REVSTAT – Statistical Journal
Volume 22, Number 4, October 2024, 433–454
https://doi.org/10.57805/revstat.v22i4.500

Estimation and Diagnostic for a Partially Linear Regression Based on an Extension of the Rice Distribution

Authors: J. C. S. Vasconcelos
– Exact Sciences Department, University of São Paulo, Brazil
juliocezarvasconcelos@hotmail.com

E. M. M. Ortega ✉
– Exact Sciences Department, University of São Paulo, Brazil
edwin@usp.br

G. M. Cordeiro
– Statistics Department, Federal University of Pernambuco, Brazil
gausscordeiro@gmail.com

J. S. Vasconcelos
– Department of Rural Engineering, Paulista State University, Brazil
julianojsv@gmail.com

M. A. M. Biaggioni
– Department of Rural Engineering, Paulista State University, Brazil
m.biaggioni@unesp.br

Received: December 2020 Revised: November 2022 Accepted: November 2022

Abstract:

- We introduce an extension of the Rice distribution and estimate its parameters by maximum likelihood. We define two regressions based on this extended distribution to model volumetric shrinkage of the wood and milk production. The performance of the parameter estimators is investigated in finite samples using Monte Carlo simulations. Also, we propose the quantile residuals for the regression models whose empirical distribution is close to normality. The usefulness of the new regressions is proved empirically through two applications to agricultural data.

Keywords:

- *cubic splines; milk production; regression extensions; simulation study; wood data.*

AMS Subject Classification:

- 62.

✉ Corresponding author

1. INTRODUCTION

The Rice distribution [17] is generally observed when the global magnitude of a vector is related to its direction components, such as when wind speed is analyzed in two directions, i.e., a two-dimensional component vector. If the components are independent and normally distributed with equal variances, the general wind speed has a Rice distribution. It is also used to model dispersion (or variability) of line-of-sight transmission between two stations which applies to FM radio waves, microwaves, magnetic resonance images in the presence of noise and satellite transmissions. It is also employed to model Rician fading, which describes how the cancellation of signals affects the propagation of radio waves. In recent years, it has been utilized by various authors for different applications. For example, [8] introduced a generalized Rice distribution based on the linearity characteristics of a system to model situations where the maximum amplitude is close to the signal's mean amplitude; [12] presented a Bayesian approach to estimate its parameters, and [26] studied a new approach to analyze images based on the maximum likelihood method that permits obtaining simultaneous estimates of the image and signal noise.

The Rice distribution is relatively unknown in the area of applied statistics. One of the objectives of this paper is to generalize the Rice distribution to be applied in different research areas. We emphasize that the papers mentioned previously do not provide regressions which have been widely employed in many fields. A fundamental conjecture that should be examined with caution regarding a data set is that when the covariates express nonlinear effects on the response variable and adopting a parametric regression may not be a suitable alternative. To overcome this circumstance, generalized semiparametric models have been proposed. For example, [7] and [9] introduced the generalized additive model (GAM) which aggregates the properties of generalized linear models with additive models; and [18] demonstrated that nonparametric regression can be considered an interesting extension of the parametric regression, and the two can be combined to produce the semiparametric regression. Another model widely applied in recent years is the generalized additive model for location, scale, and shape (GAMLSS) [19]. Various authors have published papers involving partially linear regressions. [22] introduced the symmetric generalized partial linear model; [24] presented an extension of the log-normal distribution from two perspectives, one of which was the partially linear regression; and [11] proposed a new strategy to select Bayesian models and an efficient estimation method for partially linear model.

Based on these contexts, the objectives of this paper are described below. We define the *odd log-logistic Rice* (OLLRc) distribution that can be applied to model bimodal, trimodal and asymmetric data. Based on this distribution, we introduce a parametric regression with two systematic components and illustrate its flexibility using volumetric shrinkage wood data, see for example [23]. We propose a new partially linear regression and show its utility by analyzing milk production in the Northeast of Brazil. We prove that the empirical distribution of the quantile residuals (qrs) for both regressions has approximately the standard normal distribution. We provide two applications to real data (shrinkage volume of three wood species and milk production data) to illustrate the flexibility of the OLLRc partially linear regression model.

The remaining sections are as follows. Section 2 defines the OLLRc distribution, and provides some mathematical properties. Section 3 introduces a parametric regression based on the new distribution. Section 4 proposes the OLLRc partially linear regression and performs some simulations for the distribution of the penalized maximum likelihood estimators. Simulation results for the residuals are reported in Section 5. The usefulness of the new regressions is proved by means of two real data sets in Section 6. Section 7 ends with some concluding remarks.

2. THE ODD LOG-LOGISTIC RICE DISTRIBUTION

If G denotes a baseline distribution, the cumulative distribution function (cdf) of the the odd log-logistic-G (OLL-G) (OLL-G) generator [5] is defined by

$$F(y) = \int_0^{\frac{G(y)}{G(y)}} \frac{\nu\, u^{\nu-1}}{(1+u^\nu)^2} du = \frac{G(y)^\nu}{G(y)^\nu + [1-G(y)]^\nu}, \quad y > 0, \tag{2.1}$$

where $\nu > 0$ is the shape parameter

This class of generalized distributions has been deeply investigated in the last years; see, for example, the references in [16] and [24]. The probability density function (pdf) corresponding to (2.1) can be expressed as

$$f(y) = \frac{\nu\, g(y)\{G(y)[1-G(y)]\}^{\nu-1}}{\{G(y)^\nu + [1-G(y)]^\nu\}^2}, \tag{2.2}$$

where $g(y) = d\,G(y)/d\,y$ is the baseline density.

The Rice cdf with two parameters $\mu > 0$ and $\sigma > 0$ is (for $y > 0$)

$$G_{\mu,\sigma}(y) = 1 - Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right), \tag{2.3}$$

where $Q_1(a,b)$ is the Marcum Q-function, namely

$$Q_M(a,b) = \int_b^\infty x\left(\frac{x}{a}\right)^{M-1} \exp\left(-\frac{x^2+a^2}{2}\right) I_{M-1}(ax)dx,$$

I_{M-1} is the modified Bessel function of the first kind of order $M-1$ (for $\eta \in \mathbb{R}, \eta \neq 0$),

$$I_\eta(z) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\,\Gamma(m+\eta+1)} \left(\frac{z}{\eta}\right)^{2m+\eta},$$

and $\Gamma(\cdot)$ is the gamma function. The Marcum Q-function is defined in the `VGAM` package of **R** software. For more details, see [27].

The pdf corresponding to (2.3) has the form

$$g_{\mu,\sigma}(y) = \frac{y}{\mu^2} \exp\left(-\frac{y^2+\sigma^2}{2\mu^2}\right) I_0\left(\frac{\sigma y}{\mu^2}\right), \tag{2.4}$$

where $I_0(z) = \sum_{m=0}^{\infty} z^{2m}/[4^m\,(m!)^2]$.

The Rice distribution can be obtained following a simple extension of the Rayleigh distribution. Let $X = \sqrt{T_1^2 + T_2^2}$, where $T_1 \sim \mathrm{N}(\delta_1, \mu^2)$ and $T_2 \sim \mathrm{N}(\delta_2, \mu^2)$ are independent random variables. Then, X has the Rice density (2.4), where $\sigma = \sqrt{\delta_1^2 + \delta_2^2}$. If $\sigma = 0$, then (2.4) is just the Rayleigh density. So, the parameter μ in the Rice distribution is the standard deviation of two Gaussian contributions and σ represents a distance term. The Rice distributions tends to the $\mathrm{N}(\sigma, \mu^2)$ distribution if $\sigma y/\mu^2$ goes to ∞.

The OLLRc cdf (for $y > 0$) is defined by taking $G(x)$ in (2.1) as the Rice cdf (2.3)

$$F(y) = \frac{\left[1 - Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)\right]^{\nu}}{\left[1 - Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)\right]^{\nu} + Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)^{\nu}} \tag{2.5}$$

The OLLRc density function follows by inserting (2.3) and (2.4) in Equation (2.2)

$$\begin{aligned} f(y) &= \frac{\nu\, y}{\mu^2} \exp\left\{\frac{-(y^2 + \sigma^2)}{2\mu^2}\right\} I_0\left(\frac{y\,\sigma}{\mu^2}\right) \times \left\{\left[1 - Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)\right] Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)\right\}^{\nu-1} \\ &\quad \times \left\{\left[1 - Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)\right]^{\nu} + Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu}\right)^{\nu}\right\}^{-2}, \end{aligned} \tag{2.6}$$

where all parameters are positive. The OLLRc density function can be expressed as a combination of exponentiated Rice densities (see Appendix A).

Henceforth, let $Y \sim \mathrm{OLLRc}(\mu, \sigma, \nu)$ be a random variable with density (2.6). Some shapes of (2.6) reported in Figure 1 reveal that the density of Y is very flexible for bimodal and trimodal data.

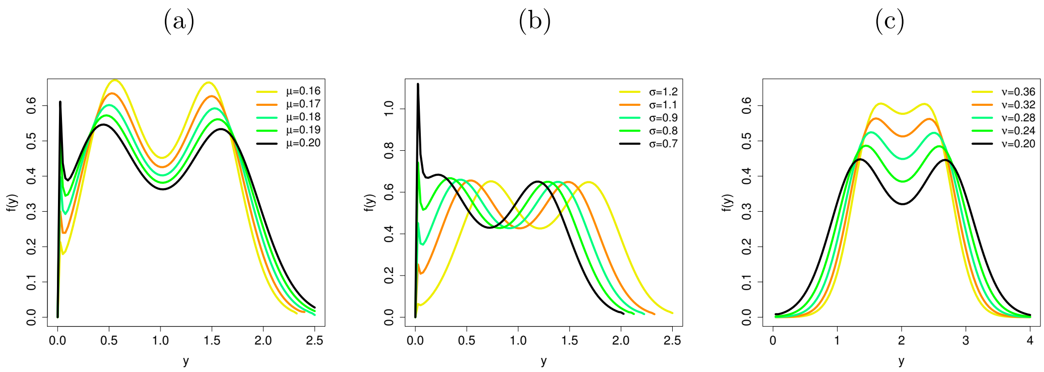

Figure 1: Plots of the OLLRc density. (a) For $\sigma = 1$ and $\nu = 0.18$ varying μ. (b) For $\mu = 0.16$ and $\nu = 0.17$ varying σ. (c) For $\mu = 0.25$ and $\sigma = 2$ varying ν.

By inverting Equation (2.5), the quantile function (qf) of Y, say $y = H(u) = F^{-1}(u)$, is

$$y = H(u) = H_{\mathrm{Rice}}\left\{\frac{u^{1/\nu}}{u^{1/\nu} + (1-u)^{1/\nu}}\right\}, \qquad u \in (0,1), \tag{2.7}$$

where $H_{\mathrm{Rice}}(u) = G^{-1}_{\mu,\sigma}(u)$ is the Rice qf.

Figure 2 displays plots of the density of Y and histograms from two simulated data sets with 100,000 replications. They show that the simulated values are consistent with the OLLRc distribution, where we note trimodal and bimodal shapes.

(a) (b) (c)

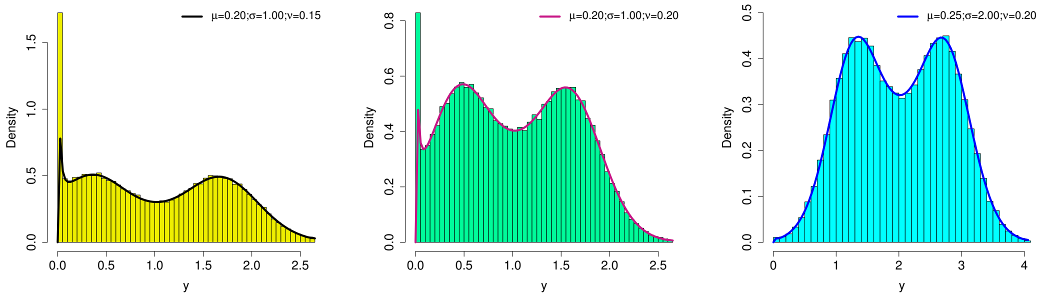

Figure 2: Histograms and plots of the OLLRc density.

The influence of the shape parameter ν on the skewness and kurtosis of Y can be easily investigated based on quantile measures. Figure 3(a) displays the Bowley skewness

$$B = \frac{H(3/4) + H(1/4) - 2\,H(2/4)}{H(3/4) - H(1/4)},$$

whereas Figure 3(b) provides the Moors kurtosis

$$M = \frac{H(3/8) - H(1/8) + H(7/8) - H(5/8)}{H(6/8) - H(2/8)}.$$

(a) (b)

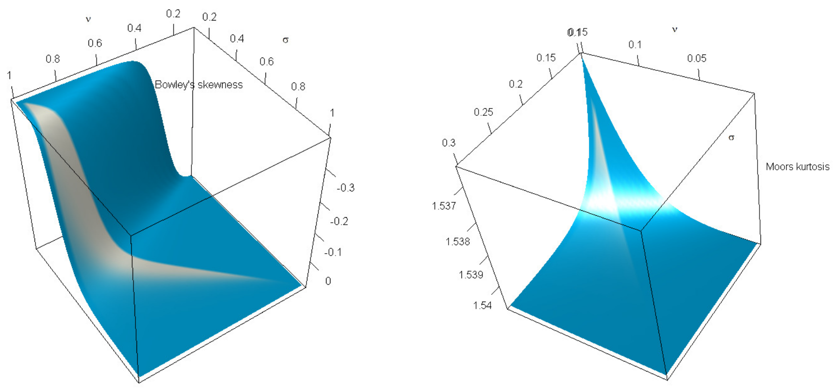

Figure 3: (a) Bowley's skewness. (b) Moors kurtosis.

By varying $\nu \in [0.1, 1]$, Figure 3(a) displays the Bowley skewness of Y for $\mu = 0.1$ and $\sigma \in [0.1, 1]$, whereas Figure 3(b) reports the Moors kurtosis of Y for $\mu = 0.1$ and $\sigma \in [0.1, 0.3]$.

3. THE OLLRC REGRESSION

The OLLRc regression is defined by two systematic components considering that the parameters μ_i and σ_i in the density (2.6) are given by (for $i = 1, ..., n$)

$$\mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1) \qquad \text{and} \qquad \sigma_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2), \tag{3.1}$$

where $\boldsymbol{\beta}_1 = (\beta_{11}, \cdots, \beta_{1p})^\top$ and $\beta_2 = (\beta_{21}, \cdots, \beta_{2p})^\top$ are vectors of unknown coefficients and $\mathbf{x}_i^\top = (x_{i1}, \cdots, x_{ip})$ is a vector of covariates associated with the ith observation.

The OLLRc regression includes two special models: the Rice (for $\nu = 1$) and Rayleigh (for $\nu = 1$ and $\sigma_i = 0$) regressions.

The log-likelihood function for the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \nu)^\top$ from a random sample $(y_1, \mathbf{x}_1), \cdots, (y_n, \mathbf{x}_n)$ has the form

$$\begin{aligned} l(\boldsymbol{\theta}) \ = \ & n\log(\nu) + \sum_{i=1}^{n} \log\left(\frac{y_i}{\mu_i}\right) - \sum_{i=1}^{n}\left(\frac{y_i^2+\sigma_i^2}{2\mu_i^2}\right) + \sum_{i=1}^{n}\log\left[I_0\left(\frac{y_i\sigma_i}{\mu_i}\right)\right] + \\ & (\nu-1)\sum_{i=1}^{n}\log\left\{\left[1-Q_1\left(\frac{\sigma_i}{\mu_i},\frac{y_i}{\mu_i}\right)\right]Q_1\left(\frac{\sigma_i}{\mu_i},\frac{y_i}{\mu_i}\right)\right\} - \\ & 2\sum_{i=1}^{n}\log\left\{\left[1-Q_1\left(\frac{\sigma_i}{\mu_i},\frac{y_i}{\mu_i}\right)\right]^{\nu} + Q_1\left(\frac{\sigma_i}{\mu_i},\frac{y_i}{\mu_i}\right)^{\nu}\right\}. \end{aligned} \tag{3.2}$$

The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be calculated by maximizing (3.2) using the **R** software and standard likelihood techniques can be adopted for inference purposes. Initial values for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ can be taken from the fitted Rice regression model (when $\nu = 1$).

Under conditions that are fulfilled for the parameter vector $\boldsymbol{\theta}$ in the interior of the parameter space but not on the boundary, the asymptotic distribution of $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is multivariate normal $N_{2p+1}(0, K(\boldsymbol{\theta})^{-1})$, where $K(\boldsymbol{\theta})$ is the information matrix. The asymptotic covariance matrix $K(\boldsymbol{\theta})^{-1}$ of $\widehat{\boldsymbol{\theta}}$ can be approximated by the inverse of the $(2p+1) \times (2p+1)$ observed information matrix $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, whose elements can be calculated numerically. The approximate multivariate normal distribution $N_{2p+1}(0, -\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ for $\widehat{\boldsymbol{\theta}}$ can be used in the classical way to construct approximate confidence regions for some parameters in $\boldsymbol{\theta}$.

We can use likelihood ratio (LR) statistics for comparing some special models with the OLLRc regression model in the usual way. Further details are given by [14].

3.1. Simulation studies

Five thousands Monte Carlo simulations are carried out in the **R** software to examine the consistency of the MLEs under two scenarios: the OLLRc distribution and the OLLRc regression. By setting $n = 25$, 80, 160 and 320, a random sample is drawn from the OLLRc(μ, σ, ν) distribution, and the MLEs are calculated in each of these replications. For the regression scenario, we also consider 700.

The OLLRc distribution

We generate observations from the OLLRc distribution using (2.7) and $u \sim U(0,1)$ with $\mu = 0.15$, $\sigma = 1$ and $\nu = 0.2$. We calculate the MLEs in each of the $5,000$ simulations and then the average estimates (AEs), biases, and means squared errors (MSEs). The results in Table 1 indicate that the estimates are accurate since their biases and MSEs converge to zero when n increases.

Table 1: Simulation findings from the OLLRc distribution.

Parameter	$n=25$			$n=80$		
	AE	Bias	MSE	AE	Bias	MSE
μ	0.177	0.027	0.017	0.160	0.010	0.004
σ	0.971	-0.029	0.024	0.994	-0.006	0.003
ν	0.269	0.069	0.073	0.227	0.027	0.018

Parameter	$n=160$			$n=320$		
	AE	Bias	MSE	AE	Bias	MSE
μ	0.154	0.004	0.001	0.152	0.002	0.000
σ	0.998	-0.002	0.001	0.998	-0.002	0.000
ν	0.211	0.011	0.006	0.208	0.008	0.002

The OLLRc regression

Consider the OLLRc regression with $\mu_i = \exp(\beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2})$ and $\sigma_i = \exp(\beta_{20} + \beta_{21}x_{i1} + \beta_{22}x_{i2})$ and fixed parameters $\beta_{10} = -2$, $\beta_{11} = 0.7$, $\beta_{12} = 1.8$, $\beta_{20} = 0.6$, $\beta_{21} = -0.8$ and $\beta_{22} = 0.4$.

For the generation process, we consider: $Y_i \sim \text{OLLRc}(\mu_i, \sigma_i, \nu)$, $x_{i1} \sim \text{Bernoulli}(0.5)$ and $x_{i2} \sim U(0,1)$. The simulation results from the fitted OLLRc regression in Table 2 indicate that the AEs go to the true parameters and that the biases and MSEs tend to vanish when n increases in agreement with first-order asymptotic theory.

4. THE OLLRC PARTIALLY LINEAR REGRESSION

The dependent variables can be influenced by explanatory variables with linear and non-linear effects in many areas. Recently, several works have been published related to regression models, for example, [2], [28], [13], [24], [22], [4], [25], among others.

In this context, considering the penalized smoothing based on the cubic-spline, we construct the partially linear regression based on the OLLRc distribution. The systematic component for the parameter μ_i in terms of the explanatory variables $\mathbf{x}_1, ..., \mathbf{x}_p$ (linear effects) and $\mathbf{t} = (t_i)$ (non-linear effect) has the form (for $i = 1, ..., n$)

$$\mu_i = \exp\Big\{\mathbf{x}_i^\top \boldsymbol{\beta}_1 + h(t_i)\Big\}, \tag{4.1}$$

where $\boldsymbol{\beta}_1 = (\beta_{11}, ..., \beta_{1p})^\top$ is the unknown parameter vector and $h(\cdot)$ is an unknown smooth function of t_i.

Table 2: Simulation findings from the OLLRc regression with $\nu = 0.6$ and $\nu = 2$.

n	Parameter	$\nu = 0.6$			$\nu = 2$		
		AE	Bias	MSE	AE	Bias	MSE
25	β_{10}	−2.438	−0.438	0.530	−2.439	−0.439	0.632
	β_{11}	0.671	−0.029	0.141	0.736	0.036	0.159
	β_{12}	2.075	0.275	0.625	2.142	0.342	0.625
	β_{20}	0.580	−0.020	0.008	0.596	−0.004	0.002
	β_{21}	−0.681	0.119	0.046	−0.737	0.063	0.018
	β_{22}	0.434	1.234	0.041	0.406	1.206	0.012
	ν	0.546	−0.054	0.086	1.951	−0.049	0.387
80	β_{10}	−2.118	−0.118	0.088	−2.081	−0.081	0.113
	β_{11}	0.673	−0.027	0.032	0.700	<0.001	0.036
	β_{12}	1.836	0.036	0.092	1.859	0.059	0.099
	β_{20}	0.597	−0.003	0.002	0.600	<0.001	<0.001
	β_{21}	−0.747	0.053	0.020	−0.800	<0.001	0.006
	β_{22}	0.405	1.205	0.016	0.400	1.200	0.003
	ν	0.568	−0.032	0.023	2.023	0.023	0.161
160	β_{10}	−2.050	−0.050	0.035	−2.016	−0.016	0.051
	β_{11}	0.691	−0.009	0.015	0.699	−0.001	0.018
	β_{12}	1.819	0.019	0.041	1.811	0.011	0.049
	β_{20}	0.600	0.000	0.001	0.600	<0.001	<0.001
	β_{21}	−0.781	0.019	0.012	−0.807	−0.007	0.004
	β_{22}	0.399	1.199	0.008	0.398	1.198	0.001
	ν	0.590	−0.010	0.009	2.036	0.036	0.078
320	β_{10}	−2.021	−0.021	0.016	−2.007	−0.007	0.025
	β_{11}	0.697	−0.003	0.007	0.703	0.003	0.008
	β_{12}	1.809	0.009	0.019	1.803	0.003	0.023
	β_{20}	0.600	<0.001	0.001	0.600	<0.001	<0.001
	β_{21}	−0.796	0.004	0.007	−0.804	−0.004	0.002
	β_{22}	0.400	1.200	0.004	0.400	1.200	0.001
	ν	0.597	−0.003	0.004	2.020	0.020	0.038
700	β_{10}	−2.007	−0.007	0.007	−2.002	−0.002	0.011
	β_{11}	0.702	0.002	0.003	0.701	0.001	0.004
	β_{12}	1.803	0.003	0.009	1.802	0.002	0.011
	β_{20}	0.600	<0.001	<0.001	0.600	<0.001	<0.001
	β_{21}	−0.805	−0.005	0.004	−0.803	−0.003	0.001
	β_{22}	0.400	1.200	0.002	0.400	1.200	<0.001
	ν	0.601	0.001	0.002	2.011	0.011	0.017

For the partially linear regression (4.1), we consider the penalty based on the second order derivative of the function $h(\cdot)$ [15].

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \sigma, \nu)^\top$ be the parameter vector related to the parametric part and $\lambda > 0$ be the smoothing parameter that controls the smoothness of the curve. Consider a smooth function $h(t)$ (second order differentiable function in the interval $[a, b]$), such that it is a cubic smoothing splines where the nodes are the ordered values of $t_1, ..., t_n$, say $t_1^0 < t_2^0 < ... < t_q^0$, and q indicates the amount of distinct values for the explanatory variable t_i that is controlled in a non-parametric way.

The penalty described above can be expressed in matrix notation [6]. Let d_i be the distance between two subsequent and different control points called nodes i and $i+1$, i.e., $d_i = t_{i+1}^0 - t_i^0$ (for $i = 1, ..., q-1$). We define the elements q_{ij} (for $i = 1, ..., q$ and $j = 2, ..., q-1$)

of the $q \times (q-2)$ tridiagonal matrix $\mathbf{A}$ by

$$q_{j-1,j} = d_{j-1}^{-1}, \quad q_{jj} = -d_{j-1}^{-1} - d_j^{-1}, \quad q_{j+1,j} = d_j^{-1} \quad \text{and} \quad q_{ij} = 0 \quad \text{for} \quad |i-j| \geq 2.$$

The elements r_{ij} (for $i = 2, ..., q-1$ and $j = 2, ..., q-1$) of the $(q-2) \times (q-2)$ symmetric matrix $\mathbf{B}$ are

$$\begin{aligned} r_{ii} &= \frac{1}{3}(d_{i-1} + d_i) \quad \text{for} \quad i = 2, ..., q-1, \\ r_{i,i+1} &= r_{i+1,i} = \frac{1}{6} d_i \quad \text{for} \quad i = 2, ..., q-2 \quad \text{and} \\ r_{ij} &= 0 \quad \text{for} \quad |i-j| \geq 2. \end{aligned}$$

Further, let $\mathbf{K} = \mathbf{A}\mathbf{B}^{-1}\mathbf{A}^{\mathbf{T}}$ be a $q \times q$ positive definite matrix. The parameters associated with the covariates in the linear and nonlinear effects ($\boldsymbol{\theta}$ and $\mathbf{h} = (h(t_1^0), ..., h(t_q^0))$, respectively) are determined by maximizing the penalized log-likelihood function

$$\begin{aligned} l_p(\boldsymbol{\theta}, \mathbf{h}) &= n \log(\nu) + \sum_{i=1}^n \log\left(\frac{y_i}{\mu_i}\right) - \sum_{i=1}^n \left(\frac{y_i^2 + \sigma^2}{2\mu_i^2}\right) + \sum_{i=1}^n \log\left[I_0\left(\frac{y_i \sigma}{\mu_i}\right)\right] + \\ & (\nu - 1) \sum_{i=1}^n \log\left\{\left[1 - Q_1\left(\frac{\sigma}{\mu_i}, \frac{y_i}{\mu_i}\right)\right] Q_1\left(\frac{\sigma}{\mu_i}, \frac{y_i}{\mu_i}\right)\right\} - \\ & 2 \sum_{i=1}^n \log\left\{\left[1 - Q_1\left(\frac{\sigma}{\mu_i}, \frac{y_i}{\mu_i}\right)\right]^\nu + Q_1\left(\frac{\sigma}{\mu_i}, \frac{y_i}{\mu_i}\right)^\nu\right\} - \frac{\lambda}{2} \mathbf{h}^T \mathbf{K}\, \mathbf{h}, \end{aligned} \tag{4.2}$$

where λ is the unknown smoothing parameter. The maximization of (4.2) is equivalent to the cubic smoothing spline. We use the gamlss($\cdot$) function from the `gamlss` [20] package to implement the OLLRc regression, and calculate the penalized maximum likelihood estimates (PMLEs). The cs($\cdot$) function is used to model the nonlinear effect based on cubic smoothing splines function [21].

4.1. Simulations for the OLLRc partially linear regression

We present a Monte Carlo study with a smooth function to verify the adequacy of the PMLEs in this regression. The covariates (linear and non-linear effects) and the response variable are generated as follows: $x_{i1} \sim U(0,1)$, $x_{i2} \sim$ Bernoulli(0.5), $t_{i3} \sim U(0, 0.045)$ (scenario 1), $t_{i3} \sim U(0, 0.03)$ (scenario 2) and $y_i \sim$ OLLRc(μ_i, σ, ν).

Further, the systematic component is $\mu_i = \exp\{0.5x_{i1} - 0.4x_{i2} + h(t_{i3})\}$, where $\beta_{11} = 0.5$, $\beta_{12} = -0.4$, $\sigma = 0.1$ and $\nu = 0.7$ (for $n = 60, 180,$ 400 and 700). In addition, 1,000 Monte Carlo samples are generated and, for each sample size, the PMLEs of the parameters are found for each replication, and then the AEs, biases and MSEs are calculated. The numbers in Table 3 indicate that the AEs tend to the true parameters and the biases and MSEs converge to zero when n increases. Thus, the sample distribution of the PMLEs is approximately normal.

Regarding the analysis of the nonlinear effect (t_{i3}), the true smooth curves $h(t_{i3}) = \sin(-50t_{i3}\pi) + \cos(30\pi t_{i3})$ and $h(t_{i3}) = \cos(100t_{i3}\pi) + \tan(15\pi t_{i3} - 1)$ and their respective estimated curves (based on 1,000 simulations) are displayed in Figure 4 for both scenario 1. We note that the estimated curves approach to the true curve for large sample sizes (as expected).

Table 3: Findings for the OLLRc partially linear regression.

n	Parameter	Scenario 1			Scenario 2		
		AE	Bias	MSE	AE	Bias	MSE
60	β_{11}	0.508	0.008	0.186	0.495	0.005	0.171
	β_{12}	-0.415	-0.015	0.064	-0.407	-0.007	0.055
	σ	0.186	0.086	0.016	0.134	0.034	0.004
	ν	0.636	-0.064	0.028	0.700	0.000	0.024
180	β_{11}	0.518	0.018	0.047	0.498	-0.002	0.048
	β_{12}	-0.409	-0.009	0.015	-0.405	-0.005	0.014
	σ	0.136	0.036	0.004	0.104	0.004	0.001
	ν	0.687	-0.013	0.005	0.730	0.030	0.007
400	β_{11}	0.510	0.010	0.018	0.504	0.004	0.019
	β_{12}	-0.404	-0.004	0.006	-0.403	-0.003	0.006
	σ	0.112	0.012	0.002	0.095	-0.005	0.001
	ν	0.703	0.003	0.002	0.737	0.037	0.004
700	β_{11}	0.505	0.005	0.011	0.494	-0.006	0.011
	β_{12}	-0.406	-0.006	0.004	-0.405	-0.005	0.003
	σ	0.100	0.000	0.001	0.090	-0.010	0.001
	ν	0.710	0.010	0.001	0.740	0.040	0.003

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

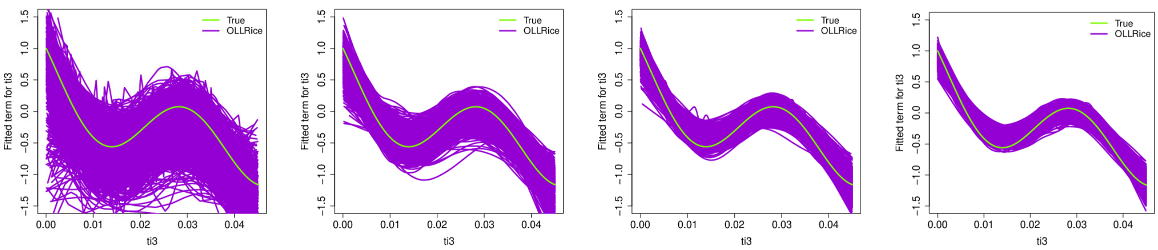

Figure 4: The generated and estimated curves for $h(t_{i3})$ (scenario 1): (a) $n = 60$, (b) $n = 180$. (c) $n = 400$. (d) $n = 700$.

5. RESIDUAL ANALYSIS AND SIMULATIONS

We define the quantile residuals (qrs) [3] for the OLLRc regression as

$$qr_i = \Phi^{-1}\left\{\frac{\left[1 - Q_1\left(\frac{\sigma_i}{\mu_i}, \frac{y_i}{\mu_i}\right)\right]^{\nu}}{\left[1 - Q_1\left(\frac{\sigma_i}{\mu_i}, \frac{y_i}{\mu_i}\right)\right]^{\nu} + Q_1\left(\frac{\sigma_i}{\mu_i}, \frac{y_i}{\mu_i}\right)^{\nu}}\right\}, \tag{5.1}$$

where $\Phi(\cdot)^{-1}$ is the inverse of the standard normal cdf and μ_i and σ_i are given in Equation (3.1).

Simulations for the OLLRc regression

Some simulations of sizes 25, 80, 160, 320 and 700 are performed using the algorithm of Section 3.1 to examine the empirical distribution of these residuals. Figure 5 (for $\nu = 0.6$)

show that this distribution becomes closer to the standard normal distribution when n increases.

(a) (b) (c) (d) (e)

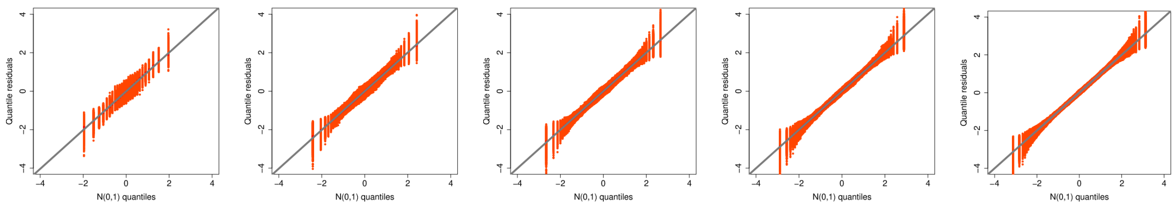

Figure 5: Normal probability plots of the qrs ($\nu = 0.6$). (a) $n = 25$. (b) $n = 80$. (c) $n = 160$. (d) $n = 320$. (e) $n = 700$.

Simulations for the OLLRc partially linear regression

Consider a simulation study to investigate the empirical distribution of the qrs for the OLLRc partially linear regression by generating 60, 180, 400 and 700 observations from Equation (4.1). Normal probability plots in Figure 6 reveal that the empirical distribution of the qrs is close to the standard normal for all samples, and the approximation becomes better when n increases.

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

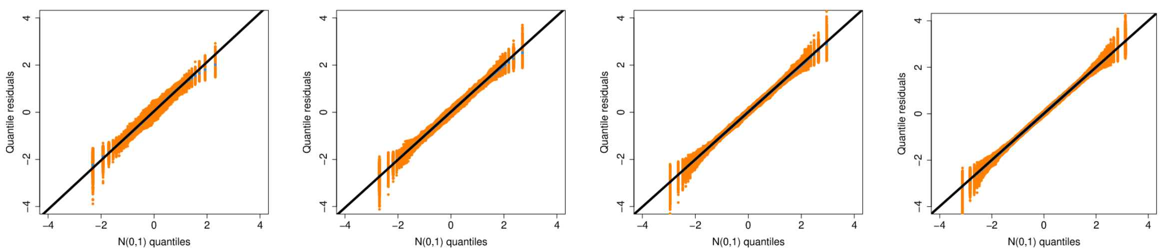

Figure 6: Normal probability plots of the qrs (scenario 1). (a) $n = 60$. (b) $n = 180$. (c) $n = 400$. (d) 700.

Thus, we use normal probability plots for the residuals (qr_i) with simulated envelopes for both models, as suggested by [1], as follows:

1. Fit the model and generate a sample of n independent observations using the fitted model as if it were the true model;
2. Fit the model to the generated sample using the data set ($\mathbf{x}_i$) and compute the values of the residuals;
3. Repeat steps (1) and (2) m times;
4. Obtain ordered values of the residuals, $qr^*_{(i)v}$, $i = 1, ..., n$ and $v = 1, ..., m$;
5. Consider n sets of the m ordered statistics and for each set compute the mean, minimum and maximum values;

6. Plot these values and the ordered residuals of the original sample against the normal scores. The minimum and maximum values of the m ordered statistics provide the envelope.

The residuals outside the limits provided by the simulated envelope require further investigation. Additionally, if a considerable proportion of points falls outside the envelope, we have evidence against the adequacy of the fitted model. Plots of the residuals against the fitted values can also be useful.

6. APPLICATIONS

We present two applications of the new regressions: the first for the OLLRc regression, and the second for the OLLRc partially linear regression.

6.1. Shrinkage volume data

We consider a data set referring to the shrinkage volume of three wood species: Cedrilho (Erismauncinatum Warm), Morototoni (SchefffleramorotoniAubl) and Pinus (Pinus spp). The shrinkage volume of wood is defined as the phenomenon related to the dimensional variation of wood due to moisture exchange with the surrounding environment until a condition of balance is attained, called the hygroscopic equilibrium moisture. The variations in the dimensions of wood specimens occur when they lose or gain moisture in relation to the saturation point of the fibers, which in general is in the range of 28% to 30% water. The dimensional variation involves either shrinkage or swelling. The shrinkage volume of wood varies widely among species, depending on the drying method and the behavior of the particular wood specimen, occasionally leading to alterations of shape and the formation of cracks and warping. Special precaution needs to be taken in situations that require wood stability. For structural framework, flooring, doors, door/window frames and furniture, cracking and warping can cause serious losses, requiring replacement. Thus correct drying methods to attain equilibrium moisture are essential. There are various explanations for the increase of contraction with higher temperature. One of them can be the reduction of the equilibrium moisture, but that factor has been experimentally found to cause an increase in contraction of less than 1%, when in reality the increase in contraction is much more than this. For these reasons, we study the effects of drying temperature and wood species on the shrinkage volume of wood specimens.

The experiment was carried out in the first half of 2020 at the School of Agronomic Sciences of Paulista State University (UNESP), located in the city of Botucatu, São Paulo, Brazil. The tests were carried out with wood specimens with volume of 20 cm^3 dried in a muffle furnace at final temperatures of 300 ^{o}C and 500 ^{o}C. A muffle furnace is type of oven that operates at high temperatures used in laboratories. The final temperature was applied for 10 minutes and the carbonization rate was 14.3 ^{o}C/min for each species. We used a pachymeter to measure the dimensions of each specimen to calculate the volume before and after carbonization in stable conditions. The variables involved are the following (for

$i = 1, ..., 36$): y_i: volumetric shrinkage (in cm^3); x_{i1}: temperature (0=300^{o}C, 1=500^{o}C) and x_{i2}: wood species (0=Cedrilho, 1=Morototoni, 2=Pinus) with two dummy variables (d_{i1}, d_{i2}).

First, we provide a marginal analysis of the response variable. Table 4 reports the MLEs (their standard errors in parentheses) of the parameters from the fitted OLLRc, Rice and Rayleigh distributions, and the statistics: Akaike Information Criterion (AIC) and Global Deviance (GD). These results indicate that the OLLRc distribution is the best model to the current data.

Table 4: Findings from the fitted distributions.

Distribution	$\log(\mu)$	$\log(\sigma)$	ν	AIC	GD
OLLRc	**−0.666**	**2.064**	**0.155**	**157.836**	**151.836**
	(0.109)	**(0.031)**	**(0.024)**		
Rice	0.804	2.021	1	162.263	158.263
	(0.124)	(0.052)	(—)		
Model	$\log(\mu)$	σ	ν		
Rayleigh	1.756	0	1	181.017	179.017
	(0.083)	(—)	(—)		

The likelihood ratio (LR) statistics in Table 8 indicate that the OLLRc distribution is the best model to these data among the three distributions. The estimated pdf of the fitted models in Figure 7 show that the OLLRc distribution gives the best fit to the shrinkage volume data.

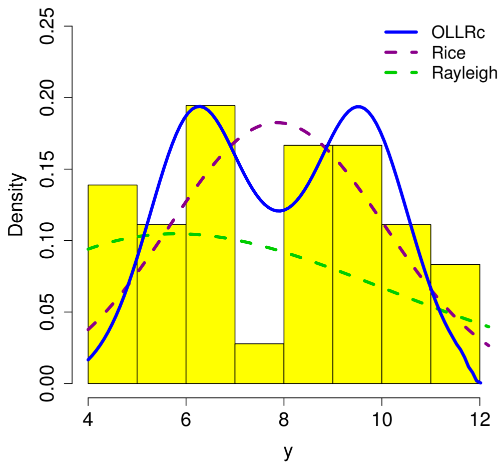

Figure 7: Estimated OLLRc, Rice and Rayleigh densities.

The OLLRc regression

The systematic components are given by

$$\mu_i = \exp(\beta_{10} + \beta_{11}x_{i1} + \beta_{12}d_{i1} + \beta_{13}d_{i2})$$

and

$$\sigma_i = \exp(\beta_{20} + \beta_{21}x_{i1} + \beta_{22}d_{i1} + \beta_{23}d_{i2}).$$

The MLEs, SEs and p-values from the fitted OLLRc regression to the current data are reported in Table 5. Some conclusions are addressed at the end of this application.

Table 5: Findings from the fitted OLLRc regression to the shrinkage volume data.

Parameter	Estimate	SE	p-value
β_{10}	−0.537	2.509	0.832
β_{11}	−0.346	0.191	0.082
β_{12}	−1.144	0.541	0.044
β_{13}	−0.786	0.219	0.001
β_{20}	1.687	0.135	<0.001
β_{21}	0.448	0.129	0.002
β_{22}	0.080	0.042	0.068
β_{23}	0.286	0.037	<0.001
ν	0.236	0.908	

The AIC and GD values in Table 6 confirm that the OLLRc regression is the best model to the shrinkage volume data.

Table 6: Adequacy statistics.

Regression	AIC	GD
OLLRc	**93.236**	**75.2356**
Rice	95.164	79.164
Rayleigh	178.373	170.373

Table 9 compares the new regression with two special models fitted to the wood volumetric retraction data, whose figures indicate that the OLLRc regression is the best model among the three.

(a) (b) (c)

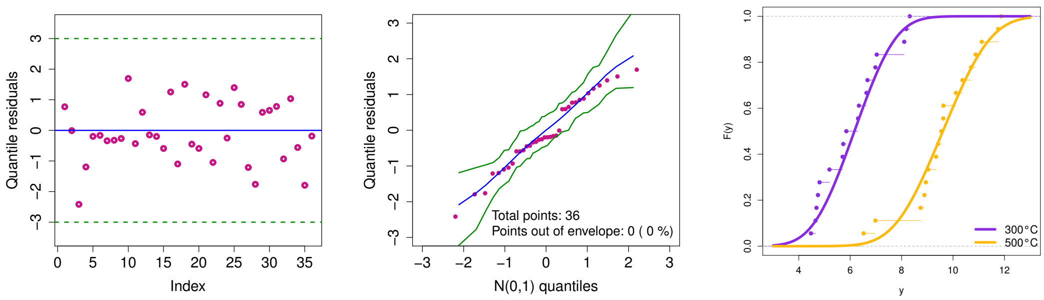

Figure 8: (a) Index plot of the qrs. (b) Normal probability plot for the qrs. (c) Empirical and estimated cdf for the temperature (300^{o}C, 500^{o}C).

Figures 8a and 8b display the index plot of the qrs and the normal probability plot with generated envelope for the OLLRc regression, respectively. These plots do not provide departure from the model assumptions. We report in Table 7 the presence of significant effects of the levels of wood from the fitted OLLRc regression to the shrinkage volume data.

Table 7: Findings for the three wood levels from the fitted OLLRc regression.

Test for link μ			
Hypotheses H_0	Estimate	SD	p-value
Cedrilho = Morototoni	−1.144	0.541	0.044
Cedrilho = Pinus	−0.786	0.219	0.001
Morototoni = Pinus	0.358	0.563	0.531

Test for link σ			
Hypotheses H_0	Estimate	SD	p-value
Cedrilho = Morototoni	0.080	0.042	0.068
Cedrilho = Pinus	0.286	0.037	<0.001
Morototoni = Pinus	0.206	0.023	<0.001

Assessing covariate effects on parameter μ

For a 5% significance level, we conclude:

The temperature levels are not statistically different for μ (see Table 5). The Cedrilho and Morototoni wood species and Cedrilho and Pinus wood species are statistically different for μ (see Table 5). The Pinus and Morototoni wood species are statistically different for μ (see Table 7).

Assessing covariate effects on parameter σ

The temperature levels are statistically different for σ (see Table 5 and Figure 8c). The Cedrilho and Morototoni wood species are not statistically different for σ (see Table 5). The Cedrilho and Pinus wood species are statistically different for σ (see Table 5). Morototoni and Pinus wood species are statistically different for σ (see Table 7).

Finally, the empirical and estimated cdf of the OLLRc regression are displayed in Figure 8c for different levels of temperatures, thus showing that this regression is suitable for the shrinkage volume data.

Table 8: LR tests (Application 1 without considering covariates).

Distributions	Hypotheses	LR statistic	p-value
OLLRc vs Rice	$H_0 : \nu = 1$ vs $H_1 : H_0$ is false	6.428	0.011
OLLRc vs Rayleigh	$H_0 : \sigma = 0$ and $\nu = 1$ vs $H_1 : H_0$ is false	27.182	<0.001

Table 9: LR statistics for three fitted regressions (Application 1).

Regressions	Hypotheses	LR statistic	p-value
OLLRc vs Rice	$H_0 : \nu = 1$ vs $H_1 : H_0$ is false	3.929	0.047
OLLRc vs Rayleigh	$H_0 : \sigma = 0$ and $\nu = 1$ vs $H_1 : H_0$ is false	95.137	<0.001

Table 10: LR tests (Application 2 without considering covariates).

Regressions	Hypotheses	LR statistic	p-value
OLLRc vs Rice	$H_0 : \nu = 1$ vs $H_1 : H_0$ is false	5.062	0.024
OLLRc vs Rayleigh	$H_0 : \nu = 0$ and $\nu = 1$ vs $H_1 : H_0$ is false	48.693	<0.001

Table 11: LR statistics for three fitted regressions (Application 2).

Models	Hypotheses	LR statistic	p-value
OLLRc vs Rice	$H_0 : \nu = 1$ vs $H_1 : H_0$ is false	96.007	<0.001
OLLRc vs Rayleigh	$H_0 : \sigma = 0$ and $\nu = 1$ vs $H_1 : H_0$ is false	148.996	<0.001

6.2. Milk production data

In the second application, the data referred to the quantity of cold milk, raw or homogenized, acquired (thousand liters) between the first quarter of 2005 until the fourth quarter of 2015 in Northeast Brazil. The Northeast is considered a new dairy farming region due to the expanded market for milk and dairy products in Brazil, including the Northeast itself, in recent years, driven by increased consumption, in turn related to rising purchasing power in the region, as well as stronger demand from other regions of Brazil and neighboring countries. To understand the advance of dairy farming in the Northeast, it is necessary to know something about the division of the region in terms of climate. Basically there are four sub-regions: the forest zone, sub-humid zone (agreste), mid-north and hinterland (sertão). Each of them has distinct physical characteristics that facilitate or hamper dairy farming. In this paper we study the states of Bahia and Pernambuco. The data set is obtained from the Brazilian Institute of Geography and Statistics [10]. The dependent variable is the production of milk produced, while the explanatory variables are: (i) state of production (Bahia and Pernambuco, two large producers of milk in the Northeast); and (ii) the quarter of production, between the first quarter of 2005 to the last quarter of 2015. This last covariable has a nonlinear effect on the quantity of cold milk. An option to analyze this data set is by means of the OLLRc partially linear regression. The variables under study are: y_i: production of cold milk (raw or homogenized) (thousand liters) (this variable was divided by 10,000); x_{i1}: states (Bahia and Pernambuco) and t_{i2}: quarter (from 1 to 44), for $i = 1, ..., 88$.

Table 12: Findings from the OLLRc partially linear regression.

Model	$\log(\mu)$	$\log(\sigma)$	ν	AIC	GD
OLLRc	**−0.146**	**1.905**	**0.326**	**372.189**	**366.189**
	(0.324)	**(0.028)**	**(0.158)**		
Rice	0.722	1.870	1	375.251	371.251
	(0.081)	(0.036)	(—)		
Model	$\log(\mu)$	σ	ν		
Rayleigh	1.616	0	1	416.881	414.881
	(0.053)	(—)	(—)		

We examine these data by studying the distribution of the response variable, that is, making a marginal analysis. Table 12 reports the results from three fitted distributions, which indicate that the OLLRc distribution can be chosen as the best model.

The numbers in Table 10 support the OLLRc distribution as the best fit model for these data. Figure 9 displays the estimated pdfs of the fitted models and shows that the wider distribution is the best for the current data.

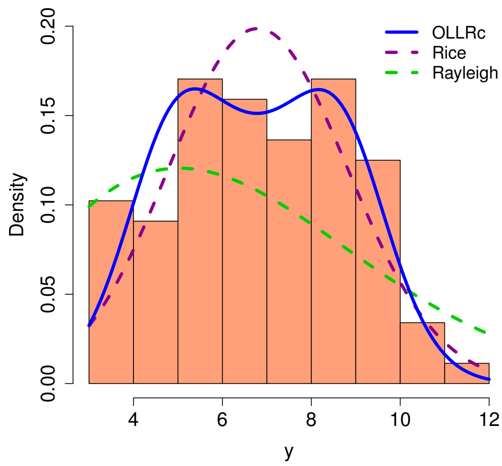

Figure 9: Estimated densities of the OLLRc, Rice and Rayleigh distributions.

Hence, the OLLRc distribution is a good candidate for modeling milk production data.

The OLLRc partially linear regression

Figure 10 displays the scatter plot between the response variable y_i and the covariate t_{i2}. So, there is a non-linear trend between these two variables, which requires the OLLRc partially linear regression model to analyze the current data.

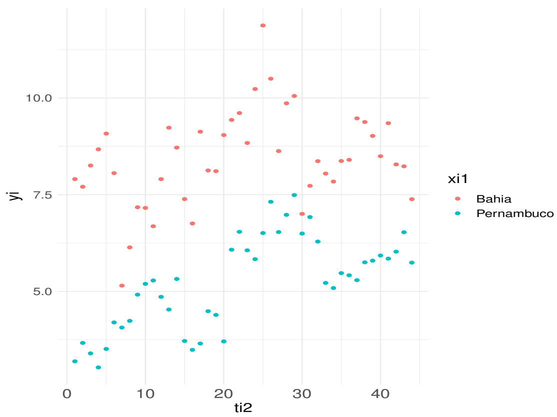

Figure 10: Scatter diagrams: production of cold milk versus quarter.

We now consider the systematic component:

$$\mu_i = \exp[\beta_{10} + \beta_{11} x_{i1} + h(t_{i2})],$$

where $h(\cdot)$ is an arbitrary smooth function associated with the explanatory variable t_{i2}; for more details, see Section 4.

Table 13 reports the generalized Akaike information criterion (GAIC), based on [19], and confirms that the OLLRc partially linear regression can be chosen as the best model.

Table 13: Model selection measures.

Model	GAIC
OLLRc	**256.802**
Rice	350.809
Rayleigh	401.797

Table 14 provides several quantities obtained from the fitted of this partially regression to the milk production data. There is a significant difference between the states of Bahia and Pernambuco in relation to milk production since the covariate x_{i1} is significant at a level of 5%.

Table 14: Findings from the fitted OLLRc partially linear regression.

Parameter	Estimate	SE	p-Value
β_{10}	1.882	0.036	<0.001
β_{11}	−0.476	0.031	<0.001
$\log(\sigma)$	−11.080	398.130	
ν	4.334	0.387	

The figures in Table 11 from two LR tests indicate that the wider regression is the best model for these data. Figure 11(a) provides the plots of the qrs against the observations index, whereas Figure 11(b) reports the normal probability plot with generated envelope. These plots support the wider linear regression for these data and that there are no observations falling outside the envelope.

Finally, Figure 11c provides the estimate of the non-linear effect. The vertical axis refers to the values of t_{i2} and the horizontal axis to the contribution of the estimated smooth curve to the values of t_i. We note from this plot that the amount of milk production is non-linear in relation to the quarter effect. In addition, a greater amount of milk production is achieved between quarters 20 to 35 (approximately).

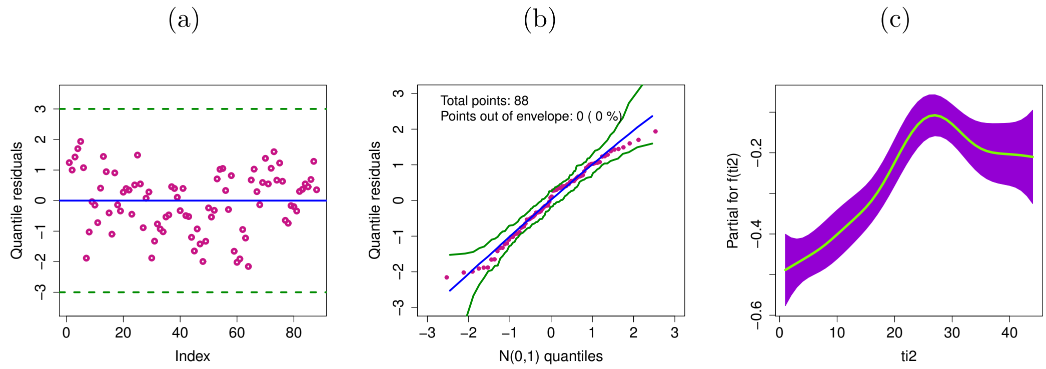

Figure 11: (a) Index plot of the qrs. (b) Normal probability plot for the qrs. (c) Smooth curve fitted from the OLLRc partially linear regression.

7. CONCLUDING REMARKS

The article presented the *odd log-logistic Rice* (OLLRc) distribution and proposed two regression models based on this distribution for the analysis of data that has no unimodal shape. We believe that this paper shows the first use of the Rice distribution in the context of a regression with two systematic components. We defined quantile residuals and provided some simulation studies. We proved the utility of the distribution and of the regressions by means of two data sets on the volumetric shrinkage of the wood and milk production. Future work can be developed using the new OLLRc model in different areas of research. Further, it may be of interest to propose a heteroscedastic semiparametric regression model based on the OLLRc distribution.

A. APPENDIX

Two power series follow for the numerator and denominator of (2.1) (for $\nu > 0$ real):

$$G(y)^\nu = \sum_{k=0}^{\infty} a_k\, G(y)^k \quad \text{and} \quad [1 - G(y)]^\nu = \sum_{k=0}^{\infty} (-1)^k \binom{\nu}{k} G(y)^k, \tag{A.1}$$

where

$$a_k = a_k(\nu) = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{\nu}{j} \binom{j}{k}.$$

Inserting (A.1) in Equation (2.1) leads to

$$F(y) = \frac{\sum_{k=0}^{\infty} a_k\, G(y)^k}{\sum_{k=0}^{\infty} b_k\, G(y)^k} = \sum_{k=0}^{\infty} c_k\, G(y)^k \tag{A.2}$$

where $b_k = a_k + (-1)^k \binom{\nu}{k}$ (for $k \geq 0$), $c_0 = a_0/b_0$ and the coefficients c_k's (for $k \geq 1$) are calculated recursively as

$$c_k = b_0^{-1} \left(a_k - \sum_{r=1}^{k} b_r\, c_{k-r} \right).$$

By differentiating (A.2), the pdf of Y follows as

$$f(y) = \sum_{k=0}^{\infty} c_{k+1}\, h_{k+1}(y), \tag{A.3}$$

where $h_{k+1}(x) = (k+1)\, G(y)^k\, g(y)$ is the exponentiated-G (exp-G) density function with power parameter $k+1$.

Hence, the exp-Rice density can be expressed from (2.3) and (2.4) as

$$h_{k+1}(y) = \left[1 - Q_1\left(\frac{\sigma}{\mu}, \frac{y}{\mu} \right) \right]^k \frac{(k+1)y}{\mu^2} \exp\left(-\frac{y^2 + \sigma^2}{2\mu^2} \right) I_0\left(\frac{y\sigma}{\mu^2} \right) \tag{A.4}$$

The mathematical properties of the OLLRc distribution can be determined numerically by combining (A.3) and (A.4) for a given number of terms (say 10) in the linear combination.

ACKNOWLEDGMENTS

We would like to thank the Editor, and two anonymous reviewers for their time and valuable remarks. This work was supported by CNPq and CAPES, Brazil.

REFERENCES

[1] Atkinson, A.C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Clarendon Press Oxford, Oxford.

[2] Cui, X.; Lu, Y. and Peng, H. (2017). Estimation of partially linear regression models under the partial consistency property, *Computational Statistics and Data Analysis*, **115**, 103–121.

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

[4] Ferreira, C.S.; Paula, G.A. and Lana, G.C. (2022). Estimation and diagnostic for partially linear models with first-order autoregressive skew-normal errors, *Computational Statistics*, **37**, 445–468.

[5] Gleaton, J.U. and Lynch, J.D. (2006). Properties of generalized log-logistic families of lifetime distributions, *Journal of Probability and Statistical Science*, **4**, 51–64.

[6] Green, P.J. and Silverman, B.W. (1993). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, CRC Press.

[7] Green, P.J. and Yandell, B.S. (1985). Semi-parametric generalized linear models, *In Generalized Linear Models*, Springer, New York, NY, 44–55.

[8] Hallbjomer, P. (2003). Modified Rice distribution for signals with limited available power, *IEEE antennas and wireless propagation letters*, **2**, 159–162.

[9] Hastie, T.J. and Tibshirani, R.J. (1990). Generalized Additive Models, CRC Press, **43**.

[10] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. (2020). Pesquisa Trimestral do Leite: Período de 2005 até 2015, disponível em: `https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/21121-primeiros-resultadios-2leite.html?=&t=series-historicas`

[11] Jeong, S.; Park, T. and van Dyk, D.A. (2022). Bayesian model selection in additive partial linear models via locally adaptive splines, *Journal of Computational and Graphical Statistics*, **31**, 324–336.

[12] Lauwers, L.; Barbé, K.; Van Moer, W. and Pintelon, R. (2009). Estimating the parameters of a Rice distribution: A Bayesian approach. In: *2009 IEEE Instrumentation and Measurement Technology Conference*, IEEE, 114–117.

[13] Manghi, R.F.; Cysneiros, F.J.A. and Paula, G.A. (2019). Generalized additive partial linear models for analyzing correlated data, *Computational Statistics and Data Analysis*, **129**, 47–60.

[14] Sen, P.K. and Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*, New York, Chapman & Hall.

[15] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems, *Statistical Science*, **1**, 502–518.

[16] Prataviera, F.; Ortega, E.M.M.; Cordeiro, G.M. and Braga, A.D.S. (2019). The heteroscedastic odd log-logistic generalized gamma regression model for censored data, *Communications in Statistics – Simulation and Computation*, **48**, 1815–1839.

[17] Rice, S.O. (1945). Mathematical analysis of random noise, *The Bell System Technical Journal*, **24**, 46–156.

[18] Ruppert, D.; Wand, M.P. and Carroll, R.J. (2003). Semiparametric Regression, Cambridge University Press, United Kingdom.

[19] Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507–554.

[20] Stasinopoulos, D.M. and Rigby, R.A. (2007). Generalized additive models for location scale and shape (GAMLSS) in **R**. *Journal of Statistical Software*, **23**, 1–46.

[21] Stasinopoulos, M.D.; Rigby, R.A.; Heller, G.Z.; Voudouris, V. and De Bastiani, F. (2017). Flexible Regression and Smoothing: Using gamlss in R, CRC Press, New York.

[22] Vasconcelos, J.C.S. and Villegas, C. (2021a). Generalized symmetrical partial linear model, *Journal of Applied Statistics*, **48**, 557–572.

[23] Vasconcelos, J.C.S.; Cordeiro, G.M.; Ortega, E.M.M. and Rezende, E.M (2021b). A new regression model for bimodal data and applications in agriculture, *Journal of Applied Statistics*, **48**, 349–372.

[24] Vasconcelos, J.C.S.; Cordeiro, G.M.; Ortega, E.M.M. and Biaggioni, M.A. (2020). The parametric and additive partial linear regressions based on the generalized odd log-logistic log-normal distribution, *Communications in Statistics – Theory and Methods*, **51**, 3480–3507.

[25] Vasconcelos, J.C.S.; Cordeiro, G.M. and Ortega, E.M.M. (2022). The semiparametric regression model for bimodal data with different penalized smoothers applied to climatology, ethanol and air quality data, *Journal of Applied Statistics*, **49**, 248–267.

[26] Yakovleva, T.V. and Kulberg, N.S. (2013). Noise and signal estimation in MRI: two-parametric analysis of rice-distributed data by means of the maximum likelihood approach, *American Journal of Theoretical and Applied Statistics*, **2**, 67–79.

[27] Yee, T.W. (2008). The VGAM package, *R News*, **8**, 28–39.

[28] Zeng, Z. and Liu, X (2018). A difference-based approach in the partially linear model with dependent errors, *Journal of Inequalities and Applications*, **267**, 1–164.