On Diagnostics in Bell–Touchard Regression Models

Artur J. Lemonte

Universidade Federal do Rio Grande do Norte

Abstract

The Bell–Touchard regression model for count response variables was introduced recently in the statistical literature. This regression model for counts is, in some aspects, similar to the generalized linear model framework, and the mean response is related to covariates and unknown regression coefficients, which allows for parameter interpretation in terms of the response variable in its original scale. However, diagnostic methods for this class of regression models were not developed. In this paper, we fill this gap and propose a variety of diagnostic tools, such as leverage measures, global influence, normal curvatures of local influence under some perturbation schemes, and Pearson residuals. All diagnostic measures developed are applied in a real data set to analyze the fitted Bell–Touchard regression model in practice.

Keywords: count data, generalized leverage, global influence, local influence, overdispersion.

1. Introduction

The two-parameter mean-parameterized Bell–Touchard ('BeTo' for short) family of distributions studied by Lemonte (2022) has probability mass function (PMF) of the form

$$\Pr(Y = y) = \exp\{\phi[1 - e^{W_0(\mu/\phi)}]\}\frac{W_0(\mu/\phi)^y T_y(\phi)}{y!}, \quad y = 0, 1, \dots,$$

where $\mu > 0$ is the location parameter, $\phi > 0$ is the precision parameter, $W_0(\cdot)$ denotes the principal branch of the Lambert function (Corless, Gonnet, Hare, Jeffrey, and Knuth 1996), $T_y(\phi) = \sum_{k=0}^{y} S(y,k)\phi^k$ are the Touchard polynomials studied in Touchard (1933), and $S(y, k)$ are the Stirling numbers of the second kind (see, for instance, Comtet 2010). In particular, it follows that $T_0(\phi) = 1$, $T_1(\phi) = \phi$, $T_2(\phi) = \phi^2 + \phi$, $T_3(\phi) = \phi^3 + 3\phi^2 + \phi$, $T_4(\phi) = \phi^4 + 6\phi^3 + 7\phi^2 + \phi$, etc. We shall use the notation $Y \sim \text{BeTo}(\mu, \phi)$ to indicate that the discrete random variate Y is BeTo distributed. We have that $\mathbb{E}(Y) = \mu$ and $\mathbb{VAR}(Y) = \mu[1 + W_0(\mu/\phi)]$, and so $\mathbb{VAR}(Y) > \mathbb{E}(Y)$, which indicates that this discrete two-parameter mean-parameterized distribution may deal with count data with overdispersion. There are other interesting properties regarding the BeTo distribution, and the reader is referred to Lemonte (2022) and Castellares, Lemonte, and Moreno-Arenas (2020) for further details.

On the basis of the mean-parameterized BeTo distribution, a new regression model was introduced by Lemonte (2022) when the response variable is a count. The mean response was

related to explanatory variables and regression parameters, which allows for parameter interpretation in terms of the response variable in its original scale. In the following, we briefly discuss the BeTo regression model. Let $Y_1, \ldots, Y_n$ denote n independent random variables BeTo distributed, where $Y_i \sim \text{BeTo}(\mu_i, \phi)$ for $i = 1, \ldots, n$. Also, consider the functional relation for the mean parameter given by $\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, where $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ are observations on p known regressors, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is a vector of regression parameters ($p < n$). Note that $\mu_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$, which ensures that $\mu_i > 0$ for all $i = 1, \ldots, n$. In general, we have that $x_{i1} = 1$ (for all $i = 1, \ldots, n$), and so β_1 corresponds to the intercept. Lemonte (2022) showed that the maximum lilekihood (ML) estimates $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^\top$ and $\widehat{\phi}$ of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and ϕ, respectively, can be obtained by iteratively solving simultaneously the equations

$$
\begin{cases}
\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{V}^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{(m)} \boldsymbol{z}^{(m)}, \\
\phi^{(m+1)} = \phi^{(m)} + \dfrac{\text{tr}(\boldsymbol{S}^{(m)})}{\text{tr}(\boldsymbol{P}^{(m)})},
\end{cases}
\tag{1}
$$

where $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]^\top$ is the model matrix with full column rank p, $\boldsymbol{V} = \text{diag}\{v_1, \ldots, v_n\}$, $\boldsymbol{P} = \text{diag}\{p_1, \ldots, p_n\}$, $\boldsymbol{S} = \text{diag}\{s_1, \ldots, s_n\}$, $\boldsymbol{z} = \boldsymbol{\eta} + \boldsymbol{V}^{-1} \boldsymbol{W}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{W} = \text{diag}\{w_1, \ldots, w_n\}$, $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, and

$$
w_i = 1 + W_0(\mu_i/\phi), \quad p_i = \frac{e^{W_0(\mu_i/\phi)}}{\phi \, w_i} + \frac{1}{\phi^2}\left[\left(\frac{T_{y_i+1}(\phi)}{T_{y_i}(\phi)}\right)^2 - \frac{T_{y_i+2}(\phi)}{T_{y_i}(\phi)}\right],
$$

$$
v_i = \frac{\mu_i}{w_i}, \quad s_i = -e^{W_0(\mu_i/\phi)} - \frac{(y_i - \mu_i)}{\phi \, w_i} + \frac{T_{y_i+1}(\phi)}{\phi \, T_{y_i}(\phi)}.
$$

Also, $\text{tr}(\cdot)$ denotes the trace operator, and $m = 0, 1, \ldots$ means the iteration counter in the iterative process (1). Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$ be the ML estimate of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$. Asymptotic properties of the ML estimators of the BeTo model parameters were also discussed by Lemonte (2022), and the author showed that $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}_{p+1}(\boldsymbol{\theta}, \boldsymbol{K}_{\boldsymbol{\theta}}^{-1})$, where $\boldsymbol{K}_{\boldsymbol{\theta}} = \text{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}}, K_\phi\}$ and so $\boldsymbol{K}_{\boldsymbol{\theta}}^{-1} = \text{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}}^{-1}, 1/K_\phi\}$, and $\boldsymbol{K}_{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{V} \boldsymbol{X}$ and $K_\phi = \text{tr}(\boldsymbol{P})$. It is interesting to note that $\boldsymbol{\beta}$ and ϕ are globally orthogonal (Cox and Reid 1987) and, hence, the ML estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\phi}$ are asymptotically independent. From the above asymptotic distribution, we have that $\text{VAR}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top \boldsymbol{V} \boldsymbol{X})^{-1}$, and $\text{VAR}(\widehat{\phi}) = 1/\text{tr}(\boldsymbol{P})$, and so asymptotic confidence intervals for the BeTo model parameters can be computed in a usual manner. The reader is referred to Lemonte (2022) for more details about the BeTo regression model.

It is evident that the parametric BeTo family of discrete distributions corresponds to an interesting, tractable, and useful distribution for count data. In particular, real data illustrations presented in Lemonte (2022) have indicated that the BeTo regression model is very useful in practice and can be an interesting alternative to the well-known Poisson and negative binomial regression models. On the other hand, Lemonte (2022) has not studied diagnostic techniques in order to detect possible influential observations under the BeTo regression model. These possible influential observations may change, for example, the inference on a particular regression parameter of the BeTo regression model fitted to the data at hand. Therefore, it is quite important to consider diagnostic techniques to identify these observations under the BeTo regression model. This paper fills this gap and develops diagnostic methods for this regression model. First, we shall derive a leverage measure to identify observations with high influence on their own predicted values. Next, we shall consider the global influence starting from the case deletion proposed by Cook (1977), and then we consider the local influence method (under different perturbation schemes) developed by Cook (1986) to detect possible influential observations under the BeTo regression model. It is worth stressing that these diagnostic methods are quite applied in the statistical literature, mainly in parametric regression models, to identify possible influential observations; see, for example, Svetliza and Paula (2001), Paula (2013), Ibacache-Pulgar, Paula, and Galea (2014), Zhu, Shi, and Liu (2015); Zhu, Liu, and Shi (2016), Ferreira, Paula, and Lana (2022) and Aoki, Bustamente, and Paula

(2022), among many others. We shall also define the Pearson residuals to assess departures from the assumptions made for the BeTo regression model, particularly for the error assumptions, as well as to detect outlying observations. Finally, the diagnostic procedures developed in this paper are applied in a real data set to analyze the fitted BeTo regression model.

2. Diagnostic analysis

2.1. Leverage

Here, we follow Pregibon (1981) and provide a leverage measure for the BeTo regression model. Basically, the idea behind the concept of leverage is that of evaluating the influence of each observed response on its own predicted value. At the convergence of the iterative process (1), we have that

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{V}} \widehat{\boldsymbol{z}},$$

where $\widehat{\boldsymbol{z}} = \widehat{\boldsymbol{\eta}} + \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{W}}^{-1} (\boldsymbol{y} - \widehat{\boldsymbol{\mu}})$. Therefore, the ML estimate $\widehat{\boldsymbol{\beta}}$ can be interpreted as the least squares solution of the linear regression of $\widehat{\boldsymbol{V}}^{1/2} \widehat{\boldsymbol{z}}$ against the columns of $\widehat{\boldsymbol{V}}^{1/2} \boldsymbol{X}$. The projection matrix of the least squares solution of the linear regression of $\widehat{\boldsymbol{z}}$ against $\boldsymbol{X}$ with weights $\widehat{\boldsymbol{V}}$ is given by

$$\boldsymbol{H} = \widehat{\boldsymbol{V}}^{1/2} \boldsymbol{X} (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{V}}^{1/2},$$

which suggests the use of the elements of the main diagonal of $\boldsymbol{H}$, say h_{ii}, to detect the presence of leverage points in this weighted linear regression model (Pregibon 1981). Note that h_{ii} can be expressed as

$$h_{ii} = \widehat{v}_i \, \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{x}_i, \quad i = 1, \dots, n,$$

where

$$\widehat{v}_i = \frac{\widehat{\mu}_i}{\widehat{w}_i}, \quad \widehat{w}_i = 1 + W_0(\widehat{\mu}_i/\widehat{\phi}), \quad \widehat{\mu}_i = \exp(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}).$$

Hence, we use h_{ii} as a leverage measure for the BeTo regression model. Therefore, the plot of h_{ii} against the index of the observations may reveal observations with a high influence on their own predicted values. In addition, observations with large values for h_{ii} may also be influential on the estimated regression coefficients (see, for example, Svetliza and Paula 2001).

2.2. Global influence

In the following, we consider the case deletion approach (Cook 1977) to assess the influence of the i-th case on the ML estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$. Let $\widehat{\boldsymbol{\theta}}_{(-i)} = (\widehat{\boldsymbol{\beta}}_{(-i)}^\top, \widehat{\phi}_{(-i)})^\top$ denote the ML estimate of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ without the i-th observation in the sample. The global influence measure, i.e. the (generalized) Cook distance, to study the effect of dropping the i-th case from data on the ML estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$ can be defined by

$$C_i(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}})^\top \widehat{\boldsymbol{K}}_{\boldsymbol{\theta}} (\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}}), \quad i = 1, \dots, n,$$

where $\widehat{\boldsymbol{K}}_{\boldsymbol{\theta}} = \text{diag}\{\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X}, \text{tr}(\widehat{\boldsymbol{P}})\}$, $\widehat{\boldsymbol{V}} = \text{diag}\{\widehat{v}_1, \dots, \widehat{v}_n\}$, $\widehat{\boldsymbol{P}} = \text{diag}\{\widehat{p}_1, \dots, \widehat{p}_n\}$, and

$$\widehat{p}_i = \frac{\mathrm{e}^{W_0(\widehat{\mu}_i/\widehat{\phi})}}{\widehat{\phi} \, \widehat{w}_i} + \frac{1}{\widehat{\phi}^2} \left[\left(\frac{T_{y_i+1}(\widehat{\phi})}{T_{y_i}(\widehat{\phi})} \right)^2 - \frac{T_{y_i+2}(\widehat{\phi})}{T_{y_i}(\widehat{\phi})} \right].$$

It is worth stressing that one can avoid the estimation of $\widehat{\boldsymbol{\theta}}_{(-i)} = (\widehat{\boldsymbol{\beta}}_{(-i)}^\top, \widehat{\phi}_{(-i)})^\top$ for all $i = 1, \dots, n$ by employing the one-step approximation given by $\widehat{\boldsymbol{\theta}}_{(-i)} \simeq \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{K}}_{\boldsymbol{\theta}}^{-1} \widehat{\boldsymbol{l}}_i$, where

$$\widehat{\boldsymbol{l}}_i = \left. \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}},$$

and $\ell_i(\boldsymbol{\theta}) = \phi[1 - \mathrm{e}^{W_0(\mu_i/\phi)}] + y_i \log[W_0(\mu_i/\phi)] + \log[T_{y_i}(\phi)]$ for $i = 1, \ldots, n$. Note that the total log-likelihood function is simply given by $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$. We have that

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sqrt{v_i} R_i \boldsymbol{x}_i, \quad \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \phi} = s_i,$$

where R_i is the Pearson residual given by

$$R_i = \frac{y_i - \mu_i}{\sqrt{\mu_i w_i}}, \quad i = 1, \ldots, n.$$

Thus, we have that

$$\widehat{\boldsymbol{\theta}}_{(-i)} - \widehat{\boldsymbol{\theta}} \simeq - \begin{bmatrix} (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} & \boldsymbol{0}_p \\ \boldsymbol{0}_p^\top & 1/\mathrm{tr}(\widehat{\boldsymbol{P}}) \end{bmatrix} \begin{pmatrix} \sqrt{\widehat{v}_i} \widehat{R}_i \boldsymbol{x}_i \\ \widehat{s}_i \end{pmatrix},$$

where

$$\widehat{R}_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i \widehat{w}_i}}, \quad \widehat{s}_i = -\mathrm{e}^{W_0(\widehat{\mu}_i/\widehat{\phi})} - \frac{(y_i - \widehat{\mu}_i)}{\widehat{\phi} \widehat{w}_i} + \frac{T_{y_i+1}(\widehat{\phi})}{\phi \, T_{y_i}(\widehat{\phi})},$$

and so

$$C_i(\boldsymbol{\theta}) = \begin{pmatrix} \sqrt{\widehat{v}_i} \widehat{R}_i \boldsymbol{x}_i \\ \widehat{s}_i \end{pmatrix}^\top \begin{bmatrix} (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} & \boldsymbol{0}_p \\ \boldsymbol{0}_p^\top & 1/\mathrm{tr}(\widehat{\boldsymbol{P}}) \end{bmatrix} \begin{pmatrix} \sqrt{\widehat{v}_i} \widehat{R}_i \boldsymbol{x}_i \\ \widehat{s}_i \end{pmatrix}, \quad i = 1, \ldots, n.$$

It then follows that

$$C_i(\boldsymbol{\theta}) = \left\{ \widehat{v}_i \widehat{R}_i^2 \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{x}_i \right\} + \left\{ \frac{\widehat{s}_i^2}{\mathrm{tr}(\widehat{\boldsymbol{P}})} \right\}, \quad i = 1, \ldots, n.$$

It is immediate to note that we can separate this distance into two components: $C_i(\boldsymbol{\theta}) = C_i(\boldsymbol{\beta}) + C_i(\phi)$, where $C_i(\boldsymbol{\beta})$ corresponds to the β-component, and $C_i(\phi)$ corresponds to the ϕ-component, with

$$C_i(\boldsymbol{\beta}) = \widehat{v}_i \widehat{R}_i^2 \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{x}_i, \quad C_i(\phi) = \frac{\widehat{s}_i^2}{\mathrm{tr}(\widehat{\boldsymbol{P}})}.$$

Hence, the plot of $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$ and $C_i(\phi)$ against the index of the observations may reveal possible influential observations on the ML estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\phi}$, respectively. It is interesting to note that $C_i(\boldsymbol{\beta}) = \widehat{R}_i^2 h_{ii}$, and so a case with a large value of Pearson residual (in absolute value) and/or a large value for h_{ii} may be an influential observation on the ML estimate $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^\top$.

2.3. Local influence

As early mentioned, the total log-likelihood function of the BeTo regression model is given by $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ is the $(p+1)$-vector of parameters, and $\ell_i(\boldsymbol{\theta}) = \phi[1 - \mathrm{e}^{W_0(\mu_i/\phi)}] + y_i \log[W_0(\mu_i/\phi)] + \log[T_{y_i}(\phi)]$ denotes the contribution of the i-th observation. Let $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ denote the perturbed log-likelihood function under a particular perturbation scheme applied in the model or data, where $\boldsymbol{\omega}$ is a q-vector of perturbations restricted to some open subset $\boldsymbol{\Omega} \subset \mathbb{R}^q$. Let $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ denote the no-perturbation vector so that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$. To measure the discrepancy between $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ and $\ell(\boldsymbol{\theta})$, one can consider the likelihood displacement given by $LD(\boldsymbol{\omega}) = 2[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})]$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the ML estimate of $\boldsymbol{\theta}$ obtained from the perturbed log-likelihood function $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. The local influence method developed by Cook (1986) considers the influence of small perturbations in the model or data on the measure $LD(\boldsymbol{\omega})$. Basically, the main idea of the local influence approach is to study the normal curvatures for $\boldsymbol{\beta}$ and ϕ in the unitary direction, $\boldsymbol{d}$ say, available at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\phi})^\top$ and $\boldsymbol{\omega}_0$. Such curvatures, for large n, can be expressed, respectively, in the forms (Cook 1986)

$$C_{\boldsymbol{d}}(\boldsymbol{\beta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{K}}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \boldsymbol{d}| \quad \text{and} \quad C_{\boldsymbol{d}}(\phi) = 2|\widehat{K}_\phi^{-1} \boldsymbol{d}^\top \boldsymbol{\Delta}_\phi^\top \boldsymbol{\Delta}_\phi \boldsymbol{d}|,$$

where $\widehat{K}_{\beta} = X^{\top}\widehat{V}X$, $\widehat{K}_{\phi} = \mathrm{tr}(\widehat{P})$, and

$$\Delta_{\beta} = \frac{\partial^2 \ell(\theta|\omega)}{\partial\beta\partial\omega^{\top}}, \quad \Delta_{\phi} = \frac{\partial^2 \ell(\theta|\omega)}{\partial\phi\partial\omega^{\top}}.$$

Note that Δ_{β} and Δ_{ϕ} are $p \times n$ and $1 \times n$ matrices, respectively. These matrices are evaluated at the ML estimate $\widehat{\theta} = (\widehat{\beta}^{\top}, \widehat{\phi})^{\top}$, and at the no-perturbation vector ω_0.

Case-weight perturbation In this perturbation scheme, it is assumed that $\ell(\theta|\omega) = \sum_{i=1}^{n} \omega_i \ell_i(\theta)$, where $0 \leq \omega_i \leq 1$ and $\omega_0 = (1, \ldots, 1)^{\top}$. After some algebra, we have that

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial\beta_r\partial\omega_i} = \sqrt{v_i}R_i x_{ir}, \quad \frac{\partial^2 \ell(\theta|\omega)}{\partial\phi\partial\omega_i} = s_i, \quad r = 1, \ldots, p, \quad i = 1, \ldots, n.$$

Let $\widehat{R} = \mathrm{diag}\{\widehat{R}_1, \ldots, \widehat{R}_n\}$ and $\widehat{S} = (\widehat{s}_1, \ldots, \widehat{s}_n)$. Hence, we can express $\Delta_{\beta} = X^{\top}\widehat{V}^{1/2}\widehat{R}$, and $\Delta_{\phi} = \widehat{S}$. It follows that $C_d(\beta) = 2|d^{\top}\widehat{R}\widehat{V}^{1/2}X(X^{\top}\widehat{V}X)^{-1}X^{\top}\widehat{V}^{1/2}\widehat{R}d|$ and $C_d(\phi) = 2|\widehat{C}_{\phi}d^{\top}\widehat{S}^{\top}\widehat{S}d|$, where $\widehat{C}_{\phi} = 1/\mathrm{tr}(\widehat{P})$. The sensitivity of the ML estimates $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^{\top}$ and $\widehat{\phi}$ under the case-weight perturbation scheme can assessed by considering the largest directions $d = d_{\max}$ in $C_d(\beta)$ and $C_d(\phi)$, which are given by the eigenvectors corresponding to the largest eigenvalues of the matrices $\widehat{R}\widehat{V}^{1/2}X(X^{\top}\widehat{V}X)^{-1}X^{\top}\widehat{V}^{1/2}\widehat{R}$ and $\widehat{C}_{\phi}\widehat{S}^{\top}\widehat{S}$, respectively.

Explanatory variable perturbation Under this perturbation scheme, the values of the j-th continuous explanatory variable are perturbed as $x_{ij\omega} = x_{ij} + \sigma_x \omega_i$, where σ_x is a scale factor which is given by the (sample) standard deviation of x_j, $\omega_i \in \mathbb{R}$, and $\omega_0 = (0, \ldots, 0)^{\top}$. This perturbation scheme leads to $\ell(\theta|\omega) = \sum_{i=1}^{n} \ell_{i\omega}(\theta)$, where $\ell_{i\omega}(\theta) = \phi[1 - e^{W_0(\mu_{i\omega}/\phi)}] + y_i \log[W_0(\mu_{i\omega}/\phi)] + \log[T_{y_i}(\phi)]$, $\mu_{i\omega} = \exp(x_{i\omega}^{\top}\beta)$, and $x_{i\omega}^{\top} = (x_{i1}, \ldots, x_{ij\omega}, \ldots, x_{in})$. After some algebraic manipulations, we have that

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial\beta_r\partial\omega_i} = \frac{\sigma_x(y_i - \mu_{i\omega})}{1 + W_0(\mu_{i\omega}/\phi)} - \frac{\sigma_x\beta_j\mu_{i\omega}(x_{ij} + \sigma_x\omega_i)}{1 + W_0(\mu_{i\omega}/\phi)}$$
$$- \frac{\sigma_x\beta_j(y_i - \mu_{i\omega})(x_{ij} + \sigma_x\omega_i)W_0(\mu_{i\omega}/\phi)}{[1 + W_0(\mu_{i\omega}/\phi)]^3},$$

where $r = 1, \ldots, p$ and $r = j$. Also, for $r = 1, \ldots, p$ and $r \neq j$, it follows that

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial\beta_r\partial\omega_i} = -\frac{\sigma_x\beta_j\mu_{i\omega}x_{ir}}{1 + W_0(\mu_{i\omega}/\phi)} - \frac{\sigma_x\beta_j(y_i - \mu_{i\omega})x_{ir}W_0(\mu_{i\omega}/\phi)}{[1 + W_0(\mu_{i\omega}/\phi)]^3}.$$

We also obtain

$$\frac{\partial^2 \ell(\theta|\omega)}{\partial\phi\partial\omega_i} = \frac{\sigma_x\beta_j(y_i - \mu_{i\omega})W_0(\mu_{i\omega}/\phi)}{\phi[1 + W_0(\mu_{i\omega}/\phi)]^3}.$$

Let $\widehat{M} = \mathrm{diag}\{\widehat{m}_1, \ldots, \widehat{m}_n\}$ and $\widehat{Z} = \mathrm{diag}\{\widehat{z}_1, \ldots, \widehat{z}_n\}$, where

$$\widehat{m}_i = \frac{\widehat{\mu}_i}{\widehat{w}_i} + \frac{W_0(\widehat{\mu}_i/\widehat{\phi})(y_i - \widehat{\mu}_i)}{\widehat{w}_i^3}, \quad \widehat{z}_i = \frac{W_0(\widehat{\mu}_i/\widehat{\phi})}{\widehat{\phi}\widehat{w}_i^3}.$$

Hence, we can express $\Delta_{\beta} = \sigma_x c_{\beta}^{(j)}(y - \widehat{\mu})^{\top}\widehat{W} - \sigma_x\widehat{\beta}_j X^{\top}\widehat{M}$ and $\Delta_{\phi} = \sigma_x\widehat{\beta}_j(y - \widehat{\mu})^{\top}\widehat{Z}$, where $\widehat{W} = \mathrm{diag}\{\widehat{w}_1, \ldots, \widehat{w}_n\}$, $\widehat{\mu} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_n)^{\top}$, $c_{\beta}^{(j)}$ denotes a p-vector with one at the j-th position and zero elsewhere, and $\widehat{\beta}_j$ (for $j = 1, \ldots, p$) denotes the j-th element of $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^{\top}$. The normal curvatures for β and ϕ in the unitary direction d are $C_d(\beta) = 2|d^{\top}\Delta_{\beta}^{\top}(X^{\top}\widehat{V}X)^{-1}\Delta_{\beta}d|$ and $C_d(\phi) = 2|\widehat{C}_{\phi}d^{\top}\Delta_{\phi}^{\top}\Delta_{\phi}d|$, respectively. Therefore, under the perturbation of the j-th continuous explanatory variable, the sensitivity of the ML estimates $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^{\top}$ and $\widehat{\phi}$ can assessed by considering the largest directions $d = d_{\max}$ in $C_d(\beta)$

and $C_{\boldsymbol{d}}(\phi)$, which are given by the eigenvectors corresponding to the largest eigenvalues of the matrices $\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{\top}(\boldsymbol{X}^{\top}\widehat{\boldsymbol{V}}\boldsymbol{X})^{-1}\boldsymbol{\Delta}_{\boldsymbol{\beta}}$ and $\widehat{C}_{\phi}\boldsymbol{\Delta}_{\phi}^{\top}\boldsymbol{\Delta}_{\phi}$, respectively.

2.4. Residual analysis

The aim of the residual analysis is to identify departures from the model assumptions, as well as to assess the overall goodness-of-fit of the fitted model to the data at hand. A simple as well as natural residual is the well-known Pearson residual early mentioned, which is based on the idea of subtracting off the mean and dividing by the standard deviation, that is,

$$R_i = \frac{y_i - \mu_i}{\sqrt{\mu_i w_i}}, \quad i = 1, \ldots, n.$$

Note that $\mathbb{E}(R_i) = 0$, and $\mathsf{VAR}(R_i) = 1$. However, in practice, we have to replace the unknown quantities with the estimated ones, and so we have that

$$\widehat{R}_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i \widehat{w}_i}}, \quad i = 1, \ldots, n.$$

It is worth emphasizing that the exact distribution of the Pearson residual $\widehat{R}_i$ is not known under the BeTo regression model. However, as suggested by Atkinson (1981), we may add envelopes into the quantile-quantile plot (QQ-plot) of the Pearson residuals to decide whether the observed residuals are consistent with the fitted BeTo regression model, as well as to detect outlying observations. Hence, if some residual values fall outside the simulated envelope, there is evidence against the adequacy of the fitted BeTo regression model to the data.

3. Application

As illustration, we shall consider a data set from Brockmann (1996). The data correspond to a study on horseshoe crabs (*Limulus polyphemus*). Let the number of males surrounding a breeding female (Y) be the response variable; that is, the number of additional males ("satellites") residing nearby each female that may fertilize her eggs. Explanatory variables are the female crab's color (1–light; 2–medium; 3–dark; and 4–darker); the female crab's spine condition (1–good; 2–middle; and 3–bad); and the female's weight (in Kg). The sample size is $n = 173$. Boxplots for the satellites according to the female's color and female's spine condition are presented in Figure 1. Note that there are a few extreme observations.
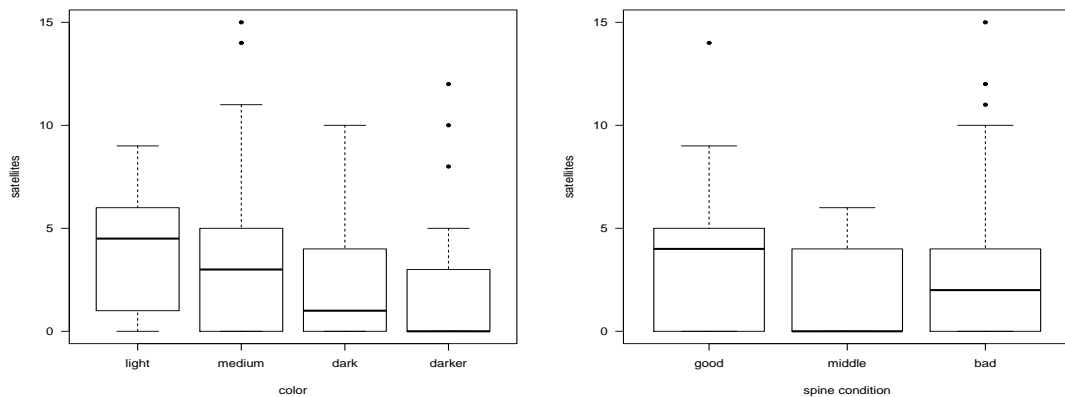

Figure 1: Boxplots of the satellites according to the female crab's color (left), and female crab's spine condition (right)

We assume $Y_{ijk} \sim \mathrm{BeTo}(\mu_{ijk}, \phi)$ so that $\log(\mu_{ijk}) = \beta_1 + \beta_2 \, \texttt{weight}_i + \beta_{3j} + \beta_{4k}$, where μ_{ijk} is the expected number of additional males ("satellites") corresponding to the j-th female's

Table 1: Parameter estimates

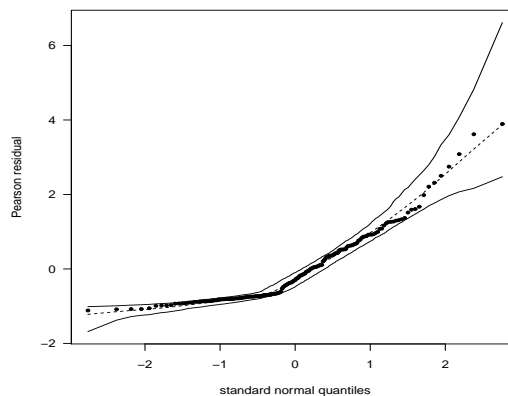| Effect	Estimate	SE	Estimate/SE	*p*-value
Intercept	−0.1042	0.4801	−0.2171	0.8281
weight	0.5781	0.1389	4.1630	0.0000
color(medium)	−0.2808	0.3181	−0.8828	0.3774
color(dark)	−0.5398	0.3647	−1.4801	0.1388
color(darker)	−0.5525	0.4186	−1.3200	0.1868
spine(middle)	−0.1807	0.3848	−0.4696	0.6386
spine(bad)	0.0828	0.2232	0.3708	0.7108

Figure 2: Pearson residuals with simulated envelope

color ($j = 1, 2, 3, 4$) and k-th female's spine condition ($k = 1, 2, 3$) for $i = 1, \ldots, 173$. Here, we assume that $\beta_{31} = \beta_{41} = 0$. The ML estimates of the regression parameters, the corresponding asymptotic standard errors (SE), and the *p*-values are listed in Table 1. The ML estimate of the precision parameter (asymptotic SE between parentheses) is $\widehat{\phi} = 0.1042\,(0.0376)$. The QQ-plot for the Pearson residuals with simulated envelopes (Figure 2) against the theoretical quantiles of the standard normal distribution does not present any unusual features, as well as there are no observations outside the envelope. Therefore, the BeTo regression model seems to be appropriate to model the horseshoe crabs data. From Table 1, note that only the female's weight seems to be statistically significant (based on the 5% nominal level, for instance) on the expected number of additional males ("satellites") that may fertilize her eggs.

We now check the presence of possible influential observations. To do that, we generate plots for the diagnostic measures derived in the previous section. Figure 3 presents the index plot of the leverage measure, where case 141 seems to have a great influence on its own predicted value. This case corresponds to a breeding female crab with 5.2Kg (being the weightiest female crab), medium color, and good spine condition, which shows the difficulty of predicting the response for weigher female crabs. Index plots of $C_i(\boldsymbol{\beta})$ and $C_i(\phi)$ are displayed in Figure 4. Figure 5 presents the index plot of the local influence under the case-weight perturbation, while Figure 6 presents the index plot of the local influence under the covariate (female crab weight) perturbation. On the basis of these plots, the cases 15, 56, 117, 134, 141, 146 and 149 have more pronounced influence than the other observations. It is worth mentioning that a common feature among these cases is that all correspond to female crabs which have a great number of males surrounding them. For example, cases 15, 56, and 117 have 14, 15, and 12 males ("satellites") residing nearby the female, respectively, and the sample mean of males is (approximately) 3.

After finding possible influential observations from the generated plots for the diagnostic measures derived in this paper, an interesting problem is to determine how to deal with
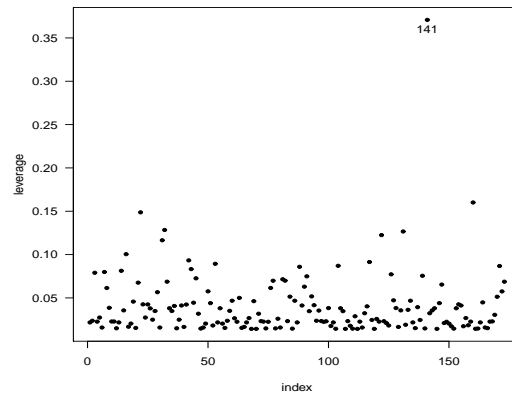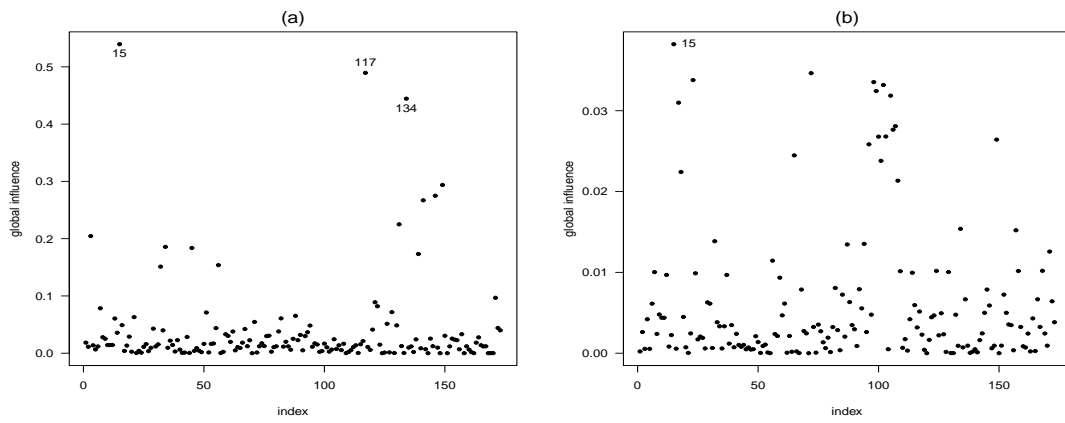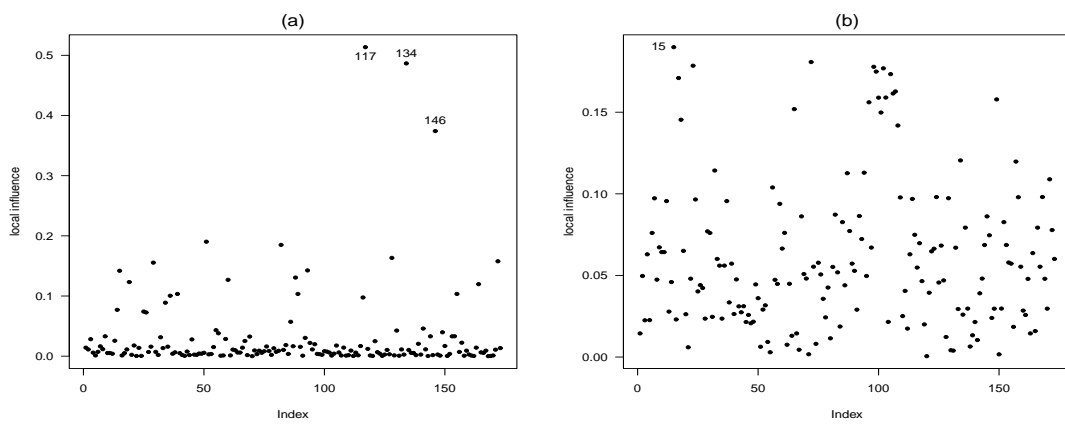
Figure 3: Leverage

Figure 4: Index plots of $C_i(\boldsymbol{\beta})$ (a), and $C_i(\phi)$ (b)

Figure 5: Index plots of $|\boldsymbol{d}_{\max}|$ under the case-weight perturbation for $\widehat{\boldsymbol{\beta}}$ (a), and $\widehat{\phi}$ (b)
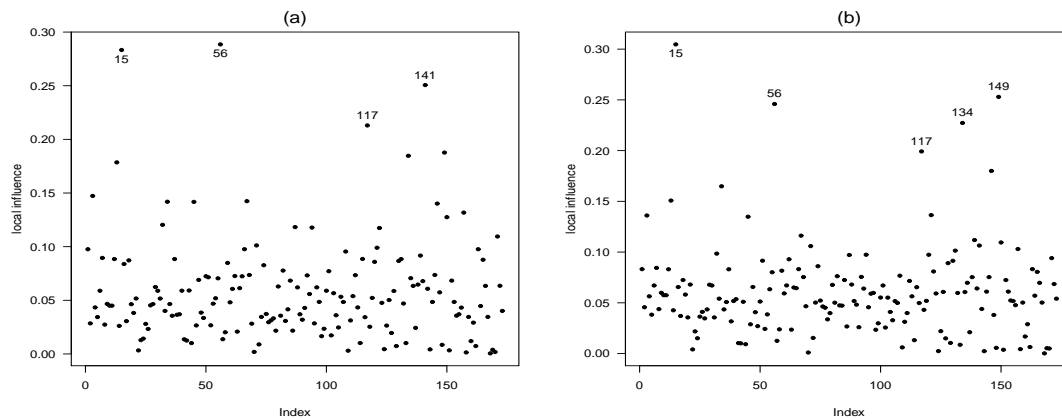
Figure 6: Index plots of $|\boldsymbol{d}_{\max}|$ under the covariate (female crab weight) perturbation for $\widehat{\boldsymbol{\beta}}$ (a), and $\widehat{\phi}$ (b)

these influential observations, that is, how to measure the effect of these observations on the parameter estimates or on the fitted model, for example. First, we have dropped each of the outstanding influential observations (cases 15, 56, 117, 134, 141, 146, and 149) and have refitted the BeTo regression model to verify their impact on the parameter estimates. Table 2 lists the relative changes of the ML estimates after removing each of the influential observation from the sample. (The relative change (in %) of an estimate $\widehat{\xi}$ of ξ is given by $(\widehat{\xi} - \widehat{\xi}_{(-i)})/\widehat{\xi} \times 100\%$, where $\widehat{\xi}_{(-i)}$ denotes the ML estimate of ξ after removing the i-th observation from the sample.) It is noteworthy that the intercept parameter was the most affected by the influential observations, followed by the parameters regarding the spine condition effect (i.e., β_{42} and β_{43}). For example, the elimination of case 15 delivers relative changes (approximately) of -169%, 78%, and -177% for β_1, β_{42}, and β_{43}, respectively. However, the elimination of these observations does not change the inference regarding the regression coefficients; that is, the covariate which is statistically significant remains significant, and the covariate that is not statistically significant remains not significant. Next, Table 3 lists the log-likelihood function evaluated at the ML estimates ($\widehat{\ell}$), the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Hannan-Quinn information criterion (HQIC) after dropping each of the outstanding influential observations (cases 15, 56, 117, 134, 141, 146, and 149). From Table 3, note that each of the outstanding influential observations has a slight impact on the BeTo regression model fitted to the data at hand; that is, the comparison criteria are near, and there is no notable improvement on the model fitting after removing each of the outstanding influential observations in comparison which the fitted model with all observations and, hence, the BeTo regression model can incorporate these outstanding influential observations.

In summary, the above analysis demonstrates the necessity of conducting a diagnostic analysis in the BeTo regression model to identify possible influential observations in the sample. Finally, on the basis of the above analysis, we assume that $Y_i \sim \mathrm{BeTo}(\mu_i, \phi)$, where $\log(\mu_i) = \beta_1 + \beta_2\,\texttt{weight}_i$ for $i = 1, \ldots, 173$. The ML estimates (asymptotic SE between parentheses) are $\widehat{\beta}_1 = -0.5233\,(0.3401)$, $\widehat{\beta}_2 = 0.6249\,(0.1256)$, and $\widehat{\phi} = 0.0994\,(0.0356)$; that is, $\widehat{\mu}_i = \exp(-0.5233 + 0.6249\,\texttt{weight}_i)$ for $i = 1, \ldots, 173$. The BeTo regression model admits the following parameter interpretation of the final selected regression model: the expected number of additional males ("satellites") surrounding a breeding female increases (approximately) 37% $[\mathrm{e}^{0.6249 \times (1/2)} - 1) \times 100\%]$ as the female's weight increases $\frac{1}{2}$Kg, for example.

4. Concluding remarks

The two-parameter discrete Bell-Touchard family of distributions introduced by Castellares

Table 2: Relative changes (in %) on the parameter estimates

Removed	β_1	β_2	β_{32}	β_{33}
15	−169.58	−9.48	−44.23	−16.93
56	60.94	3.70	−2.03	5.65
117	137.52	8.89	5.48	1.07
134	−4.03	−0.26	1.32	0.89
141	−249.14	−16.88	15.14	9.08
146	−19.96	−1.28	0.53	0.67
149	−84.17	−5.38	1.35	−16.08
Removed	β_{34}	β_{42}	β_{43}	ϕ
15	−16.24	78.64	−177.66	−8.45
56	5.57	−11.95	58.57	−4.13
117	−37.45	−16.55	30.56	−3.36
134	−36.52	2.32	7.13	−5.47
141	12.74	0.77	44.69	−0.29
146	−27.73	4.10	2.66	−3.36
149	3.62	20.31	5.56	−6.91

Table 3: Model selection

Removed	$\widehat{\ell}$	AIC	BIC	HQIC
None	−358.36	732.72	757.94	742.95
15	−351.76	719.52	744.70	729.74
56	−353.14	722.27	747.45	732.49
117	−353.59	723.17	748.35	733.39
134	−353.07	722.15	747.33	732.36
141	−355.03	726.07	751.25	736.28
146	−353.93	723.86	749.04	734.07
149	−352.67	721.34	746.52	731.55

et al. (2020) and reparameterized in Lemonte (2022), and its associated regression model for count response variables, has proved to be useful to deal with overdispersion in count data. We have derived useful diagnostic methods for the Bell-Touchard class of regression models. It is worth stressing that all diagnostic measures developed in this paper are simple, compact, and can easily be implemented in statistical software, once only simple matrix operations have to be performed. We have also considered the diagnostic measures developed in a real data set to illustrate their usefulness in finding possible influential observations when using the Bell-Toucherd regression model in practice. Finally, we have developed R codes (Team 2022) to perform the diagnostic measures regarding the Bell-Touchard regression model in Section 3. The R code may be obtained upon request.

Acknowledgments

References

Aoki R, Bustamente JPM, Paula GA (2022). "Local Influence Diagnostics with Forward Search in Regression Analysis." *Statistical Papers*, **63**, 1477–1497. URL https://doi.org/10.1007/s00362-021-01279-4.

Atkinson AC (1981). "Two Graphical Display for Outlying and Influential Observations in Regression." *Biometrika*, **68**, 13–20. URL https://doi.org/10.2307/2335801.

Brockmann HJ (1996). "Satellite Male Groups in Horseshoe Crabs, *Limulus polyphemus*." *Ethology*, **102**, 1–21. URL https://doi.org/10.1111/j.1439-0310.1996.tb01099.x.

Castellares F, Lemonte AJ, Moreno-Arenas G (2020). "On the Two-Parameter Bell–Touchard Discrete Distribution." *Communications in Statistics - Theory and Methods*, **49**, 4834–4852. URL https://doi.org/10.1080/03610926.2019.1609515.

Comtet L (2010). *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. Reidel Publishing Company. ISBN 978-94-010-2198-2, URL https://doi.org/10.1007/978-94-010-2196-8.

Cook RD (1977). "Detection of Influential Observation in Linear Regression." *Technometrics*, **19**, 15–18. URL https://doi.org/10.2307/1268249.

Cook RD (1986). "Assessment of Local Influence." *Journal of the Royal Statistical Society B*, **48**, 133–169. URL https://doi.org/10.1111/j.2517-6161.1986.tb01398.x.

Corless RM, Gonnet GH, Hare DEG, Jeffrey D, Knuth DE (1996). "On the Lambert W Function." *Advances in Computational Mathematics*, **5**, 329–359. URL https://doi.org/10.1007/BF02124750.

Cox DR, Reid N (1987). "Parameter Orthogonality and Approximate Conditional Inference." *Journal of the Royal Statistical Society B*, **40**, 1–39. URL https://doi.org/10.1111/j.2517-6161.1987.tb01422.x.

Ferreira CS, Paula GA, Lana GC (2022). "Estimation and Diagnostic for Partially Linear Models with First-Order Autoregressive Skew-Normal Errors." *Computational Statistics*, **37**, 445–468. URL https://doi.org/10.1007/s00180-021-01130-2.

Ibacache-Pulgar G, Paula GA, Galea M (2014). "On Influence Diagnostics in Elliptical Multivariate Regression Models with Equicorrelated Random Errors." *Statistical Methodology*, **16**, 14–31. URL https://doi.org/10.1016/j.stamet.2013.06.001.

Lemonte AJ (2022). "On the Mean-Parameterized Bell-Touchard Regression Model for Count Data." *Applied Mathematical Modelling*, **105**, 1–16. URL https://doi.org/10.1016/j.apm.2021.12.035.

Paula GA (2013). "On Diagnostics in Double Generalized Linear Models." *Computational Statistics and Data Analysis*, **68**, 44–51. URL https://doi.org/10.1016/j.csda.2013.06.008.

Pregibon D (1981). "Logistic Regression Diagnostics." *Annals of Statistics*, **9**, 705–724. URL https://doi.org/10.1214/aos/1176345513.

Svetliza CF, Paula GA (2001). "On Diagnostics in Log-Linear Negative Binomial Models." *Journal of Statistical Computation and Simulation*, **71**, 231–244. URL https://doi.org/10.1080/00949650108812145.

Team RC (2022). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Touchard J (1933). "Propriétés Arihtmétiques de Certains Nombres Récurrents." *Annales de la Société Scientifique de Bruxelles*, **53**, 21–31.

Zhu F, Liu S, Shi L (2016). "Local Influence Analysis for Poisson Autoregression with an Application to Stock Transaction Data." *Statistica Neerlandica*, **70**, 4–25. URL https://doi.org/10.1111/stan.12071.

Zhu F, Shi L, Liu S (2015). "Influence Diagnostics in Log-Linear Integer-Valued GARCH Models." *AStA Advances in Statistical Analysis*, **99**, 311–335. URL https://doi.org/10.1007/s10182-014-0242-4.

Affiliation:

Artur J. Lemonte
Departamento de Estatística, CCET
Universidade Federal do Rio Grande do Norte
Lagoa Nova, Natal/RN 59078970, Brazil
E-mail: arturlemonte@gmail.com