Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 2, 402–424
https://doi.org/10.1214/18-BJPS395

Influence measures for the Waring regression model

Luisa Rivas[a] **and Manuel Galea**[b]

[a]*Universidad de Concepción*
[b]*Pontificia Universidad Católica de Chile*

Abstract. In this paper, we present a regression model where the response variable is a count data that follows a Waring distribution. The Waring regression model allows for analysis of phenomena where the Geometric regression model is inadequate, because the probability of success on each trial, p, is different for each individual and p has an associated distribution. Estimation is performed by maximum likelihood, through the maximization of the Q-function using EM algorithm. Diagnostic measures are calculated for this model. To illustrate the results, an application to real data is presented. Some specific details are given in the Appendix of the paper.

1 Introduction

It is common that, in diverse statistical applications, the response variable corresponds to count. The possible distributions that can describe the behavior of this type of data are varied. One of them is the Geometric distribution, which has been widely studied in fertility applications; for example, Weinberg and Gladen (1986) consider that the number of menstrual cycles necessary for conception has a Geometric distribution with parameter p. They mention that, in this situation, this parameter is not necessarily constant, i.e., it varies from individual to individual. With this, they have suggested that the Geometric regression model is not suitable when there is heterogeneity in the study. Furthermore, they assume that the parameter p, $0 < p < 1$ follows a Beta distribution, thus generating the Beta-Geometric mixture, known later as the Waring distribution. In this context, p has a very clear interpretation, since it represents the probability of a conception occurring for any randomly-chosen couple. This idea was originally designed by Henry (1957). This model was also previously studied by Miller (1961), who proposes that the number of vehicles in a traffic jam follows a Geometric distribution, and that the passing rate has a Beta distribution. Pielou (1962) uses this distribution to analyze the behavior of certain plant species, where the parameter p represents the proportion of segregation between species. Later, Ridout and Morgan (1991) extended the research by Weinberg and Gladen (1986) and used that said model to separate the sample in women who smoke and who do not smoke, and in women who take birth

Key words and phrases. EM algorithm, beta-geometric distribution, generalized Cook's distance, appropriate perturbation, global and local influence.
Received July 2017; accepted February 2018.

control and those who do not. The main objective was to report expected values and to conclude that women who smoke reduce their fertility. This conclusion was concordant to Baird and Wilcox's (1985) study, who carried out an epidemiologic study with the same objective. Since the Waring model has one more parameter than the Geometric model, it is necessary to analyze if this parameter is significant. With this purpose, Paul (2005) proposed a test that can justify the use of one or another distribution, jointly with other goodness of fit measures. This test is based on the statistic of ratio of likelihood.

Recently, Singh, Pudir and Maheshwari (2014) obtained the moment estimators, of maximum likelihood and the estimators from a bayesian point of view. They compare them, for which two applications have been presented in the study of fertility: the first of them in a data set from the government of India regarding health statistics, and the second of them in a simulated data set. Extensions of the Waring regression are presented by Rodríguez-Avi et al. (2007, 2009).

An important aspect in the analysis of a data set is the influence diagnostics, because it allows to identify a lack of adjustment or the presence of influential observations. Zhu et al. (2001) and Zhu and Lee (2001) proposed a diagnostic approach based on the Cook proposal (Cook (1986)), but using the Q-function, this method is called global and local influence. The first approach consists of the study the change in MLE by eliminating an observation of the data set, while the second evaluates the effect of small perturbations in the model and/or data on the parameter estimates. Several authors have studied the Waring regression model, however, neither of them have referred to the influence analysis for this model. Thus, the main objective of this work is to develop estimation method and diagnostics analysis, based on case delation and the local influence approach for the Waring regression model through the Q-function.

This paper is organized as follows. In Section 2, we present the Waring regression models and discussed some of their properties. In Section 3, we discussed the EM algorithm to calculated the maximum likelihood estimators. Diagnostic measures are discussed in Section 4. Section 5 contains applications of the proposed Waring regression model. Concluding remarks are given in Section 6. Some technical details are present in the Appendix.

2 The Waring regression model

The Waring regression model with parameters α and $\beta > 0$, denoted by Waring(α, β), is used to describe count data, y_i, which assumes nonnegative integer values and whose probability function has been presented by several authors, who use different parameterizations. In this paper, we used the proposal by Singh, Pudir and Maheshwari (2014), which was previously used by Crouchley and Dassios (1998), but in the context of censored data. Then, the probability function is given by

$$P(y_i|\alpha,\beta) = \frac{\alpha\Gamma(\alpha+\beta)\Gamma(y_i+\beta)}{\Gamma(\beta)\Gamma(y_i+\alpha+\beta+1)}, \qquad y_i \in \{0,1,2,\dots\}. \tag{2.1}$$

Singh, Pudir and Maheshwari (2014) also indicate that (2.1) is a parametrization of the probability function originally proposed by Weinberg and Gladen (1986).

The mean and the variance of the variable $y_i \sim \text{Waring}(\alpha,\beta)$ are, respectively

$$\begin{aligned} \mu_i = \mu = E(y_i) &= \frac{\beta}{\alpha-1}, \qquad \alpha > 1,\\ \text{Var}(y_i) &= \frac{\alpha\beta(\alpha+\beta-1)}{(\alpha-2)(\alpha-1)^2}, \qquad \alpha > 2. \end{aligned} \tag{2.2}$$

The model (2.1), in terms of mixture of distributions, can be written as

$$\begin{cases} y_i|p_i \sim \text{Geo}(p_i),\\ p_i \sim \text{Beta}(\alpha,\beta), \end{cases} \tag{2.3}$$

where $\text{Geo}(p_i)$ is the Geometric distribution with probability function

$$P(y_i|p_i) = p_i(1-p_i)^{y_i}, \qquad y_i \in \{0,1,2,\dots\},$$

with $0 < p_i < 1$, and whose density function is given by (Crouchley and Dassios (1998))

$$f(p_i|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1}(1-p_i)^{\beta-1}, \qquad \alpha,\beta > 0.$$

The moment generating function (MGF) of this distribution is given in the Appendix.

In this paper we propose a new parametrization for (2.1), including a regression structure, that is,, $\mu \neq \mu_i \forall i = 1,\dots,n$, then, the probability function reparameterized in terms of μ_i and ϕ is

$$P(y_i|\mu_i,\phi) = \frac{\phi_1\Gamma(\phi_1+\mu_i\phi_2)\Gamma(y_i+\mu_i\phi_2)}{\Gamma(\mu_i\phi_2)\Gamma(y_i+\phi_1+\mu_i\phi_2+1)}, \qquad y_i \in \{0,1,2,\dots\},$$

where $\phi_1 = \frac{2\phi}{\phi-1}$ and $\phi_2 = \frac{\phi+1}{\phi-1}$, with $\mu_i > 0$ and $\phi > 1$. From the latter, we can write, $y_i \sim \text{Waring}(\mu_i,\phi)$ whose mean and variance, respectively, are

$$E(y_i) = \mu_i \quad \text{and} \quad \text{Var}(y_i) = \phi(\mu_i^2+\mu_i) = \phi V(\mu_i),$$

where $V(\mu_i) = \mu_i(\mu_i+1)$ denotes the variance function. In this way, μ_i is the mean of the response variable, and ϕ can be interpreted as a dispersion parameter, that is, for fixed μ_i, if ϕ increases the variance of y_i also increases. In addition to allowing the inclusion of the regression structure, it is possible to clarify the role of each parameter in the model.

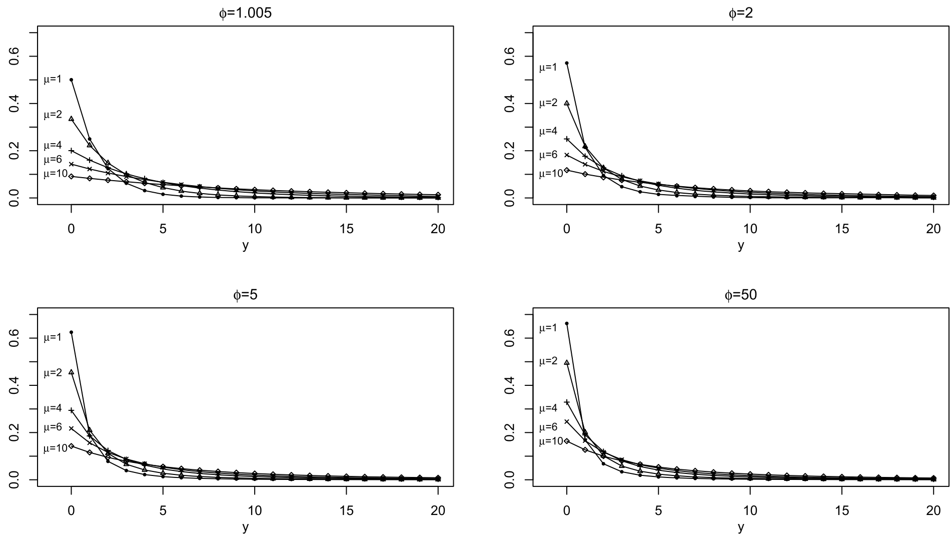

Figure 1 *Probability function of Waring distribution for different values of the parameter μ, with fixed ϕ.*

Figure 1 presents the probability function of the Waring distribution, $P(y_i|\mu, \phi)$, for different values of parameter μ with fixed ϕ; we note that when μ increases, the tail of the distribution is more heavy. This pattern is repeated for the different values of the parameter ϕ.

Figure 2 shows the probability function of the Waring distribution when μ is fixed; we observe that when ϕ decreases the tail of the distribution is more heavy. The decrease of the probability values is more drastic when μ increases.

Figure 3(a) shows the probability function of the Geometric distribution, $P(y|p)$, with parameter $p = 1/5$, where $p = 1/(1+\mu)$; compared with the Waring distribution for $\mu = 4$ and different values of ϕ, we observe that when $\phi \to 1$, the curve best fits the Geometric distribution. In this way, Figure 3(b), where $p = 1/11$. Both in the Figure 3(a) and (b) it can be seen that the Geometric distribution is a particular case of the Waring distribution, when $\phi \to 1$.

Let $y_1, y_2, \dots, y_n$ be independent random variables, where $y_i \sim \text{Waring}(\mu_i, \phi)$. The model is obtained by assuming that the mean of y_i can be written as

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i, \qquad i = 1, \dots, n,$$

in this paper in particular, we used the link function log, in which case we can write, $\mu_i = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}$.

The log-likelihood function based on a sample of n independent observations is

$$l = l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\mu_i, \phi),$$

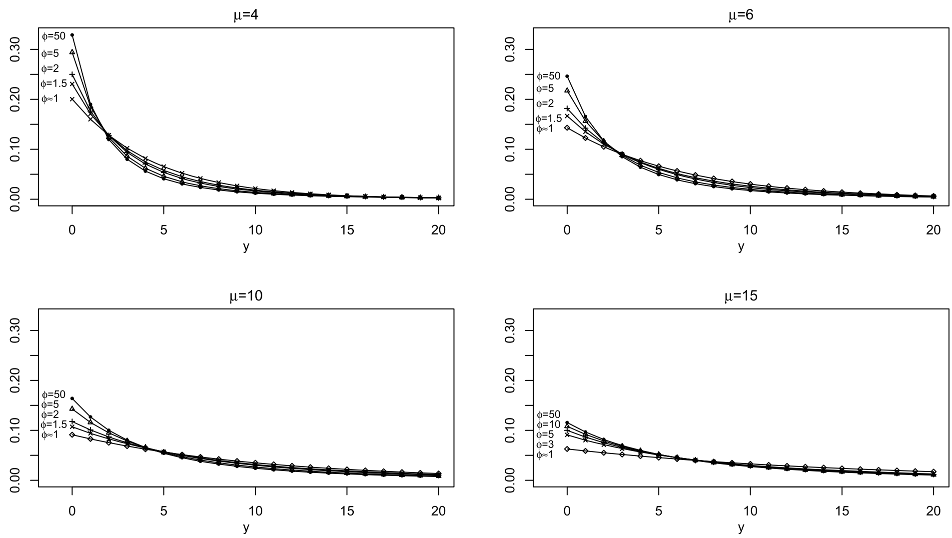

Figure 2 *Probability function of Waring distribution for different values of the parameter* ϕ, *with fixed* μ.

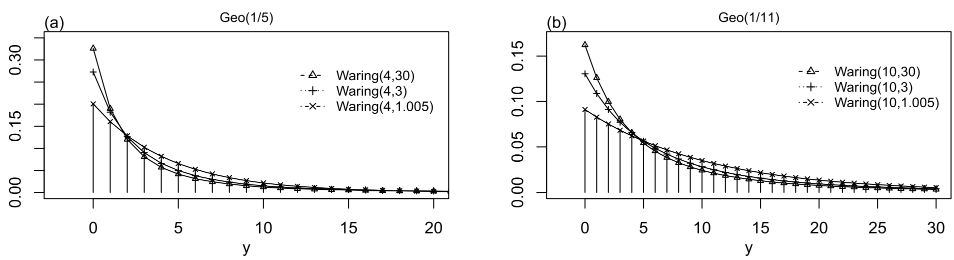

Figure 3 *Probability function of Geometric distribution*, $P(y|p)$, *versus the Waring distribution*, $P(y|\mu, \phi)$.

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ is a q-dimensional vector of unknown parameters, and the contribution of the ith observation to the log-likelihood function is equal to

$$\begin{aligned} l_i(\mu_i, \phi) = {} & \log \phi_1 + \log \Gamma(\phi_1 + \mu_i \phi_2) + \log \Gamma(y_i + \mu_i \phi_2) - \log \Gamma(\mu_i \phi_2) \\ & - \log \Gamma(y_i + \phi_1 + \mu_i \phi_2 + 1), \end{aligned} \tag{2.4}$$

for $i = 1, \ldots, n$. In this case, $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ is a function of $\boldsymbol{\beta}$, the vector of regression parameters.

On the other hand, from the Beta-Geometric mixture model, the complete-data, $\mathbf{y}_c = (\mathbf{y}_0, \mathbf{y}_m)$, the complete log-likelihood function can be written as

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^{n} \{ & \phi_1 \log p_i + (\mu_i \phi_2 + y_i - 1) \log(1 - p_i) \\ & + \log\Gamma(\phi_1 + \mu_i\phi_2) - \log\Gamma(\phi_1) - \log\Gamma(\mu_i\phi_2) \}, \end{aligned} \tag{2.5}$$

where y_i denote the observation data ($\mathbf{y}_0 = (y_1, y_2, \ldots)$) and p_i can be treated as missing data, that is, $\mathbf{y}_m = (p_1, p_2, \ldots)$.

3 Maximum likelihood estimation

We may easily obtain the maximum likelihood estimation (MLE) of $\boldsymbol{\theta}$ based on the complete-data log-likelihood function and the EM algorithm, presented by Dempster, Laird and Rubin (1977). A standard EM algorithm consists of two steps:

- **E-step** (Expectation): Compute the Q-function as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = E\{l_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_0, \boldsymbol{\theta}^{(m)}\}.$$

In fact, from (2.5) it follows that

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{n} \{ & \phi_1 E(\log p_i|\mathbf{y}_0, \boldsymbol{\theta}^{(m)}) + (\mu_i\phi_2 + y_i - 1) \\ & \cdot E(\log(1 - p_i)|\mathbf{y}_0, \boldsymbol{\theta}^{(m)}) + \log\Gamma(\phi_1 + \mu_i\phi_2) \\ & - \log\Gamma(\phi_1) - \log\Gamma(\mu_i\phi_2) \}. \end{aligned} \tag{3.1}$$

Since $p_i \sim \text{Beta}(\phi_1, \mu_i\phi_2)$, then $p_i|\mathbf{y}_0, \boldsymbol{\theta} \sim \text{Beta}(\phi_1 + 1, y_i + \mu_i\phi_2)$, then, using the properties of the Beta distribution, by Johnson, Kotz and Balakrishnan (1970), we obtain that

$$E(\log p_i|\mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = e_i \quad \text{and} \quad E(\log(1 - p_i)|\mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = s_i,$$

where $e_i = \psi(\phi_1^{(m)} + 1) - \psi(y_i + \mu_i^{(m)}\phi_2^{(m)} + \phi_1^{(m)} + 1)$ and $s_i = \psi(y_i + \mu_i^{(m)}\phi_2^{(m)}) - \psi(y_i + \mu_i^{(m)}\phi_2^{(m)} + \phi_1^{(m)} + 1)$. Substituting in (3.1) we have to

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{n} \{ & \phi_1 e_i + (\mu_i\phi_2 + y_i - 1)s_i + \log\Gamma(\phi_1 + \mu_i\phi_2) \\ & - \log\Gamma(\phi_1) - \log\Gamma(\mu_i\phi_2) \}. \end{aligned}$$

- **M-step**: Update the parameters $\boldsymbol{\theta}^{(m)}$ maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$, i.e.,

$$\boldsymbol{\theta}^{(m+1)} = \underset{\boldsymbol{\theta}}{\text{ArgMaxQ}}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}).$$

Then, $\boldsymbol{\theta}^{(m)}$ is updated using the pseudovalues e_i and s_i, where the parameters $\boldsymbol{\theta}$ are estimated through the Newton-Raphson algorithm, where la score function is given by

$$\mathbf{U}_{\boldsymbol{\theta}}(Q) = \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{U}_{\boldsymbol{\beta}}^T(Q) \\ \mathbf{U}_{\phi}(Q) \end{pmatrix}, \tag{3.2}$$

with

$$\mathbf{U}_{\boldsymbol{\beta}}(Q) = \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} = \phi_2 \mathbf{X}^T \mathrm{R}_1,$$

where R_1 is a n-dimensional vector, $\mathrm{R}_{1i} = (s_i + \psi(\phi_1 + \mu_i\phi_2) - \psi(\mu_i\phi_2))\mu_i$, $i = 1, \dots, n$, and

$$\mathbf{U}_{\phi}(Q) = \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \phi} = \frac{-2}{(\phi-1)^2}\mathrm{F}_1,$$

where $\mathrm{F}_1 = \sum_{i=1}^n [e_i + \mu_i s_i + \psi(\phi_1 + \mu_i\phi_2)(1+\mu_i) - \psi(\phi_1) - \psi(\mu_i\phi_2)\mu_i]$.

Whereas the matrix of second derivatives is given by

$$\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T} = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}(Q) & \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\phi}}(Q) \\ \mathbf{I}_{\boldsymbol{\phi}\boldsymbol{\beta}}(Q) & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(Q) \end{pmatrix}, \tag{3.3}$$

where

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}(Q) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}^T} = \phi_2 \mathbf{X}^T \,\mathrm{Diag}(\mathrm{R}_1)\mathbf{X} + \phi_2 \mathbf{X}^T \mathbf{R}_2 \mathbf{X},$$

with $\mathbf{R}_2$ being a diagonal matrix, $\mathrm{R}_{2i} = (\psi'(\phi_1 + \mu_i\phi_2) - \psi'(\mu_i\phi_2))\mu_i$, and $\mathrm{Diag}(\mathrm{R}_1)$ denotes the diagonal matrix whose elements are precisely the components of vector R_1.

$$\begin{aligned} \mathbf{I}_{\phi\phi}(Q) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \phi^2} \\ &= \frac{4}{(\phi-1)^3}\mathrm{F}_1 + \frac{4}{(\phi-1)^4}\sum_{i=1}^n [\psi'(\phi_1 + \mu_i\phi_2)(1+\mu_i)^2 - \psi'(\phi_1) \\ &\quad - \psi'(\mu_i\phi_2)\mu_i^2] \\ &= \frac{4}{(\phi-1)^3}\mathrm{F}_1 + \frac{4}{(\phi-1)^4}\mathrm{F}_2, \end{aligned}$$

with $\mathrm{F}_2 = \sum_{i=1}^n [\psi'(\phi_1 + \mu_i\phi_2)(1+\mu_i)^2 - \psi'(\phi_1) - \psi'(\mu_i\phi_2)\mu_i^2]$; and

$$\mathbf{I}_{\boldsymbol{\beta}\phi}(Q) = \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}\, \partial \phi} = -\frac{2}{(\phi-1)^2}\mathbf{X}^T \mathrm{R}_1 + \phi_2 \mathbf{X}^T \mathrm{R}_3,$$

where the elements of vector R_3 are, $\mathrm{R}_{3i} = (\psi'(\phi_1 + \mu_i\phi_2)(1 + \mu_i) - \psi'(\mu_i\phi_2)\mu_i)\mu_i$, $i = 1, \ldots, n$.

Then, by substituting into (3.3), we obtain

$$\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \phi_2\mathbf{X}^T\,\mathrm{Diag}(\mathrm{R}_1)\mathbf{X} + \phi_2\mathbf{X}^T\mathbf{R}_2\mathbf{X} & -\dfrac{2}{(\phi-1)^2}\mathbf{X}^T\mathrm{R}_1 + \phi_2\mathbf{X}^T\mathrm{R}_3 \\ -\dfrac{2}{(\phi-1)^2}\mathrm{R}_1^T\mathbf{X} + \phi_2\mathrm{R}_3^T\mathbf{X} & \dfrac{4}{(\phi-1)^3}\mathrm{F}_1 + \dfrac{4}{(\phi-1)^4}\mathrm{F}_2 \end{pmatrix}, \tag{3.4}$$

evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Then, (3.2) and (3.4) are used to estimate the vector of parameters $\boldsymbol{\theta}$. If the criterion of convergence is satisfied then we can stop iterating, else go back to **E**-step. Note that the maximum likelihood estimators of $\boldsymbol{\theta}$ are obtained from the equation $\mathbf{U}_{\boldsymbol{\theta}}(Q) = 0$, and do not have closed-form.

To obtain the asymptotic standard error of the MLEs of $\boldsymbol{\theta}$ we use $(-\mathbf{I}_{\boldsymbol{\theta}})^{-1}$, where $\mathbf{I}_{\boldsymbol{\theta}}$ is the observed-data information matrix given by

$$\mathbf{I}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\beta}\phi} \\ \mathbf{I}_{\phi\boldsymbol{\beta}} & \mathbf{I}_{\phi\phi} \end{pmatrix},$$

with

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^T} = \phi_2^2\mathbf{X}^T\,\mathrm{Diag}(\mathfrak{R}^*)\mathbf{X} + \phi_2\mathbf{X}^T\,\mathrm{Diag}(\mathfrak{R})\mathbf{X},$$

where the elements of the vectors $\mathfrak{R}$ and $\mathfrak{R}^*$ are respectively given by, $\mathfrak{R}_i = (\mathrm{a}_i - \mathrm{b}_i)\frac{1}{g'(\mu_i)}$, $\mathfrak{R}_i^* = (\mathrm{a}_i^* - \mathrm{b}_i^*)\frac{1}{g'(\mu_i)^2}$, with $\mathrm{a}_i = \psi(\phi_1 + \mu_i\phi_2) - \psi(y_i + \phi_1 + \mu_i\phi_2 + 1)$, $\mathrm{b}_i = \psi(\mu_i\phi_2) - \psi(y_i + \mu_i\phi_2)$, $\mathrm{a}_i^* = \psi'(\phi_1 + \mu_i\phi_2) - \psi'(y_i + \phi_1 + \mu_i\phi_2 + 1)$, $\mathrm{b}_i^* = \psi'(\mu_i\phi_2) - \psi'(y_i + \mu_i\phi_2)$ and $\psi'(\cdot)$ denote the trigamma function and $l(\boldsymbol{\theta})$ is given by (2.4)

$$\mathbf{I}_{\boldsymbol{\beta}\phi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\,\partial\phi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\phi\,\partial\boldsymbol{\beta}^T} = \mathbf{I}_{\phi\boldsymbol{\beta}}^T.$$

Then,

$$\mathbf{I}_{\boldsymbol{\beta}\phi} = \frac{-2}{(\phi-1)^2}\mathbf{X}^T\mathrm{J}_3,$$

with J_3 being a n-dimensional vector with elements given by $\mathrm{J}_{3i} = [(\mathrm{a}_i - \mathrm{b}_i) + \phi_2(\mathrm{a}_i^*(1 + \mu_i) - \mathrm{b}_i^*\mu_i)]\frac{1}{g'(\mu_i)}$, and

$$\begin{aligned}\mathbf{I}_{\phi\phi} &= \frac{\partial^2 l(\boldsymbol{\beta},\phi)}{\partial\phi^2} \\ &= \sum_{i=1}^n\biggl(-\frac{1}{\phi^2} + \frac{1}{(\phi-1)^2} + 4\frac{(\mu_i+1)^2}{(\phi-1)^4}\mathrm{a}_i^* + 4\frac{\mu_i+1}{(\phi-1)^3}\mathrm{a}_i \\ &\qquad - 4\frac{\mu_i^2}{(\phi-1)^4}\mathrm{b}_i^* - 4\frac{\mu_i}{(\phi-1)^3}\mathrm{b}_i\biggr).\end{aligned}$$

To test linear hypotheses of interest we can use the three classic tests based on the likelihood function, including the Wald test, likelihood-ratio test, and score test; see, for instance, Boos and Stefanski (2013).

4 Influence measures

Xie and Wei (2008) obtained the MLE of the parameters for the Poisson inverse Gaussian regression model employing the EM algorithm, and analyzed the presence of influential observations using the generalized Cook's distance, the Q-distance, and under different perturbation schemes of the original model: perturbations of case weights and of the response and the explanatory variables, respectively. Barreto-Souza and Simas (2015) propose the use of the EM algorithm for the estimation of the parameters of the Poisson mixture regression models. They also derive some expressions of global influence analysis, as are the generalized Cook's distance and the Q-distance. They also make corrections to the work of Xie and Wei (2008), since the discrete nature of the response variable makes its perturbation infeasible.

In this section, we present the expressions that allow us to detect possible influential observations both for global and local influence for the Waring regression model using the Q-function.

4.1 Global influence

Let $l(\boldsymbol{\theta}|\mathbf{y}_c)$ and $l(\boldsymbol{\theta}|\mathbf{y}_{c[i]})$, be the log-likelihood function of the q-dimensional vector of parameters $\boldsymbol{\theta}$ for the complete-data and for the data with the deletion of the ith case, respectively, where a subscript $[i]$ means the original quantity with the ith case deleted.

Then, the function $Q(\bullet)$ for the data set without the ith observation is given by

$$Q_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E\{l(\boldsymbol{\theta}|\mathbf{y}_{c[i]})|\mathbf{y}_{0[i]}, \hat{\boldsymbol{\theta}}\},$$

whose maximum is denoted by $\hat{\boldsymbol{\theta}}_{[i]}$, $i = 1, \ldots, n$ and $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$.

Zhu et al. (2001) have proposed the use of the generalized Cook's distance from the Q-function. In this case distance between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{[i]}$ is given by

$$GD_i = \frac{(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]})^T\{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{[i]})}{q}, \qquad \forall i = 1, \ldots, n. \tag{4.1}$$

Using one-step linear approximations $\hat{\boldsymbol{\theta}}^{(1)}_{[i]}$ of $\hat{\boldsymbol{\theta}}_{[i]}$ given by Pregibon (1981), we get

$$\hat{\boldsymbol{\theta}}^{(1)}_{[i]} = \hat{\boldsymbol{\theta}} + \{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}), \tag{4.2}$$

where $\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\ddot{Q}_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ of the estimation algorithm is replaced by $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$.

By substituting (4.2) in (4.1), the one-step approximation of the GD_i is obtained in the context of EM

$$GD_i^{(1)} = \frac{(\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}))^T\{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1}(\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}))}{q} \approx (\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}))^T\{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1}(\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})). \tag{4.3}$$

Then, large values of $GD_i^{(1)}$ have a great impact on the MLEs. This proposal is later used by Pan, Fei and Foster (2014), who analyze the detection of influential observation in the mixed linear models.

Zhu et al. (2001) define the Q-distance as

$$QD_i = 2\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_{[i]}|\hat{\boldsymbol{\theta}})\};$$

by (4.2) one obtains the one-step approximation of QD_i

$$QD_i^{(1)} = 2\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_{[i]}^{(1)}|\hat{\boldsymbol{\theta}})\}. \tag{4.4}$$

Zhu et al. (2001) prove that, under certain regularity conditions, $GD_i^{(1)} \approx QD_i^{(1)}$. For the mixed Beta-Geometric model we have that

$$\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \left.\begin{pmatrix} \phi_2 \mathbf{X}_{[i]}^T \mathrm{R}_{[i]} \\ \dfrac{-2}{(\phi-1)^2} D \end{pmatrix}\right|_{\theta=\hat{\theta}} \tag{4.5}$$

is a q-dimensional vector, where $\mathrm{R}_{[i]}$ is a $(n-1)$-dimensional vector whose elements are given by $\mathrm{R}_{[i]} = (s_j + \psi(\phi_1 + \mu_j\phi_2) - \psi(\mu_j\phi_2))\mu_j$, with $\forall j \neq i$ and $D = \sum_{i \neq j}[e_j + \mu_j s_j + \psi(\phi_1 + \mu_j\phi_2)(1+\mu_j) - \psi(\phi_1) - \psi(\mu_j\phi_2)\mu_j]$. The expressions (3.4) and (4.5) allow to obtain the estimates at one-step from $\boldsymbol{\theta}_{[i]}$ as in (4.2), and thus the one-step generalized Cook's distance, given in (4.3).

The one-step Q-distance is calculated from (4.4) using one-step estimates of the parameters without the ith observation.

4.2 Local influence

Let $\boldsymbol{\omega}$ be a perturbation vector, $l_0(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_0)$ and $l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)$ the log-likelihood function of the observed data and of complete-data perturbed by $\boldsymbol{\omega}$, respectively, where $\boldsymbol{\theta}$ is a q-dimensional vector of parameters and the perturbed statistical model $\mathcal{M} = \{P(\mathbf{y}_c, \boldsymbol{\theta}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ be the MLE which maximize $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E(l_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_0, \hat{\boldsymbol{\theta}})$ and $Q(\boldsymbol{\theta}|\boldsymbol{\omega}, \hat{\boldsymbol{\theta}}) = E(l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)|\mathbf{y}_0, \hat{\boldsymbol{\theta}})$, respectively. Zhu et al. (2001) and Zhu and Lee (2001) propose the displacement of the Q-function to measure the difference between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$, given by

$$f_Q(\boldsymbol{\omega}) = 2[Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\hat{\boldsymbol{\theta}})],$$

where $Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = Q(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0, \hat{\boldsymbol{\theta}})$ and $Q(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\hat{\boldsymbol{\theta}}) = Q(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega}_0, \hat{\boldsymbol{\theta}})$. Note that $f_Q(\boldsymbol{\omega}) \geq 0$ $\forall \boldsymbol{\omega} \in \Omega$ and $f_Q(\boldsymbol{\omega}_0) = 0$.

Zhu and Lee (2001) define the influence graph of the Q-displacement function as $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}^T, f_Q(\boldsymbol{\omega}))^T$.

Then, $C_{Q,\mathbf{h}}(\boldsymbol{\theta})$ denotes the normal curvature of the surface $\alpha(\boldsymbol{\omega})$ in $\boldsymbol{\omega}_0$, in the direction of unitary vector $\mathbf{h}$ ($\|\mathbf{h}\| = 1$), which is given by

$$C_{Q,\mathbf{h}}(\boldsymbol{\theta}) = 2\left|\mathbf{h}^T \ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{h}\right|,$$

where $\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\omega}_0}^T\{-\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\}^{-1}\Delta_{\boldsymbol{\omega}_0}$, $\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\Delta_{\boldsymbol{\omega}} = \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\omega}^T}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}} \text{ and } \boldsymbol{\omega}=\boldsymbol{\omega}_0}$.

4.2.1 *Selection of the perturbation scheme.* Zhu et al. (2007) and Chen, Zhu and Lee (2009) discuss the need to be rigorous in the selection of perturbation schemes, based in the work of Cook (1986) on the use the normal curvature as a tool to evaluate the local influence of small perturbations of a statistical model. The proposal of Zhu et al. (2007) is based on the expected Fisher information matrix with respect to the perturbation vector, given by

$$\mathbf{G}(\boldsymbol{\omega}) = \big(G_{ir}(\boldsymbol{\omega})\big) = E_{\boldsymbol{\omega}}\big(\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\omega})\mathbf{U}_{\boldsymbol{\theta}}^T(\boldsymbol{\omega})\big) = -E_{\boldsymbol{\omega}}\left(\frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial\boldsymbol{\omega}\,\partial\boldsymbol{\omega}^T}\right),$$

where $E_{\boldsymbol{\omega}}$ denotes the expectation regarding the density of the perturbed model, $P(\mathbf{y}_c, \boldsymbol{\theta}, \boldsymbol{\omega})$, and $\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\omega})$ corresponds to the score function of the perturbed model.

For Zhu et al. (2007), an appropriate perturbation to a statistical model should satisfy the following conditions: $\mathbf{G}(\boldsymbol{\omega})$ is positive definite in a small neighborhood of $\boldsymbol{\omega}_0$ and the off-diagonal elements of $\mathbf{G}(\boldsymbol{\omega})$, evaluated in $\boldsymbol{\omega}_0$ should be as small as possible.

Then, a perturbation $\boldsymbol{\omega}$ is appropriate if, when evaluating the expected Fisher information matrix at $\boldsymbol{\omega}_0$, one has that $\mathbf{G}(\boldsymbol{\omega}_0) = \mathrm{c}\mathbf{I}_n$, where $\mathrm{c} = E_{\boldsymbol{\omega}}(\partial \log P(y_{ic}, \boldsymbol{\theta}, \omega_i)/\partial\omega_i)^2 > 0$ evaluated in $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, which warrants that the elements of $\boldsymbol{\omega}$ are asymptotically independent. Moreover, if the matrix $\mathbf{G}(\boldsymbol{\omega}_0) = \mathrm{Diag}(G_{11}(\boldsymbol{\omega}_0), \dots, G_{nn}(\boldsymbol{\omega}_0))$, it is always possible to choose another perturbation vector $\tilde{\boldsymbol{\omega}}$ defined as follows:

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}_0 + \mathrm{c}^{-1/2}\mathbf{G}(\boldsymbol{\omega}_0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}_0),$$

such that $\mathbf{G}(\tilde{\boldsymbol{\omega}})$ evaluated in $\boldsymbol{\omega}_0$ is equal to $\mathrm{c}\mathbf{I}_n$. For more details, see Rivas (2017).

Zhu et al. (2007) present various examples to find the appropriate perturbation under different schemes. Giménez and Galea (2013) analyze three schemes of perturbation from the comparative calibration model, such as: case weight, additive perturbation of the referent and an alternative instrument.

In this paper, the predictor additive perturbation is considered. Then, the log-likelihood function of the complete-data, under this scheme of perturbation, is given by

$$l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c) = \sum_{i=1}^{n} \{\log p_{i\omega} + y_i \log(1 - p_{i\omega}) + \log \Gamma(\phi_1 + \mu_i \phi_2) + (\phi_1 - 1) \log p_i + (\mu_i \phi_2 - 1) \log(1 - p_i) - \log \Gamma(\phi_1) - \log \Gamma(\mu_i \phi_2)\}, \tag{4.6}$$

where $p_{i\omega} = \frac{1}{1+\mu_{i\omega}}$, with $\mu_{i\omega} = \exp\{\mathbf{x}^T \boldsymbol{\beta} + S_i \omega_i \beta_j^*\}$, for $i = 1, \ldots, n$ and j^* representing the perturbed predictor. First, S_i is determined so that the proposed perturbation is appropriate. From (4.6), one has that

$$\frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i} = -(y_i + 1)\left(\frac{\mu_{i\omega}}{1 + \mu_{i\omega}}\right) S_i \beta_j^* + y_i S_i \beta_j^*,$$

$$\frac{\partial l_c^2(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i^2} = -(y_i + 1)\left(\frac{\mu_{i\omega}}{(1 + \mu_{i\omega})^2}\right) S_i^2 \beta_j^{*2},$$

$$\frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i \, \partial \omega_r} = 0, \qquad i \neq r.$$

Then, the elements of $\mathbf{G}$ matrix are given by

$$G_{ii}(\boldsymbol{\omega}) = -E\left(\frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i^2}\right) = \left(\frac{2 + \mu_{i\omega}}{1 + \mu_{i\omega}}\right) \frac{\mu_{i\omega}}{(1 + \mu_{i\omega})^2} S_i^2 \beta_j^{*2},$$
$$G_{ir}(\boldsymbol{\omega}) = E\left(\frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i} \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_r}\right) = 0. \tag{4.7}$$

Evaluating each element of (4.7) in $\boldsymbol{\omega}_0$, we obtain

$$\mathbf{G}(\boldsymbol{\omega}_0) = \beta_j^{*2} \operatorname{Diag}\left(\frac{2 + \mu_1}{(1 + \mu_1)^3} S_1^2, \ldots, \frac{2 + \mu_n}{(1 + \mu_n)^3} S_n^2\right) \neq \mathrm{c}\mathbf{I}_n.$$

Therefore, for the perturbation to be adequate, S_i is defined as

$$S_i = \left(\frac{2 + \mu_i}{(1 + \mu_i)^3}\right)^{-1/2}, \qquad \text{for } i = 1, \ldots, n.$$

Rewriting $\mu_{i\omega}$, one has that

$$\mu_{i\omega} = \exp\left\{\mathbf{x}_i^T \boldsymbol{\beta} + \left(\frac{2 + \mu_i}{(1 + \mu_i)^3}\right)^{-1/2} \omega_i \beta_j^*\right\}.$$

Therefore, the elements in the matrix Δ are given by, $\tau_{\phi_i} = 0$, $i = 1, \ldots, n$, while $\tau_{\beta_{ij}}$ for $j = j^*$ are given by

$$\tau_{\beta_{ij}} = -\frac{(y_i + 1)}{(1 + \mu_{i\omega})^2} \mu_{i\omega} \big[\beta_j^* S_i \big(x_{ij} + S_{i\beta_j} \omega_i \beta_j^* + S_i \omega_i\big) + \beta_j^* S_{i\beta_j} (1 + \mu_{i\omega} + y_i) + S_i (1 + \mu_{i\omega} + y_i)\big].$$

And for $j \neq j^*$,

$$\tau_{\beta_{ij}} = -\frac{(y_i+1)}{(1+\mu_{i\omega})^2}\mu_{i\omega}\big[\beta_j^* S_i\big(x_{ij} + S_{i\beta_j}\omega_i\beta_j^*\big) + \beta_j^* S_{i\beta_j}(1+\mu_{i\omega})\big] + \beta_j^* y_i S_{i\beta_j}.$$

Note that in this context, $S_{i\beta_j}$ is given by

$$S_{i\beta_j} = \frac{\partial S_i}{\partial \beta_j} = -\frac{1}{2}S_i^3\left(\frac{-2\mu_i^2 x_{ij} - 5\mu_i x_{ij}}{(1+\mu_i)^4}\right), \qquad \forall i = 1,\dots,n, \text{and } j = 0,\dots,k.$$

All the expressions must be evaluated in MLEs of $\boldsymbol{\theta}$ and in $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (0,\dots,0)^T$.

5 Application

In this section, we illustrate the methodology developed in the paper using simulated data and a real data set. For comparative purposes, we also fit the geometric regression model.

5.1 Simulated data

We present the results of a Monte Carlo simulation experiment in order to investigate the finite-sample performance of the MLEs. This experiment considers the Waring regression model, that is, $y_i \sim \text{Waring}(\mu_i, \phi)$, where $\phi = 2.5$ and $\mu_i = \exp(\beta_0 + \beta_1 x_i)$, whose covariate $x_i \sim \text{Uniforme}(0, 1)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^T = (2, 1)^T$. We choose five different sample size specifically $n = 50, 100, 500, 800,$ and 1000. For each sample size $R = 1000$, Monte Carlo replications were generated.

In order to analyze the results, we computed the average (*AVE*), the standard deviation (*sd*), the *Bias* and the root mean squared error ($\sqrt{MSE}$) of the estimates. The mean and the standard deviation of any estimate from $\boldsymbol{\theta} = (\beta_0, \beta_1, \phi)^T$, namely $\hat{\theta}$, are estimated as $AVE = R^{-1}\sum_{r=1}^{R}\hat{\theta}^r$ and $sd = \sqrt{(R-1)^{-1}\sum_{r=1}^{R}(\hat{\theta}^r - AVE)^2}$, respectively, with $\hat{\theta}^r$ being the estimate of $\boldsymbol{\theta}$ in the rth replication, for $r = 1,\dots,R$. The *Bias* and mean squared error is estimated as $Bias = AVE - \theta$ and $MSE = sd^2 + Bias^2$.

Table 1 presents simulation results, for the smallest sample size considered ($n = 50, 100$), the estimation algorithm failed to converge en 3%, and 1%, respectively. For large sample sizes the algorithm converged for all the samples.

For all the sample size considered, the β's MLE are very close to the real values and for ϕ it approaches a 2.5 (real value) as the sample size increases. Also and as expected, the *sd*, *Bias* and $\sqrt{MSE}$ decrease when the sample size increases.

Table 1 *Results from simulations, including average parameter estimates (AVE), standard deviation (sd), Bias and root mean squared error ($\sqrt{MSE}$)*

n	Parameter	*AVE*	*sd*	*Bias*	$\sqrt{MSE}$
50	β_0	1.9691	0.4090	−0.0309	0.4101
	β_1	0.9945	0.6815	−0.0054	0.6815
	ϕ	3.5156	3.9162	1.0156	4.0457
100	β_0	1.9726	0.2887	−0.0274	0.2900
	β_1	1.0163	0.4706	0.0163	0.4709
	ϕ	3.4078	3.4978	0.9078	3.6137
500	β_0	1.9931	0.1272	−0.0069	0.1274
	β_1	1.0078	0.2085	0.0078	0.2086
	ϕ	2.6864	1.2000	0.1864	1.2139
800	β_0	2.0021	0.1034	0.0021	0.1034
	β_1	0.9958	0.1670	−0.0042	0.1671
	ϕ	2.6112	0.7748	0.1112	0.7827
1000	β_0	1.9996	0.0920	−0.0004	0.0920
	β_1	0.9977	0.1510	−0.0023	0.1510
	ϕ	2.5868	0.6731	0.0868	0.6786

5.2 Visits to the doctor data

Next, as illustration, we consider the data set given in Hilbe (2011), called *rwm5yr* and available in the *COUNT* library of the packages of the same name in the *R* software. In that book, this data set was used for to compare the Negative Binomial regression model with the Generalized Waring regression models, but considering all the years in which they were measured. In our case, for simplicity, we only use the data for the year 1987.

The response variable (y_i) are the number of visits to the doctor of 1755 German women between 25 and 64 years old, during 1987, and the regressors variables are annual household income divided in 10 (in Marks) *INC*, the age (in years) *AGE*, and education (in years) *ED*. Figure 4 shows the behavior of the y_i per individual. There it can be observed the high number of women who did not go to the doctor during that year and how the frequency is reduced when the number of visits increases. Note that there are two patients that went to the doctor 82 and 90 times during 1987. These patients have the number 37 and 291, respectively.

Figure 5 shows the behavior of the response variable with respect to each of the covariates, from which is concluded that there are three observations away from the point cloud. Patients 37 and 291 correspond to observations with a high number of visit to the doctor and low family income, while the third observation, patient 285, corresponds to a woman with high family income and high number of visits. It can also be noted that patients 37 and 291 are over 50 years old and only has 9 years of education.

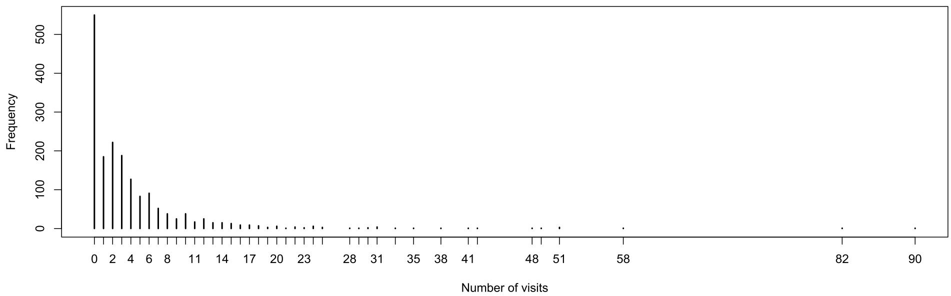

Figure 4 *Histogram number of visits to the doctor of* 1755 *German woman during* 1987.

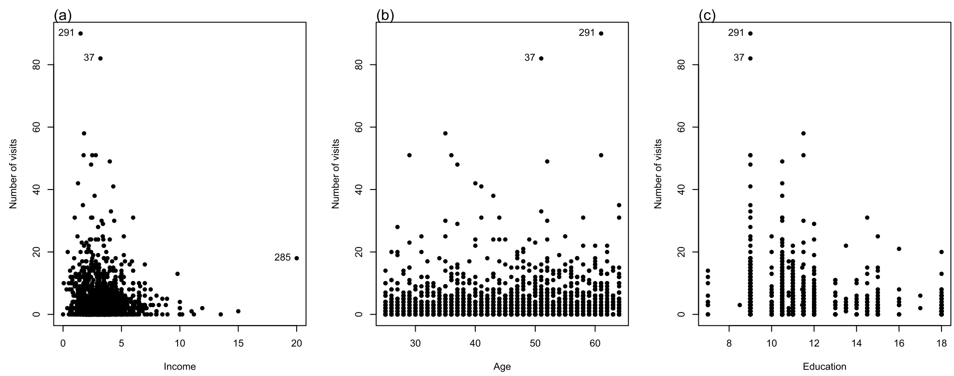

Figure 5 *Scatter plot of the number the visits to the doctor versus Income* (a), *Age* (b) *and Education* (c).

Table 2 presents the parameter estimates and their respective standard deviations (sd) for data on doctor visits obtained from the fitted Geometric and Waring regression models. It is observed that, although the estimates of the parameters are not very different in both models, the value of the log-likelihood function increases, while the AIC is reduced in the Waring regression model, giving indications that this model is more suitable than the Geometric regression model to describe the behavior of this data set. In addition, when performing the likelihood ratio test on the models to be compared, a value of $LR_c = 166.0824$ is obtained, with an associated p-value < 0.05, considering that the statistic corresponding to the test follows a distribution χ^2 with one degree of freedom, since the Geometric model is nested in the Waring model. This supports the foregoing conclusion. In the context of the problem, it becomes relevant that the it is probability, p, that each of these women go to the doctor change from individual to individual, as is natural to think, thus justifying the use of the Waring regression. Note that parameters β_1 and β_3, associated with *INC* and *ED*, respectively, are not significant. The expected Fisher information matrix was used to calculate the asymptotic standard deviations.

Table 2 *Summary of Geometric and Waring regression models adjustments for visits to doctor data*

Model	Estimation	
Geometric	$\hat{\beta}_0 = 1.4142\ (0.3017)$	$l_{\text{geo}} = -4345.632$
	$\hat{\beta}_1 = -0.0364\ (0.0246)$	$AIC = 8699.263$
	$\hat{\beta}_2 = 0.0099\ (0.0035)$	
	$\hat{\beta}_3 = -0.0352\ (0.0198)$	
Waring	$\hat{\beta}_0 = 1.1092\ (0.2662)$	$l_{\text{Waring}} = -4265.269$
	$\hat{\beta}_1 = -0.0258\ (0.0216)$	$AIC = 8540.538$
	$\hat{\beta}_2 = 0.0130\ (0.0031)$	
	$\hat{\beta}_3 = -0.0225\ (0.0174)$	
	$\hat{\phi} = 2.3701\ (0.3552)$	
$LR_c = 2(l_{\text{Waring}} - l_{\text{geo}}) = 166.0824$, p-value < 0.05		

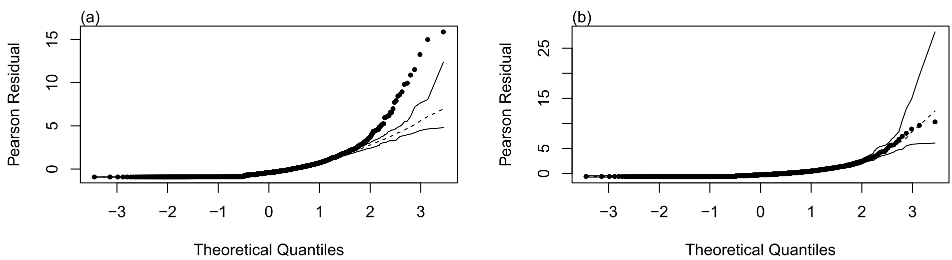

Figure 6 *Simulated envelope for the Pearson Residual in the Geometric model* (a) *and Waring model* (b).

Figure 6 presents the envelope simulations for the Pearson residuals, for both the Geometric regression model (a) and the Waring regression model (b), one can observed that the latter shows a better fit, as explained before, in the sense that the Waring regression model is more adequate to describe these data than the Geometric regression model.

Figure 7 shows the behavior of the Q-distance to one-step for $\hat{\boldsymbol{\theta}}$, $QD_i^{(1)}$. It is clearly observed that observation 285 is the one that has the greatest influence on the estimation of the parameters. Note that the graph of $GD_i^{(1)}$ is not included, since $QD_i^{(1)} \approx GD_i^{(1)}$.

Figure 8 shows that under the additive perturbation scheme of the variable *INC* the observations 37 and 291 are the most influential in the MLE of the parameters, in both graphs, $|\mathbf{h}_{max}|$ (a) and total local curvature C_i (b). The corresponding graphs for the variables *AGE* and *ED* were not included, since they present a similar behavior.

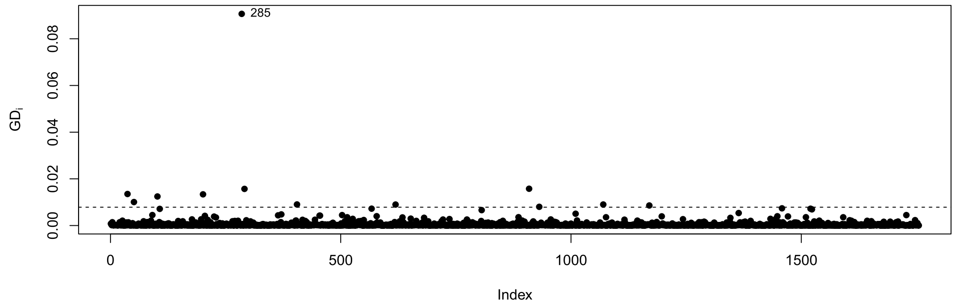

Figure 7 *Index plot of the generalized Cook's distance one-step for $\hat{\boldsymbol{\theta}}$.*

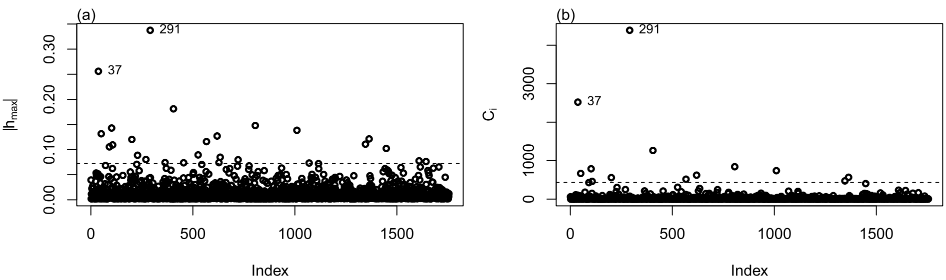

Figure 8 *Index plot of* $|\mathbf{h}_{\max}|$ (a) *and total local curvature* C_i (b) *under additive perturbation of covariable* INC.

Table 3 shows the rate of change (RC) of the MLEs, given by $RC_j = |\frac{\hat{\theta}_j - \hat{\theta}_{j(I)}}{\hat{\theta}_j}| * 100$, $\forall j$ for the Waring regression model, since the possible influential observations detected in the previous graphs and their combinations are deleted (37, 285 and 291). From the above table it can be concluded that β_2 in all cases proved to be significant and β_3 always remained non-significant for the model. However, β_1 changed its significance when the observation 285 was involved.

Similarly, it can be seen from Table 3 that the rates of changes of the MLEs with respect to β_1 are greater when a subset containing observation 285 is deleted from the data set. In addition, when this observation is deleted, it causes a change in the significance of parameter β_1, i.e., when this observation is included in the data set the hypothesis $H_0 : \beta_1 = 0$ is not rejected. However, when it is eliminated, and the hypothesis is contrasted again, it is rejected with a level of significance of 5%. In the remaining cases involving the elimination of observation 285, there is also a change in the significance of the parameter, but with a significance level of 10%.

It should be noted that the analysis was repeated only with the variables income and age, since the parameter associated with the variable education was never significant. The conclusions are analogous to what is presented in Table 3, therefore, the *ED* variable can be extracted from the model.

Table 3 *Rate of change of the MLEs by eliminating subset of observations in the Beta-Geometric regression model*

Observations deleted	RC β_0	β_1	β_2	β_3	ϕ	Significance
–	–	–	–	–	–	β_2 significant
37	1.21	1.55	0.77	4.44	4.34	β_2 significant
285	1.13	60.85	0.00	14.22	1.79	β_1, β_2 significant
291	0.15	6.20	3.08	1.78	4.13	β_2 significant
37,285	0.06	63.95	1.54	19.11	6.07	β_1^*, β_2 significant
37,291	1.34	4.26	3.85	6.22	8.32	β_2 significant
285,291	0.99	55.04	3.08	16.00	5.74	β_1^*, β_2 significant
37,285,291	0.17	57.75	3.85	20.89	9.88	β_1^*, β_2 significant

From Table 3, it can be seen that observation 285 exerts changes in the MLE of the parameters, which corresponds to a woman with high income, 42 years of age and with only 9 years of education.

It can be stated that, for this application, it is justified to choose the Waring model over the Geometric model, since it is natural that the probability associated with going to the doctor of any person is individual. Therefore, it is logical to think that this varies from subject to subject and, as such, the probability can have an associated distribution, in this case a Beta distributions.

Since observation 285 is detected as a potential influential observation, it is advisable to study its nature and/or apply a robust estimation method.

6 Conclusions

Evaluating the sensitivity (robustness) of the results obtained with the available data set is an important step in any statistical analysis, since outlying cases can distort estimators and test statistics, leading to, in some cases, wrong decisions. The aim of influence diagnostic methods is to identify outlying observations that may affect the values of statistics of interest under the proposed model. Thus, the goal of this paper was to propose influence measures to detect outlying observations that may distort some statistics of interest in the Waring regression model, a statistical model for count data, which is useful in many areas of knowledge.

We derive in closed-form expression the Fisher information matrix for the full parameter vector and a iterative process based on the Newton-Raphson algorithm is developed for maximum likelihood estimation.

Using the Q-function we calculated expressions for global influence as generalized Cook's distance and Q-distance; closed form expressions were obtained for local influence measures under the additive perturbation scheme of an explicatory variable and we defined an appropriate perturbation vector.

Empirical results show that the influence measures developed in the paper are useful for evaluating the effect of the observations in the estimation process and in hypothesis testing.

The codes used were written in R and are available and can be requested from the author via email.

Appendix

In this appendix, we obtain the moment generating function (MGF) of the Waring distribution (α, β) and the mean and variance. The MGF is given by

$$M_y(t) = E(e^{ty}) = \sum_{y=0}^{\infty} e^{ty} P(y|\alpha, \beta),$$

where $P(y|\alpha, \beta)$ is the probability function of the Waring distribution, which is obtained from the mixture given in (2.3), that is,

$$P(y|\alpha, \beta) = \int_0^1 p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp,$$

then, the MGF is given by

$$M_y(t) = \sum_{y=0}^{\infty} e^{ty} \int_0^1 p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp,$$

from where

$$\begin{aligned} M_y(t) &= \lim_{n\to\infty} \sum_{y=0}^{n} e^{ty} \int_0^1 p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp \\ &= \int_0^1 \lim_{n\to\infty} \sum_{y=0}^{n} e^{ty} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp. \end{aligned}$$

Given that

$$e^{ty} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \geq 0,$$

it is possible to define an increasing sequence, such that

$$A_n = \sum_{y=0}^{n} e^{ty} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1},$$

then we have that $0 \leq A_n \leq A_{n+1}$, therefore, by monotone convergence theorem

$$M_y(t) = \int_0^1 \sum_{y=0}^{\infty} e^{ty} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp.$$

Then,

$$M_y(t) = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\left(\sum_{y=0}^{\infty} e^{ty} p(1-p)^y\right) dp$$
$$= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \frac{p}{1-(1-p)e^t}\, dp,$$

given that

$$M_{y|p}(t) = \frac{p}{1-(1-p)e^t}, \qquad \text{as } y|p \sim \mathrm{Geo}(p).$$

Finally,

$$M_y(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{p^{\alpha}(1-p)^{\beta-1}}{1-(1-p)e^t}\, dp. \tag{A.1}$$

From (A.1), it is possible to attain mean and variance of the Waring distribution, given in (2.2). One can notice that

$$\int_0^1 \frac{\partial}{\partial t} \frac{p^{\alpha}(1-p)^{\beta-1}}{1-(1-p)e^t}\bigg|_{t=0} dp = \int_0^1 \frac{p^{\alpha}(1-p)^{\beta}e^t}{(1-(1-p)e^t)^2}\bigg|_{t=0} dp$$
$$= \int_0^1 p^{\alpha-2}(1-p)^{\beta}\, dp$$
$$= \frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} T_1$$
$$= \frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)},$$

where $T_1 = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha-1)\Gamma(\beta+1)} p^{\alpha-2}(1-p)^{\beta}\, dp$ is $\mathrm{Beta}(\alpha-1, \beta+1)$.

Therefore, $\frac{\partial}{\partial t} \frac{p^{\alpha}(1-p)^{\beta-1}}{1-(1-p)e^t}|_{t=0}$ is integrable. Then,

$$E(y) = \frac{\partial M_y(t)}{\partial t}\bigg|_{t=0} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{\partial}{\partial t} \frac{p^{\alpha}(1-p)^{\beta-1}}{1-(1-p)e^t}\bigg|_{t=0} dp = \frac{\beta}{\alpha-1}. \tag{A.2}$$

In a similar way, we can show that $\frac{\partial^2}{\partial t^2} \frac{p^{\alpha}(1-p)^{\beta-1}}{1-(1-p)e^t}|_{t=0}$ is integrable, and therefore the second moment is given by

$$E(y^2) = \frac{\partial^2 M_y(t)}{\partial t^2}\bigg|_{t=0}$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{\partial^2}{\partial t^2} \frac{p^{\alpha}(1-p)^{\beta-1}}{1-(1-p)e^t}\bigg|_{t=0} dp$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \bigg[\frac{p^{\alpha}(1-p)^{\beta}e^t(1-(1-p)e^t)^2}{(1-(1-p)e^t)^4}$$

$$
\begin{aligned}
&+ \frac{2p^{\alpha}(1-p)^{\beta+1}(1-(1-p)e^t)e^{2t}}{(1-(1-p)e^t)^4}\Bigg]\Bigg|_{t=0} dp \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{p^{\alpha+2}(1-p)^{\beta} + 2p^{\alpha+1}(1-p)^{\beta+1}}{p^4}\, dp \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \left(p^{\alpha-2}(1-p)^{\beta} + 2p^{\alpha-3}(1-p)^{\beta+1}\right) dp \qquad \text{(A.3)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \Bigg[\frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} \\
&\quad \cdot \underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha-1)\Gamma(\beta+1)} p^{\alpha-2}(1-p)^{\beta}\, dp}_{\sim \mathrm{Beta}(\alpha-1,\beta+1)} \\
&\quad + \frac{2\Gamma(\alpha-2)\Gamma(\beta+2)}{\Gamma(\alpha+\beta)} \\
&\quad \cdot \underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha-2)\Gamma(\beta+2)} p^{\alpha-3}(1-p)^{\beta+1}\, dp}_{\sim \mathrm{Beta}(\alpha-2,\beta+2)} \Bigg] \\
&= \frac{\beta}{\alpha-1} + \frac{2(\beta+1)\beta}{(\alpha-1)(\alpha-2)}.
\end{aligned}
$$

From (A.2) and (A.3), it is possible to attain the variance of the distribution,

$$
\begin{aligned}
\mathrm{Var}(y) = E\left(y^2\right) - E^2(y) &= \frac{\beta}{\alpha-1} + \frac{2(\beta+1)\beta}{(\alpha-1)(\alpha-2)} - \frac{\beta^2}{(\alpha-1)^2} \\
&= \frac{\alpha\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)},
\end{aligned} \qquad \text{(A.4)}
$$

where (A.2) and (A.4) coincide with (2.2).

Acknowledgments

The authors acknowledgments the partial support from Projects VRID 215.014. 025-1IN, Universidad de Concepción and FONDECYT 1150325, Chile.

References

Baird, D. D. and Wilcox, A. J. (1985). Cigarette smoking associated with delayed conception. *Journal of the American Medical Association* **253**, 2979–2983. DOI:10.1001/jama.1985.03350440057031.

Barreto-Souza, W. and Simas, A. B. (2015). General mixed Poisson regression models with varying dispersion. *Statistics and Computing* **26**, 1–18. MR3538636

Boos, D. and Stefanski, L. (2013). *Essential Statistical Inference. Theory and Methods*. New York: Springer. MR3024617

Chen, F., Zhu, H.-T. and Lee, S.-Y. (2009). Perturbation selection and local influence analysis for nonlinear structural equation model. *Psychometrika* **74**, 493–516. MR2551673

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B* **48**, 133–169. MR0867994

Crouchley, R. and Dassios, A. (1998). Interpreting the beta geometric in comparative fecundability studies. *Biometrics* **54**, 161–167. MR2534004

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38. MR0501537

Giménez, P. and Galea, M. (2013). Influence measures on corrected score estimators in functional heteroscedastic measurement error models. *Journal of Multivariate Analysis* **114**, 1–15. MR2993869

Henry, L. (1957). Fécondité et famille. *Population* **12**, 413–444. Available at http://www.jstor.org/stable/1524504. DOI:10.2307/1524504.

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge: Cambridge University Press. MR2797563

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1970). *Distributions in Statistics: Continuous Univariate Distributions*. New York: John Wiley & Sons. MR0418337

Miller, A. J. (1961). A queueing model for road traffic flow. *Journal of the Royal Statistical Society, Series B* **23**, 64–90. MR0128554

Pan, J., Fei, Y. and Foster, P. (2014). Case-deletion diagnostics for linear mixed models. *Technometrics* **56**, 269–281. MR3238065

Paul, S. R. (2005). Testing goodness of fit of the geometric distribution: An application to human fecundability data. *Journal of Modern Applied Statistical Methods* **4**, 425–433. Available at http://digitalcommons.wayne.edu/jmasm/vol4/iss2/8. DOI:10.22237/jmasm/1130803620.

Pielou, E. (1962). Runs of one species with respect to another in transects through plant populations. *Biometrics* **18**, 579–593. MR2527903

Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics* **9**, 705–724. MR0619277

Ridout, M. S. and Morgan, B. J. (1991). Modelling digit preference in fecundability studies. *Biometrics* **47**, 1423–1433. MR2532396

Rivas, L. (2017). Análisis de Influencia Para el Modelo de Regresión Waring. Tesis doctoral, Pontificia Universidad Católica de Chile, Chile. Available at http://www.mat.uc.cl/archivos/analisis-de-influencia-para-el-modelo-de-regresion-waring.pdf.

Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A. and Olmo-Jiménez, M. (2007). A new generalization of the Waring distribution. *Computational Statistics and Data Analysis* **51**, 6138–6150. MR2407704

Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A., Olmo-Jiménez, M. and Martínez-Rodríguez, A. (2009). A generalized Waring regression model for count data. *Computational Statistics and Data Analysis* **53**, 3717–3725. MR2751960

Singh, B., Pudir, P. and Maheshwari, S. (2014). Parameter estimation of beta-geometric model with application to human fecundability data. arXiv preprint. Available at arXiv:1405.6392, https://arxiv.org/ftp/arxiv/papers/1405/1405.6392.pdf.

Weinberg, C. R. and Gladen, B. C. (1986). The beta-geometric distribution applied to comparative fecundability studies. *Biometrics* **43**, 547–560. MR2531205

Xie, F.-C. and Wei, B.-C. (2008). Influence analysis for Poisson inverse Gaussian regression models based on the EM algorithm. *Metrika* **67**, 49–62. MR2353794

Zhu, H., Ibrahim, J. G., Lee, S. and Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics* **35**, 2565–2588. MR2382658

Zhu, H., Lee, S.-Y., Wei, B.-C. and Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika* **88**, 727–737. MR1859405

Zhu, H.-T. and Lee, S.-Y. (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society, Series B* **63**, 111–126. MR1811994

Departamento de Estadística
Universidad de Concepción
Concepción
Chile
E-mail: lrivas@udec.cl

Departamento de Estadística
Pontificia Universidad Católica de Chile
Santiago
Chile
E-mail: mgalea@mat.uc.cl