{S_{xw}}{S_x^2}+\frac{S_{xy}}{S_x^2},\qquad\beta_0=-\eta_e(\alpha-1)\overline{w}+\overline{y}-\beta_1\overline{x},\qquad\alpha=-\frac{1}{\overline{u}},$$

and

$$\eta_e = \frac{(1-\alpha)(\overline{wy}-\beta_0\overline{w}-\beta_1\overline{wx})}{2n} + \frac{\sqrt{(1-\alpha)^2(\overline{wy}-\beta_0\overline{w}-\beta_1\overline{wx})^2+4n\sum_{i=1}^n(y_i-\beta_0-\beta_1x_i)^2}}{2n}$$

with

$$S_x^2=\sum_{i=1}^n(x_i-\overline{x})^2/n,\ \ S_{xy}=\sum_{i=1}^n(x_i-\overline{x})(y_i-\overline{y})/n\ \text{ and }\ S_{wx}=\sum_{i=1}^n(w_i-\overline{w})(x_i-\overline{x})/n,$$

where $\overline{w}=\sum_{i=1}^n w_i/n$, $\overline{u}=\sum_{i=1}^n u_i/n$, $\overline{x}=\sum_{i=1}^n x_i/n$, $\overline{x^2}=\sum_{i=1}^n x_i^2/n$, $\overline{y}=\sum_{i=1}^n y_i/n$, $\overline{xy}=\sum_{i=1}^n x_iy_i/n$, $\overline{wx}=\sum_{i=1}^n w_ix_i/n$ and $\overline{wy}=\sum_{i=1}^n w_iy_i/n$.

For $\alpha=1$, the model with normal error terms follow and the estimators reduce to the well known $\hat{\beta}_1=\frac{S_{xy}}{S_x^2}$, $\hat{\beta}_0=\overline{y}-\hat{\beta}_1\overline{x}$ and $\hat{\eta}_e=\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2}$. To initialize the likelihood approach, we can take as initial values the vector $\hat{\boldsymbol{\beta}}$ and for parameter η_e the ones obtained by the least squares approach. They can be computed as follows: for $\epsilon_i^*=\epsilon_i-\mu_\epsilon$, we have that $E(\epsilon^*)=0$ and $Var(\epsilon^*)=\eta_e^2\Phi_2(\alpha)$, where Φ_2 is the variance of the random variable $\mathrm{PN}(0,1,\alpha)$.

Hence, after minimizing the error sum of squares, namely,

$$\sum_{i=1}^n \epsilon_i^{*2}=\sum_{i=1}^n\left(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}^*\right)^2$$

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

$$\hat{\boldsymbol{\beta}}^*=(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}\ \text{ and }\ \hat{\eta}_e^2=\frac{\Phi_2^{-1}(\hat{\alpha})}{n-2}\sum_{i=1}^n\left(y_i-\hat{\beta}_0^*-\hat{\beta}_1x_i\right)^2.$$

On the other hand, an initial value for α can be obtained by fitting the PN model for the errors obtained or by using the elemental percentile approach of Castillo and Hadi (1995), assuming β and η_e known (usually computed using the least-squares approach).

The elements needed to compute the observed information matrix are given in Appendix 10.1. The expected (Fisher) information matrix follows then by taking expectations of those components (multiplied by n^{-1}).

Approximation $N_{p+3}(\boldsymbol{\theta},\Sigma_{\hat{\boldsymbol{\theta}}})$ can be used to construct confidence intervals for θ_r, which are given by $\hat{\boldsymbol{\theta}}_r\mp z_{1-\delta/2}\sqrt{\hat{\sigma}(\hat{\boldsymbol{\theta}}_r)}$, where $\hat{\sigma}(\cdot)$ corresponds to the r-th diagonal element of the matrix $\Sigma_{\hat{\boldsymbol{\theta}}}$ and $z_{1-\delta/2}$ denotes $100(\delta/2)$-quantile of the standard normal distribution.

For the simple linear regression model, that is, $p = 1$, denoting the elements of the observed information matrix by

$$i_{\beta_0\beta_0},\ i_{\beta_1\beta_0},\ i_{\eta_e\beta_0},\ i_{\alpha\beta_0},\dots,i_{\eta_e\eta_e},\ i_{\alpha\alpha},$$

and making $a_{jk} = E\left(W^j Y^k\right)$ for $k = 0,1,2,3$ and $j = 0,1,2$, we obtain the expected information matrix, the elements of which are given in the appendix.

5. The alpha-power nonlinear regression model

A more general model can be defined replacing the linearity assumption by a nonlinear one. Therefore, we define the nonlinear alpha-power model as

$$y_i = f(\boldsymbol{\beta},\mathbf{x}_i) + \epsilon_i, \quad i = 1,2,\dots,n,$$

where y_i is the response variable, f is an injective continuous and twice differentiable function with respect to the parameter $\boldsymbol{\beta}$, $\mathbf{x}_i$ is an explanatory variable vector and ϵ_i are independent and identically distributed $P_F(0,\eta,\alpha)$ random variables with

$$\mu_\epsilon = \alpha\eta\int_0^1 z^{\alpha-1}F^{-1}(z)dz.$$

As in the linear case, $E(Y_i) = f(\boldsymbol{\beta},\mathbf{x}_i) + \mu_\epsilon$, so that corrections are required so that the error term is unbiased for zero, that is,

$$y_i \sim P_F(f(\boldsymbol{\beta},\mathbf{x}_i),\eta,\alpha).$$

In the PN situation we have the density function

$$\varphi(y_i;\boldsymbol{\beta},\alpha) = \frac{\alpha}{\eta}\phi\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)\left\{\Phi\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)\right\}^{\alpha-1}. \tag{12}$$

which we denote by $y_i|\mathbf{x}_i \sim \mathrm{PN}(f(\boldsymbol{\beta},\mathbf{x}_i),\eta,\alpha)$. The log-likelihood function (disregarding constants) for the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\mathsf{T},\eta,\alpha)^\mathsf{T}$ for a random sample of size n from y_i with distribution $\mathrm{PN}(f(\boldsymbol{\beta},\mathbf{x}_i),\eta,\alpha)$, is given by

$$\ell(\boldsymbol{\theta};\mathbf{X},\mathbf{y}) = n\log\left(\frac{\alpha}{\eta}\right) - \frac{1}{2\eta^2}\sum_{i=1}^n (y_i - f(\boldsymbol{\beta},\mathbf{x}_i))^2 + (\alpha-1)\sum_{i=1}^n \log\left\{\Phi\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)\right\}.$$

The score function $U(\boldsymbol{\theta}) = (U(\boldsymbol{\beta}), U(\eta), U(\alpha))^{\mathsf{T}}$ is given by

$$U(\beta_i) = \frac{1}{\eta^2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_i} - \frac{\alpha - 1}{\eta}\sum_{i=1}^{n} w_i \frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_i}, \quad U(\alpha) = n\left(\frac{1}{\alpha} + \overline{u}\right),$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta^3}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))^2 - \frac{\alpha - 1}{\eta^2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))w_i,$$

with $u_i = \Phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)$ and $w_i = \frac{\phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)}{\Phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)}$.

Differentiating the scores above, we arrive at the observed information matrix, see appendix. Hence, the maximum likelihood estimator for θ, can be obtained by implementing the following Newton-Raphson type iterative procedure:

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \hat{\boldsymbol{\theta}}(k) + [J(\hat{\boldsymbol{\theta}}^{(k)})]^{-1}U(\hat{\boldsymbol{\theta}}^{(k)}), \tag{13}$$

where $J(\boldsymbol{\theta}) = -\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathsf{T}}}$.

6. Nonlinear autoregressive alpha-power-normal model

We consider now the extension of the nonlinear-normal model with autoregressive errors to the PN distribution. Hence, the stochastic representation for the nonlinear PN model with autoregressive errors is given by

$$y_i = f(\boldsymbol{\beta}, \mathbf{x}_i) + \epsilon_i, \quad \text{with} \quad \epsilon_i = \rho\epsilon_{i-1} + a_i, \quad i = 1, 2, \cdots, n, \tag{14}$$

where y_i, $i = 1, \ldots, n$ are the observed responses, the $\mathbf{x}_i$, $i = 1, \ldots n$ are known covariate vectors with ρ as the autoregressive coefficient satisfying $|\rho| < 1$; $\boldsymbol{\beta}$ is an unknown p-dimensional vector of real parameters, f is a known continuous and twice differentiable function with respect to $\boldsymbol{\beta}$, a_i are independent random variables with $a_i \sim \text{PN}(0, \eta^2, \alpha)$ and $\epsilon_0 = 0$.

It then follows that the expectation of the random response is

$$E(Y_i) = f(\boldsymbol{\beta}, \mathbf{x}_i) + E(a_i)\sum_{k=0}^{i-1}\rho^k, \tag{15}$$

$i = 1, \ldots, n$, where $E(a_i)$ is the expectation of a random variable with $\mathrm{PN}(0, \eta^2, \alpha)$ distribution.

6.1. Maximum likelihood estimation

Given a random sample of size n from the above model, the log-likelihood function for parameter vector $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}^\top, \eta^2, \alpha)^\top$, can be written as

$$\ell_n(\boldsymbol{\theta}; \mathbf{y}) = n\left\{\log(\alpha) - \log(\eta) - \frac{1}{2}\log(2\pi)\right\} - \sum_{i=1}^{n} \frac{(\epsilon_i - \rho\epsilon_{i-1})^2}{2\eta^2} + (\alpha - 1)\sum_{i=1}^{n}\log\{\Phi(z_i)\},$$

with $z_i = \frac{\epsilon_i - \rho\epsilon_{i-1}}{\eta}$. Therefore, for $w_i = \frac{\phi(z_i)}{\Phi(z_i)}$, $D_i = -\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \boldsymbol{\beta}} + \rho\frac{\partial f(\boldsymbol{\beta}, x_{i-1})}{\partial \boldsymbol{\beta}}$ and $Q_i = -w_i(z_i + w_i)$, $i = 1, 2, \ldots, n$, the score function $U_{\boldsymbol{\theta}} = (U_\rho, U_{\boldsymbol{\beta}}^\top, U_{\eta^2}, U_\alpha)^\top$ has elements:

$$U(\rho) = \sum_{i=1}^{n}\left[\frac{a_i}{\eta^2} - \frac{\alpha - 1}{\eta}w_i\right]\epsilon_{i-1}, \qquad U(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[\frac{a_i}{\eta^2} - \frac{\alpha - 1}{\eta}w_i\right]D_i,$$

$$U(\eta^2) = \sum_{i=1}^{n}\left[-\frac{1}{2\eta^2} + \frac{a_i^4}{2\eta^4} - \frac{\alpha - 1}{2\eta^3}a_i w_i\right], \qquad U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n}\log\{\Phi(z_i)\},$$

where $a_i = \epsilon_i - \rho\epsilon_{i-1}$ and $\epsilon_i = y_i - f(\boldsymbol{\beta}, x_i)$. Hence, taking $G_i = -\frac{\partial^2 f(\boldsymbol{\beta}, x_i)}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}^\top} + \rho\frac{\partial^2 f(\boldsymbol{\beta}, x_{i-1})}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}^\top}$, we obtain the Hessian matrix, see Appendix, from which the expected information can be obtained.

Therefore, the maximum likelihood estimators can be obtained by iteratively solving the equation:

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \hat{\boldsymbol{\theta}}^{(k)} + [J(\hat{\boldsymbol{\theta}}^{(k)})]^{-1}U(\hat{\boldsymbol{\theta}}^{(k)}), \tag{16}$$

where $J(\boldsymbol{\theta}) = -\dfrac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$.

6.2. Score statistics for testing ρ

In the particular case where $\rho = 0$, the autoregressive model (14) reduces to the nonlinear PN regression model. Hence, it is important to verify whether this is the case or not. Considering $\boldsymbol{\beta}$, η^2 and α as nuisance parameters, we want to test the hypotheses

$$H_0 : \rho = 0 \;\text{ versus }\; H_1 : \rho \neq 0.$$

It can be shown that the score statistics (Cox and Hinkley, 1974) for testing H_0 is given by :

$$\mathrm{SC}_1 = [U_\rho^2 J^{\rho\rho}(\theta)]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}, \tag{17}$$

where $J^{\rho\rho}$ is the block of J^{-1} corresponding to ρ and $\hat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Under H_0, statistics (17) follows, asymptotically the chi-square distribution (χ_1^2) with one degree of freedom.

7. Simulation study

We report next results of a simulation study designed at investigating the performance of the maximum likelihood estimators for parameters β_0, β_1 and η_e. We simulated 1000 samples of sizes $n = 50$, 75 and 100. Without loss of generality we took $\eta_e = 1$. Values for X were generated from the $U(0,1)$, the uniform distribution on the $(0,1)$ interval and $p = 1$, with $\beta_0 = 1.5$ and $\beta_1 = -2.5$. Moreover, we took $\epsilon_i \sim \mathrm{PN}(0, \eta_e, \alpha)$. Estimators performance were evaluated by computing the relative empirical bias (RB = empirical bias/parameter value) and the square root of the empirical mean squared error ($\sqrt{\mathrm{MSE}}$) and the covering probability of the 95% large sample intervals (discussed above) or, equivalently, the rejection rate for testing $\beta_1 = 0$ at the 5% significance level. This study was implemented using software R.

Table 1: *Empirical RB and $\sqrt{MSE}$ for the simple PNR model.*

		$\alpha = 0.75$			$\alpha = 1.50$			$\alpha = 2.25$		
n	$\hat{\theta}$	RB(%)	$\sqrt{\mathrm{MSE}}$	$1-\delta$	RB(%)	$\sqrt{\mathrm{MSE}}$	$1-\delta$	RB(%)	$\sqrt{\mathrm{MSE}}$	$1-\delta$
	$\hat{\beta}_0$	7.96	1.25	0.66	4.06	1.26	0.70	5.93	1.32	0.74
	$\hat{\beta}_1$	0.21	0.16	0.77	0.13	0.13	0.79	0.11	0.11	0.83
50	$\hat{\eta}$	15.58	0.52	0.85	12.08	0.46	0.83	12.05	0.46	0.84
	$\hat{\alpha}$	58.81	1.36	0.62	84.57	3.22	0.67	101.80	5.51	0.70
	$\hat{\beta}_0$	5.07	1.11	0.66	2.19	1.12	0.76	3.89	1.12	0.82
	$\hat{\beta}_1$	0.16	0.13	0.83	0.12	0.11	0.87	0.10	0.09	0.90
75	$\hat{\eta}$	10.85	0.44	0.82	5.53	0.38	0.83	4.29	0.36	0.88
	$\hat{\alpha}$	46.53	1.17	0.65	76.44	2.86	0.72	92.73	4.84	0.79
	$\hat{\beta}_0$	1.10	0.47	0.80	1.01	0.51	0.92	2.64	0.53	0.93
	$\hat{\beta}_1$	0.06	0.04	0.94	0.02	0.04	0.94	0.04	0.03	0.95
500	$\hat{\eta}$	0.56	0.16	0.80	0.46	0.16	0.92	0.47	0.15	0.93
	$\hat{\alpha}$	12.08	0.47	0.78	16.09	1.13	0.88	21.51	1.88	0.88

Table 2: *Empirical RB and $\sqrt{MSE}$ for simple regression model with contaminated model.*

		$\alpha = 0.75$			$\alpha = 1.50$			$\alpha = 2.25$		
n	$\hat{\theta}$	RB(%)	$\sqrt{\text{MSE}}$	$1-\delta$	RB(%)	$\sqrt{\text{MSE}}$	$1-\delta$	RB(%)	$\sqrt{\text{MSE}}$	$1-\delta$
	$\hat{\beta}_0$	6.61	1.22	0.63	8.74	1.28	0.68	5.98	1.28	0.71
	$\hat{\beta}_1$	0.24	0.16	0.75	0.21	0.13	0.80	0.15	0.12	0.83
50	$\hat{\eta}$	15.12	0.51	0.85	13.69	0.48	0.81	11.06	0.44	0.84
	$\hat{\alpha}$	56.72	1.32	0.59	75.31	3.12	0.64	92.29	5.29	0.67
	$\hat{\beta}_0$	5.82	1.12	0.68	7.11	1.14	0.73	3.77	1.13	0.76
	$\hat{\beta}_1$	0.16	0.13	0.82	0.16	0.10	0.86	0.18	0.08	0.85
75	$\hat{\eta}$	10.94	0.44	0.83	9.47	0.40	0.82	7.49	0.37	0.85
	$\hat{\alpha}$	47.76	1.22	0.65	60.14	2.72	0.70	70.91	4.39	0.72
	$\hat{\beta}_0$	1.66	0.49	0.80	2.20	0.53	0.92	3.68	0.54	0.93
	$\hat{\beta}_1$	0.06	0.04	0.93	0.06	0.04	0.94	0.04	0.03	0.94
500	$\hat{\eta}$	0.46	0.16	0.80	0.13	0.15	0.92	0.54	0.15	0.94
	$\hat{\alpha}$	13.70	0.50	0.78	19.67	1.22	0.89	24.79	2.01	0.89

Results in Table 1 show that the relative bias and $\sqrt{\text{MSE}}$ for the maximum likelihood estimators for parameters η_e, β_0 and β_1 decreases as the sample sizes increase which is expected. It can also be noted that the relative bias can be large in small and moderate sample sizes situations. As parameter α increases, relative bias also increases for parameters η_e and $\hat{\beta}_0$ which is also expected. Relative bias for $\hat{\beta}_1$ is below 1.5%. To reduce bias for β_0 procedures such as bootstrap and jacknife could be implemented.

We also developed a simulation study designed at evaluating the robustness of the estimation procedure under the PN regression model obtained by contaminating the error terms with a skew-normal random variable. It was considered that the first observation was generated according to the distribution $\text{SN}(0,1,-1)+\text{PN}(0,1,\alpha)$. Maximum likelihood estimators were than computed for each generated sample, as described above and Table 1 presents the results. It can be deduced from the table that empirical RB and $\sqrt{\text{MSE}}$ does not seem affected by changes in the model generating the data.

8. Numerical illustrations

8.1. Linear model

The following illustration is based on the Australian athletes data set available for downloading at the directory http://azzalini.stat.unipd.it/SN/. The linear model considered is

$$Bfat_i = \beta_0 + \beta_1 Wt_i + \beta_2 sex_i + \epsilon_i, \ \ i = 1,2,\ldots,202,$$

where $Bfat_i$ is the body fat percentage for the i-th athlete, and covariates Wt_i and sex_i the weight and sex, respectively, for the i-th athlete; variable sex is dichotomized with 1 for male and zero for female. A residual analysis has indicated that symmetric models may not be the most adequate ones and that an asymmetric model can present a better fit, see Table 3, where quantities $\sqrt{b_1}$ and b_2 indicate sample asymmetry and kurtosis coefficients.

Table 3: *Summary statistics for estimated residuals under normality.*

n	Mean	Variance	$\sqrt{b_1}$	b_2
202	0.0050	11.8431	0.6030	3.9321

We fitted linear regression models under the assumption that model errors follow an asymmetric distribution, namely the the skew-normal (SNR), the skew-t$_\nu$ (StR), the student con ν degrees of freedom and power-normal (PN) distributions. For estimating under skew-normal and skew-Student-t R Development Core Team (2014) package is used, which uses the centred parametrization (CP), namely $E(Y) = \mathbf{x}^\top \boldsymbol{\beta}$ and $\mathrm{Var}(Y) = \eta^2$ (see Chiogna (2005) and Pewsey (2000)), whereas for model PN we use the optim program in the R package.

We use the AIC (Akaike, 1974), written as $\mathrm{AIC} = -2\hat{\ell}(\cdot) + 2k$ and BIC, written as $\mathrm{BIC} = -2\hat{\ell}(\cdot) + (\log(n))k$, where k is the number of unknown parameters, for comparing the normal and power-normal which are nested models. The best model is the one with the smallest AIC or BIC.

Moreover, the results in Table 4 present estimates for model parameters. It also reveals that, according to the PN regression model, % of body fat depends on weight and sex of the athlete. Estimating β_0^* in the PN regression model leads to $\hat{\beta}_0^* = 0.39$.

Table 4: *Estimates (standard error) for normal and PN linear models.*

Parameters	Normal model	SNC model	St_{14} model	PN model
β_0	1.62 (1.43)	2.91 (1.34)	−0.52 (1.35)	−5.97(2.00)
β_1	0.24 (0.02)	0.21 (0.02)	0.21 (0.02)	0.24 (0.02)
β_2	−12.25 (0.57)	−11.10 (0.71)	−11.09 (0.68)	−11.25 (0.60)
η	3.43 (0.17)	3.43 (0.18)	4.47 (0.75)	5.29 (0.48)
α		0.57 (0.14)	2.07 (0.50)	5.38 (1.83)

The model $\epsilon_i \sim \mathrm{PN}(0, 5.29, 5.38)$ seems to present a good fit for the data set under study. A more emphatic justification for using a PN type model comes from testing the normality assumption, that is, the hypotheses

$$H_0 : \alpha = 1 \;\text{ versus }\; H_1 : \alpha \neq 1,$$

by using the likelihood ratio statistics,

$$\Lambda = \frac{\ell_{\mathrm{N}}(\widehat{\boldsymbol{\theta}})}{\ell_{\mathrm{PN}}(\widehat{\boldsymbol{\theta}})},$$

which, for the data set under study, leads to $-2\log(\Lambda) = 4.97$, so that p-value $=$ $\mathrm{Prob}(\chi_1^2 > 4.97) < 0.05$. with strong indication against the null hypothesis.

Computing AIC and BIC for normal and PN regression models lead to AIC $=$ 1079.54 and BIC $=$ 1092.77 and AIC $=$ 1076.56 and BIC $=$ 1093.10, respectively. According to the values obtained for AIC and BIC, the power-normal (PN) linear regression model presents the better fit when compared with normal linear model.

We use Voung (1989) approach (generalized LR statistic) for comparing the skew-normal (SNR), skew-Student-t (StR) and power-normal (PNR) linear non-nested models fitted to the data. A description of the procedure is described next. Being F_θ and G_ζ two non-nested models and $f(y_i|\mathbf{x}_i,\boldsymbol{\theta})$ and $g(y_i|\mathbf{x}_i,\boldsymbol{\zeta})$ the corresponding densities, the likelihood ratio statistics to compare both models is given by

$$\mathrm{LR}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\zeta}}) = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \log \frac{f(y_i|\mathbf{x}_i,\widehat{\boldsymbol{\theta}})}{g(y_i|\mathbf{x}_i,\widehat{\boldsymbol{\zeta}})} \right\},$$

which does not follow a chi-square distribution. To overcome this problem, Vuong (1989) proposed an alternative approach based on the Kullback-Liebler divergence criterion. Based on the divergence between each model and the true process generating the data, namely the model $h^0(y|x)$, one arrives at the statistics

$$T_{\mathrm{LR,NN}} = \frac{1}{\sqrt{n}} \frac{\mathrm{LR}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\zeta}})}{\widehat{w}}, \tag{18}$$

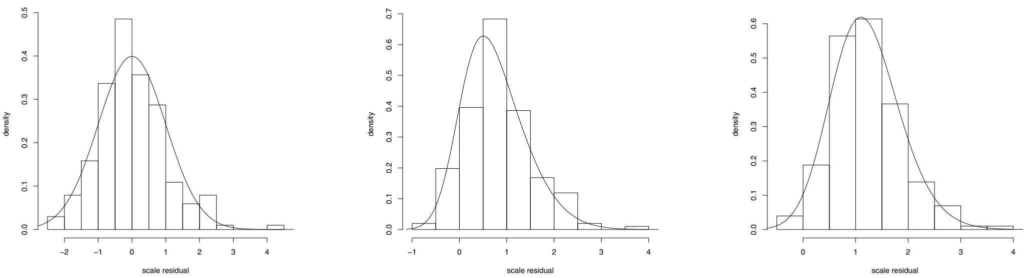
where

$$\widehat{w}^2 = \frac{1}{n}\sum_{i=1}^{n} \left(\log \frac{f(y_i|\mathbf{x}_i,\widehat{\boldsymbol{\theta}})}{g(y_i|\mathbf{x}_i,\widehat{\boldsymbol{\zeta}})} \right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} \log \frac{f(y_i|\mathbf{x}_i,\widehat{\boldsymbol{\theta}})}{g(y_i|\mathbf{x}_i,\widehat{\boldsymbol{\zeta}})} \right)^2.$$

For strictly non-nested models, it can be shown that the statistic $T_{\mathrm{LR,NN}}$ converges in distribution to a standard normal distribution under the null hypothesis. Thus, the null hypothesis is not rejected if $|T_{\mathrm{LR,NN}}| \leq z_{p/2}$. On the other hand, we reject at significance level p the null hypothesis of equivalence of the models in favor of model F_θ being better (or worse) than model G_ζ if $T_{\mathrm{LR,NN}} > z_p$ (or $T_{\mathrm{LR,NN}} < -z_p$).

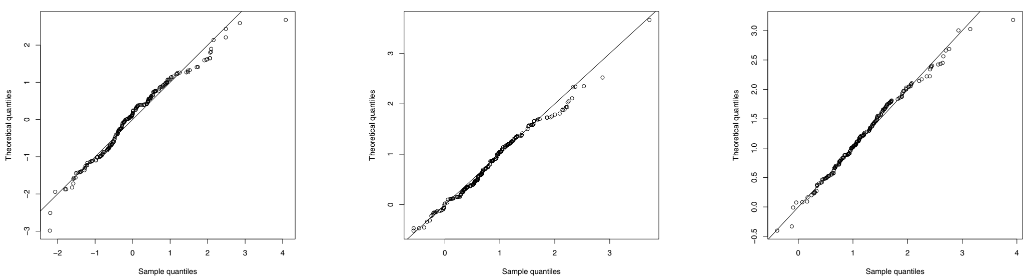
For testing PNR versus SNR, we obtain $T_{\text{LR,NN}} = 22.59$ (p-value < 0.05) and for the PNR versus RSt_{14} model, $T_{\text{LR,NN}} = 0.61$ (p-value > 0.05). Therefore, the PNR model is significantly better than the SNR model according to the generalized LR statistic. In a similar fashion it can be concludes that there is no significant difference between models PNR and RSt_{14}. However, favouring model PNR we have the fact that it involves one less parameter. Authors Lange et al. (1989), Berkane et al. (1994), Fernández and Steel (1999), Taylor and Verbyla (2004) and Leiva et al. (2008), all reported dificulties in estimating the degrees of freedom parameter.

We also computed the scaled residuals $e_i = (y_i - \mathbf{x}_i^{\top}\hat{\boldsymbol{\beta}})/\hat{\eta}$ to investigate model fit. Figures 2-(a), (b) and (c) and 3-(a), (b) and (c) depict the histograms and Q-Q plots for the scaled residuals under normal, SNR and PNR models, which also indicate a good fit for the PNR model.



(a) (b) (c)

Figure 2: *Graphs for residuals, of the fitted models. (a) Normal, (b) SN and (c) PN.*



(a) (b) (c)

Figure 3: *Q-qplots for the scaled residuals Z, from the fitted models. (a) Normal, (b) SN and (c) PN.*

8.2. Nonlinear model with correlated errors

In the following we present an application of the PN model fitting to the palm oil data set presented in Foong (1999) and studied in Xie et al. (2009) using a skew-normal nonlinear model. This data set was previously analysed in Azme et al. (2005), were parameter estimates are obtained under nonlinear growth curve models using Marquardat's iterative procedure. They found that the best fit is presented by the logistic growth curve model (see, Ratkowsky, 1983), followed by the Gompertz model, which was followed by the Morgan-Mercer-Flodin, Chapman-Richard model. Cancho et al. (2008) also analysed the model using a nonlinear skew-normal model with logistic growth. We focus now on analyzing the data set under a PN nonlinear regression model with logistic growth. Therefore, the model considered can be written as

$$y_i = \frac{\beta_1}{1+\beta_2 \exp(-\beta_3 x_i)} + \epsilon_i \tag{19}$$

with $\epsilon_i = \rho\epsilon_{i-1} + a_i$, $a_i \sim PN(0, \eta^2, \alpha)$, $i = 1, \ldots, n$.

We are now implementing the correlated nonlinear normal model with normally distributed errors (NLCM) and the correlated nonlinear model with errors PN distributed (NLCPN). As Table 5 reveals, according to both criteria (AIC and BIC), the nonlinear PN model with correlated errors fits the data better.

Table 5: *AIC and BIC for the oil palm data.*

Statistics	Log-likelihood	AIC	BIC
Normal	−41.2656	92.5312	97.2534
PN	−39.1004	90.2008	95.8674

Table 6: *Parameter estimates (standard errors) for the following models: NLCN, NLCPN and NLPN.*

	NLCN	NLCPN	NLPN
Parameter	estimate	estimate	estimate
ρ	0.3222(0.2757)	0.2574(0.2114)	—
β_1	37.5699(0.3038)	37.9163(0.4041)	38.8798(0.2485)
β_2	11.4310(0.8327)	17.5880(1.2504)	17.5833(1.7888)
β_3	0.5092(0.0227)	0.6140(0.0135)	0.6079(0.0172)
η^2	5.5559(0.7392)	2.6815(0.3658)	1.2010(0.1550)
α	—	0.7010(0.1564)	0.2547(0.0589)

We consider now testing the hypotheses $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$, that is, a nonlinear normal model with correlated errors against a nonlinear PN model with correlated errors. The likelihood ratio statistics for testing the above hypotheses, namely, $\Lambda = \frac{\ell_{\text{NLCN}}(\hat{\boldsymbol{\theta}})}{\ell_{\text{NLCPN}}(\hat{\boldsymbol{\theta}})}$, leads to $-2\log(\Lambda) = 4.3304$, a value greater than the corresponding 5%

chi-square critical values which is $\chi^2_{1,5\%} = 3.8414$. Hence there is strong evidence that the nonlinear PN model with correlated errors fit the oil palm data set far better than the corresponding normal one.

Parameter estimates and standard errors for models NLPN, NLCN and NLCPN are presented in Table 6.

Figure 4(a), presents the nonlinear fitted models graphs and (b), and the fitted residuals for model PN, $\hat{r}_i$ against $\hat{r}_{i-1} = \hat{r}(1)$, under the assumption that $\rho = 0$; which does not reveal presence of correlation. Therefore, we implement a nonlinear model with errors PN$(0, \eta, \alpha)$, (NLPN) for which parameter estimates are given in Table 6.

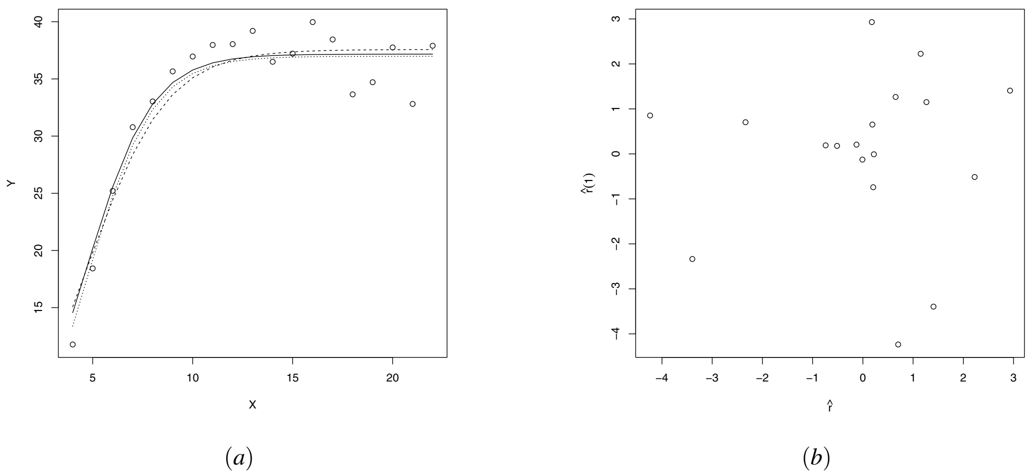


Figure 4: (a) Graphs for fitted models, NLCN (dashed line), NLCPN (solid line) and NLPN (mixed (dashed-dotted) line); (b) graph for $\hat{r}_i$ against $\hat{r}_{i-1}$.

9. Final discussion

In this paper we extended the power models in Pewsey et al. (2012) for the case of regression models. Linear models were considered as well as a non-linear extension. Emphasis was placed on the PN regression model situation. Estimation was performed by implementing the maximum likelihood approach. Large sample point and interval estimates were obtained by using the observed information (minus the inverse of the Hessian matrix evaluated at the maximum likelihood estimates). The exact Fisher information matrix is also derived and shown to be non-singular, so that large sample distribution for the alternative likelihood ratio statistics is central chisquare. For some comparisons, models are not nested so that an appropriate statistics with limiting normal distribution is considered.

The methodology implemented presented satisfactory results when applied to real data sets. Results of a small scale simulation indicate that the estimation approach leads to good parameter recovery and that for large sample sizes bias and mean square error are significantly reduced. One of the applications is to a linear model applied to the Australian athletes data set (available for downloading from the internet) previously analysed by several other authors. It was seen that data present moderate to large skewness so that the PN regression model can be a viable alternative. The second data set that was analysed is the palm oil data set previously analysed by several authors. It turned out that the non-linear model with PN errors fitted the data better than the ordinary normal model.

10. Appendix

In this section we present in closed form the elements of the observed and expected (Fisher) information matrices for the PNR type models considered in this. Their derivation (requiring extensive algebraic manipulations) extends results in Pewsey et al. (2012). The relevance of the results rely on the fact one can conclude they are nonsingular so that large sample properties of the maximum likelihood estimators hold for such models. A similar discussion for skew-normal type models is considered in Azzalini (2013).

10.1. Observed information matrix for the PNR model

In this section we present the observed information matrix for the general PNR model.

$$j_{\boldsymbol{\beta}^\top \boldsymbol{\beta}} = \frac{1}{\eta_e^2}\mathbf{X}^\top\mathbf{X} + \frac{\alpha-1}{\eta_e^2}\mathbf{X}^\top\Lambda_2\mathbf{X}, \qquad j_{\eta_e\boldsymbol{\beta}} = \frac{2}{\eta_e^3}\mathbf{X}^\top(y-\mathbf{X}\boldsymbol{\beta}) + \frac{\alpha-1}{\eta_e^2}\mathbf{X}^\top\Lambda_3,$$

$$\begin{aligned} j_{\eta_e\eta_e} = &-\frac{n}{\eta_e^2} + \frac{3}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)^2 - \frac{2(\alpha-1)}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)w_i \\ &+ \frac{\alpha-1}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)^3 w_i + \frac{\alpha-1}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)^2 w_i^2 \end{aligned}$$

$$j_{\alpha\boldsymbol{\beta}} = \frac{1}{\eta_e}\mathbf{X}^\top\Lambda_1, \qquad j_{\alpha\eta_e} = \frac{1}{\eta_e^2}(y-\mathbf{X}\boldsymbol{\beta})^\top\Lambda_1, \qquad j_{\alpha\alpha} = n/\alpha^2,$$

where

$$\Lambda_2 = \operatorname{diag}\left\{\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\eta_e}\right)w_i + w_i^2\right\}_{i=1,2,\ldots,n}$$

and $\Lambda_3 = (a_1, a_2, \ldots, a_n)^\top$ with

$$a_i = \left\{ \left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\eta_e} \right)^2 w_i + \left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\eta_e} \right) w_i^2 - w_i \right\}_{i=1,2,\ldots,n}.$$

10.2. Information matrix for the simple PNR model

The elements of the FIM for the case $p = 1$ are given by

$$i_{\beta_0\beta_0} = \left\{ 1 + \frac{\alpha-1}{\eta_e} [a_{11} - a_{10}(\beta_0 + \beta_1 \overline{x})] + (\alpha-1)a_{20} \right\} / \eta_e^2,$$

$$i_{\beta_1\beta_0} = \left\{ \overline{x} + \frac{\alpha-1}{\eta_e} \left[\overline{x}(a_{11} - \beta_0 a_{10}) - \beta_1 a_{10} \overline{x^2} \right] + (\alpha-1)a_{20}\overline{x} \right\} / \eta_e^2,$$

$$\begin{aligned} i_{\eta_e\beta_0} = {} & \frac{1-\alpha}{\eta_e^2} a_{10} + \frac{1}{\eta_e^3} [2a_{01} + (\alpha-1)a_{21} - (2 + (\alpha-1)a_{20})(\beta_0 + \beta_1\overline{x})] + \\ & + \frac{\alpha-1}{\eta_e^4} \left\{ a_{12} + a_{10}(\beta_0^2 + \beta_1^2 \overline{x^2} + 2\beta_0\beta_1\overline{x}) - 2a_{11}(\beta_0 + \beta_1\overline{x}) \right\}, \end{aligned}$$

$$i_{\beta_1\beta_1} = \left\{ \overline{x^2}(1 + (\alpha-1)a_{20}) + \frac{\alpha-1}{\eta_e} \left[a_{11}\overline{x^2} - a_{10}(\beta_0\overline{x^2} + \beta_1\overline{x^3}) \right] \right\} / \eta_e^2,$$

$$\begin{aligned} i_{\eta_e\beta_1} = {} & \frac{1-\alpha}{\eta_e^2} a_{10}\overline{x} + \frac{1}{\eta_e^2} \left[\overline{x}(2(a_{01} - \beta_0) + (\alpha-1)(a_{21} - \beta_0 a_{20})) - \beta_1(2 + (\alpha-1)a_{20})\overline{x^2} \right] \\ & + \frac{\alpha-1}{\eta_e^4} \left[a_{12}\overline{x} + a_{10}(\beta_0^2\overline{x} + 2\beta_0\beta_1\overline{x^2} + \beta_1^2\overline{x^3}) - 2a_{11}(\beta_0\overline{x} + \beta_1\overline{x^2}) \right], \end{aligned}$$

$$\begin{aligned} i_{\eta_e\eta_e} = {} & -\frac{1}{\eta_e^2} + \frac{1}{\eta_e^4} [3a_{02} + (\alpha-1)a_{22} - 2(\beta_0 + \beta_1\overline{x})(3a_{01} + (\alpha-1)a_{21})] \\ & + \frac{1}{\eta_e^4}(3 + (\alpha-1)a_{20})(\beta_0^2 + 2\beta_0\beta_1\overline{x} + \beta_1^2\overline{x^2}) - 2\frac{\alpha-1}{\eta_e^3}(a_{11} - a_{10}(\beta_0 + \beta_1\overline{x})) \\ & + \frac{\alpha-1}{\eta_e^5} \left[a_{13} - 3a_{12}(\beta_0 + \beta_1\overline{x}) + 3a_{11}(\beta_0^2 + 2\beta_0\beta_1\overline{x} + \beta_1^2\overline{x^2}) \right] \\ & - \frac{\alpha-1}{\eta_e^5} a_{10}(\beta_0^3 + \beta_1^3\overline{x^3} + 3\beta_0\beta_1\overline{x^2} + 3\beta_0^2\beta_1\overline{x}), \end{aligned}$$

$$i_{\alpha\beta_0} = a_{10}/\eta_e,\ \ i_{\alpha\beta_1} = a_{10}\overline{x}/\eta_e,\ \ i_{\alpha\eta_e} = [a_{11} - a_{10}(\beta_0 + \beta_1\overline{x})]/\eta_e^2 \ \text{ and } \ i_{\alpha\alpha} = 1/\alpha^2.$$

The above expressions can be computed numerically.

10.3. Observed information matrix for the nonlinear PNR model

The elements of the observed information matrix for the nonlinear PNR model are given by

$$j_{\beta_k\beta_j} = \frac{1}{\eta^2}\sum_{i=1}^{n}[(y_i - f(\boldsymbol{\beta},\mathbf{x}_i))w_i + w_i^2 + \alpha - 1]\frac{\partial f(\boldsymbol{\beta},\mathbf{x}_i)}{\partial \beta_k}\frac{\partial f(\boldsymbol{\beta},\mathbf{x}_i)}{\partial \beta_k} +$$

$$\frac{1}{\eta^2}\sum_{i=1}^{n}[-(y_i - f(\boldsymbol{\beta},\mathbf{x}_i)) + \eta(\alpha-1)w_i]\frac{\partial^2 f(\boldsymbol{\beta},\mathbf{x}_i)}{\partial \beta_k \partial \beta_j},$$

$$j_{\eta\beta_j} = \frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\left\{-w_i + \frac{(y_i - f(\boldsymbol{\beta},\mathbf{x}_i))}{\eta}[(y_i - f(\boldsymbol{\beta},\mathbf{x}_i))w_i + w_i^2]\frac{\partial f(\boldsymbol{\beta},\mathbf{x}_i)}{\partial \beta_j}\right\} +$$

$$\frac{2}{\eta^3}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta},\mathbf{x}_i))\frac{\partial f(\boldsymbol{\beta},\mathbf{x}_i)}{\partial \beta_j},$$

$$j_{\alpha\beta_j} = \frac{1}{\eta}\sum_{i=1}^{n} w_i \frac{\partial f(\boldsymbol{\beta},\mathbf{x}_i)}{\partial \beta_j}, \qquad j_{\alpha\eta} = \frac{1}{\eta}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)w_i, \qquad j_{\alpha\alpha} = n/\alpha^2,$$

$$j_{\eta\eta} = -\frac{n}{\eta^2} + \frac{3}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)^2 - \frac{2(\alpha-1)}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)w_i$$

$$+\frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)^3 w_i + \frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta},\mathbf{x}_i)}{\eta}\right)^2 w_i^2.$$

10.4. Hessian matrix for the nonlinear PNR model with correlated errors

For the case of the nonlinear model with correlated errors, we have the following elements for the Hessian matrix:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho^2} = \frac{1}{\eta^2}\sum_{i=1}^{n}[-1 + (\alpha-1)Q_i]\,\epsilon_{i-1}^2,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^\top \partial \rho} = \sum_{i=1}^{n} \left[[1-(\alpha-1)Q_i] \frac{\epsilon_{i-1}}{\eta^2} D_i^\top - \left[\frac{a_i}{\eta^2} - \frac{1}{\eta} w_i \right] \frac{\partial f(\boldsymbol{\beta}, \mathbf{x}_{i-1})}{\partial \boldsymbol{\beta}^\top} \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \eta^2 \partial \rho} = \sum_{i=1}^{n} \left[-\frac{a_i}{\eta^4} + \frac{\alpha-1}{2\eta^2} \left[\frac{a_i Q_i}{\eta^2} + \frac{w_i}{\eta} \right] \right] \epsilon_{i-1}, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \rho} = -\frac{1}{\eta} \sum_{i=1}^{n} w_i \epsilon_{i-1},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \, \partial \boldsymbol{\beta}^\top} = \sum_{i=1}^{n} \left[\frac{1}{\eta^2} [-1+(\alpha-1)Q_i] D_i D_i^\top - \frac{a_i}{\eta^2} G_i + \frac{\alpha-1}{\eta} w_i G_i \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \eta^2 \partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left[\frac{a_i}{\eta^4} D_i - \frac{\alpha-1}{2\eta^2} \left[\frac{a_i}{\eta^2} Q_i D_i + \frac{1}{\eta} w_i D_i \right] \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \boldsymbol{\beta}} = \frac{1}{\eta} \sum_{i=1}^{n} w_i D_i, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \eta^2} = -\frac{1}{2\eta^3} \sum_{i=1}^{n} a_i w_i, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha^2} = -\frac{n}{\alpha^2}.$$

Acknowledgments

We thank two referees for comments and suggestions that substantially improved the presentation. The research of H. Bolfarine was supported by CNPq-Brasil. The research of H. W. Gómez was supported by FONDECYT-Chile 1130495.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Azme, K., Ismail, Z., Haron, K. and Ahmad, T.M. (2005). Nonlinear growth models for modeling oil palm yield growth. *Journal of Mathematical Statistics*, I(3), 225–233.

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46, 199–208.

Azzalini, A. (2013). *Skew Normal and Related Families*. Cambridge University Press.

Berkane, M., Kano, Y. and Bentle, P. M. (1994). Pseudo maximum likelihood estimation in elliptical theory: effects of misspecification. *Computational Statistics and Data Analysis*, 18, 255–267.

Cancho, V. G., Lachos, V. H. and Ortega, E. M. (2008). A nonlinear regression model with skew-normal errors. *Statistical Papers*, 51, 547–558.

Castillo, E. and Hadi, A. S. (1995). A method for estimating parameters and quantiles of distributions of continuous random variables. *Computational Statistics and Data Analysis*, 20, 421–439.

Chiogna, M. (2005). Notes on estimation problems with scalar skew-normal distributions. *Statistical Methods and Applications*, 14, 331–341.

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

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical statistics*. Chapman and Hall, London.

DiCiccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, 99, 439–450.

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

Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31, 497–512.

Fernández, C. and Steel, M. (1999). Multivariate Student-t regression models: pitfalls and inference. *Biometrika*, 86, 153–167.

Foong, F. S. (1999). Impact of mixture on potential evapotranspiration, growth and yield of palm oil. Pro 1999, *PORIM International Palm Oil Congress (Agric.)*. 265–287.

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

Gómez, H. W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18, 395–407.

Gupta, A. K. and Nadarajah, S. (2004). On the moments of the beta normal distribution. *Communications in Statistics-Theory and Methods*, 33, 1–13.

Gupta, D. and Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, 17, 197–210.

Hutson, A. D. (2004). Utilizing the flexibility of the epsilon-skew-normal distribution for common regression problems. *Journal of Applied Statistics*, 31, 673–683.

Lange, K. L., Little, J. A. and Taylor, M. G. J. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84, 881–896.

Lehmann, E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics*, 24, 23–43.

Leiva, V., Riquelme, M., Balakrishnan, N. and Sanhueza, A. (2008). Lifetime analysis based on the generalized Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis*, 52, 2079–2097.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, 83, 291–309.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 859–870.

Pewsey, A., Gómez, H. W. and Bolfarine, H. (2012). Likelihood-based inference for power distributions. *Test*, 21, 775–789.

R Development Core Team (2014). A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Razzaghi, M. (2009). Beta-normal distribution in dose-response modeling and risk assessment for quantitative responses. *Enviromental and Ecological Statistics*, 16, 25–36.

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

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

Taylor, J. and Verbyla, A. (2004). Joint modeling of location and scale parameters of t distribution. *Statistical Modelling*, 4, 91–112.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.

Xie, F. C., Lin, J. G. and Wei, B.C. (2009). Diagnostics for skew nonlinear regression models with AR(1) errors. *Computational Statistics and Data Analysis*, 53, 4403–4416.