Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 4, 826–860
https://doi.org/10.1214/19-BJPS447

Bayesian approach for the zero-modified Poisson–Lindley regression model

Wesley Bertoli[a], Katiane S. Conceição[b], Marinho G. Andrade[b] and Francisco Louzada[b]

[a]*Federal University of Technology—Paraná*
[b]*University of São Paulo*

Abstract. The primary goal of this paper is to introduce the zero-modified Poisson–Lindley regression model as an alternative to model overdispersed count data exhibiting inflation or deflation of zeros in the presence of covariates. The zero-modification is incorporated by considering that a zero-truncated process produces positive observations and consequently, the proposed model can be fitted without any previous information about the zero-modification present in a given dataset. A fully Bayesian approach based on the *g-prior* method has been considered for inference concerns. An intensive Monte Carlo simulation study has been conducted to evaluate the performance of the developed methodology and the maximum likelihood estimators. The proposed model was considered for the analysis of a real dataset on the number of bids received by 126 U.S. firms between 1978–1985, and the impact of choosing different *prior* distributions for the regression coefficients has been studied. A sensitivity analysis to detect influential points has been performed based on the Kullback–Leibler divergence. A general comparison with some well-known regression models for discrete data has been presented.

1 Introduction

The standard Poisson ($\mathcal{P}$) distribution is the most adopted discrete model for the analysis of count data in several research fields, mainly due to its great simplicity and by having its computational implementation available for many statistical packages. However, it is well-known that such model is not a suitable choice for the analysis of counts in which the variance-to-mean ratio is not (at least) close to 1, that is, when the equidispersion property is violated. Apart from data transformation, the most popular way to circumvent such an issue is the use of finite mixture models (McLachlan and Peel, 2000) that can accommodate, for example, overdispersion (Karlis and Xekalaki, 2005). The Negative Binomial ($\mathcal{NB}$) distribution (that may arise as a $\mathcal{P}$ mixture model by using a Gamma distribution for the continuous part) is undoubtedly the most famous alternative to model extra-$\mathcal{P}$ variability. However, the literature concerning discrete models, which can handle different

Key words and phrases. Bayesian inference, influential points, overdispersion, Poisson–Lindley distribution, regression models, zero-modified data.
Received April 2018; accepted April 2019.

levels of overdispersion, is extensive and provides several other mixed distributions as the Poisson–Lindley (Sankaran, 1970), the Poisson–Lognormal (Bulmer, 1974), the Poisson–Inverse Gaussian (Shaban, 1981), the Negative Binomial–Lindley (Zamani and Ismail, 2010), the Poisson–Shanker (Shanker, 2016a), the Poisson–Sujatha (Shanker, 2016b), among others.

Unfortunately, there is a significant drawback regarding such mixture models, which is the fact that they do not fit well when data presents a modification in the frequency of zeros. Excess of zeros is the most common case in practice and, in the sense of address such a problem, several zero-inflated/hurdle approaches were proposed for the traditional $\mathcal{P}$ distribution (Cohen, 1960; Umbach, 1981; Mullahy, 1986; Lambert, 1992; Zorn, 1996; McDowell, 2003; Wagh and Kamalja, 2018). Various authors have considered these approaches for real applications, and here we point out a few. Bohara and Krieg (1996) have shown that the modeling of migratory frequency data can be improved by using zero-inflated Poisson ($\mathcal{ZIP}$) models. Gurmu and Trivedi (1996) have sought to deal with the excess of zeros on data from recreational trips. Ridout, Demétrio and Hinde (1998) have exploited the apple shoot propagation data, and they have addressed the modeling task by using several $\mathcal{ZIP}$ regression models. In the social sciences, Bahn and Massenburg (2008) have considered the hurdle version of the $\mathcal{P}$ model for the number of homicides in Chicago (State of Illinois, U.S.). Further applications of these models were considered for quantitative studies about HIV-risk reduction (Heilbron and Gibson, 1990; Hu, Pavlicova and Nunes, 2011), for the analysis of DNA sequencing data (Beuf et al., 2012) and the modeling of several datasets on chromosomal aberrations induced by radiation (Oliveira et al., 2016). Ngatchou-Wandji and Paris (2011) have provided a comprehensive discussion about zero-inflated models. A Bayesian approach for the zero-inflated Poisson distribution was considered by Rodrigues (2003) and by Ghosh, Mukhopadhyay and Lu (2006) in a regression framework. Usually, zero-deflated data are frequently observed in practice, but there are very few studies addressing this case solely (Dietz and Böhning, 2000; Angers and Biswas, 2003), even if this situation is often referred to in papers dealing with zero-inflation.

Zero-modified models provide a more general tool to handle count data with inflation or deflation at zero when no information about the nature of this phenomenon is available. Dietz and Böhning (2000) have introduced the zero-modified Poisson ($\mathcal{ZMP}$) regression model for zero-inflated/deflated samples and Conceição, Andrade and Louzada (2013) have considered a Bayesian approach for this model as an alternative to model Brazilian *leptospirosis* notification data. Bertoli et al. (2018) have proposed the zero-modified Poisson–Shanker regression model, whose usefulness was illustrated through its application to a real dataset on fetal deaths notification. Since zero-inflated/deflated models may also be useful to deal with data presenting overdispersed, this paper aims to introduce and present the usefulness of a regression model based on the zero-modified version

of the $\mathcal{PL}$ distribution ($\mathcal{ZMPL}$, for short). The Poisson–Lindley ($\mathcal{PL}$) distribution was proposed by Sankaran (1970) as an alternative to model overdispersed count datasets. The model was derived by mixing the $\mathcal{P}$ with the Lindley ($\mathcal{L}$) distribution (Lindley, 1958). A strictly positive random variable ψ is said to have $\mathcal{L}$ distribution if its probability density function (pdf) has the form

$$f(\psi;\theta) = \frac{\theta^2}{(\theta+1)}(\psi+1)e^{-\theta\psi}, \quad \psi \in \mathbb{R}_+,$$

where $\theta \in \mathbb{R}_+$ is the shape parameter. The author has shown that this model is defined by a 2-component mixture of an Exp(θ) and a Gamma$(2,\theta)$ distribution with mixing proportions given, respectively, by $\theta(\theta+1)^{-1}$ and $(\theta+1)^{-1}$.

Naturally, the proposed model is more flexible than the $\mathcal{ZIP}$ model, since it takes into account inflation or deflation of zeros, which is a feature often encountered when analyzing count data, besides modeling datasets with overdispersion that does not come only from the inflation or deflation of zeros. In a general regression framework, discrepant points (*outliers*) can be identified and, through a careful sensitivity analysis, one can quantify the influence of such observations. However, since the $\mathcal{PL}$ distribution accounts for several levels of overdispersion, its zero-modified version is naturally more robust than the $\mathcal{ZMP}$ model as it may accommodate discrepant points that do not significantly influence the parameter estimates.

For this paper, we consider a reparameterization of the original $\mathcal{ZMPL}$ probability mass function (pmf), which allow the likelihood function to be partitioned due to orthogonality between the vectors of fixed effects. Inference procedures are conducted under a fully Bayesian perspective, considering an adaptation of the *g-prior* method. The Random-walk Metropolis algorithm is applied for estimation via Markov Chain Monte Carlo (MCMC) sampling. A local influence measure based on the Kullback–Leibler divergence is considered for the task of detecting influential points. A Monte Carlo simulation study was performed, and the obtained results are presented to illustrate the performance of the developed methodology. For comparative purposes, the performance of the maximum likelihood estimators (MLEs) was also assessed. Moreover, an application using a real dataset on the number of bids received by 126 U.S. firms between 1978–1985 is presented as a way to assess the usefulness and the competitivity of the proposed model.

This paper is organized as follows. In Section 2, we briefly present the $\mathcal{PL}$ distribution and some of its main statistical properties. In Section 3, we introduce the $\mathcal{ZMPL}$ model, demonstrating its flexibility to deal with zero-inflated/deflated data. In Section 4, we present a regression framework based on the hurdle version of the $\mathcal{ZMPL}$ distribution. In Section 5, we describe all the Bayesian procedures and methodologies that are considered for inferential purposes in this paper. In Section 6, we discuss the results of a Monte Carlo simulation study and in Section 7, an application of the proposed model is exhibited. General comments and concluding remarks are addressed in Section 8.

2 The PL distribution

The $\mathcal{PL}$ distribution is a probabilistic model that may arise when it is believed that the rate parameter (ψ) of the $\mathcal{P}$ model behaves according to a $\mathcal{L}$ random variable into the subset of positive real numbers $\mathbb{R}_+$. In other words, a random variable Y is said to follow the $\mathcal{PL}$ law if the stochastic representation $Y|\psi \sim \mathcal{P}(\psi)$ and $\psi \sim \mathcal{L}(\theta)$ holds for all $\theta \in \mathbb{R}_+$. In this case, the $\mathcal{PL}$ distribution is defined by the equation

$$\mathrm{P}(Y = y; \theta) = \frac{\theta^2(\theta + y + 2)}{(\theta + 1)^{y+3}}, \quad y \in \mathcal{Y}_0, \tag{2.1}$$

for $\theta \in \mathbb{R}_+$ and $\mathcal{Y}_0 = \{0, 1, \ldots\}$ is the set of nonnegative integers. Using the gamma integral, the equation (2.1) can be easily obtained by integrating $\mathrm{P}(Y = y|\psi) f(\psi; \theta)$ respect to ψ over $\mathbb{R}_+$, in which $\mathrm{P}(Y = y|\psi)$ is the conditional pmf of a $\mathcal{P}$ random variable.

The unconditional distribution of the random variable Y can be denoted by $\mathcal{PL}(\theta)$. The pmf (2.1) does not involve complicated expressions, and therefore, the probabilities can be easily computed over $\mathcal{Y}_0$ as

$$\mathrm{P}(Y = 0; \theta) = \frac{\theta^2(\theta + 2)}{(\theta + 1)^3}.$$

From the results provided by Sankaran (1970), the mean and the variance of Y are given, respectively, by

$$\mu = \frac{\theta + 2}{\theta(\theta + 1)} \quad \text{and} \quad \sigma^2 = \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2}, \tag{2.2}$$

for $\theta \in \mathbb{R}_+$. It can be easily shown that $\mu \propto \theta^{-1}$. Also, the expression of σ^2 can be straightforwardly rearranged as

$$\sigma^2 = \mu\left[1 + \frac{\theta^2 + 4\theta + 2}{\theta(\theta + 1)(\theta + 2)}\right],$$

where the term inside the brackets correspond to the index of dispersion ($\mathrm{ID} = \sigma^2\mu^{-1}$). One can notice that the ratio involving θ is always positive. This implies that the $\mathcal{PL}$ distribution is always overdispersed, that is, $\{\theta \in \mathbb{R}_+ : \sigma^2 \leq \mu\} = \varnothing$. In addition, the index of dispersion is clearly greater than 1, also implying overdispersion. Conversely, we have that $\mathrm{ID} \to 1$ ($\sigma^2 \to \mu$) as $\theta \to \infty$, that is, the $\mathcal{PL}$ distribution has the property of equidispersion for large values of θ.

Another useful measure to characterize a discrete distribution is the zero-modification (ZM) index

$$\mathrm{ZM} = 1 + \mu^{-1} \log[\mathrm{P}(Y = 0)],$$

which is defined based on the $\mathcal{P}$ distribution. This measure can be easily interpreted since $\mathrm{ZM} > 0$ indicates zero-inflation, $\mathrm{ZM} < 0$ indicates zero-deflation and

ZM = 0 indicates no zero-modification. For the $\mathcal{PL}$ distribution, the ZM index is given by

$$\text{ZM} = 1 + \frac{\theta(\theta+1)[2\log(\theta) + \log(\theta+2) - 3\log(\theta+1)]}{\theta+2},$$

for $\theta \in \mathbb{R}_+$. When analysing the ZM index more deeply, we have obtained that ZM $\to -1$ as $\theta \to \infty$ and ZM $\to 0$ as $\theta \to 0$. This implies that, besides the usual case (ZM = 0), the $\mathcal{PL}$ distribution is suitable to deal with zero-deflation, but is not indicated to model zero-inflated datasets.

Now, let us reparameterize the pmf (2.1) in terms of the mean μ. It can be particularly useful since our interest is to derive a regression model based on the $\mathcal{PL}$ distribution, in which the influence of fixed and random effects can be evaluated directly over the mean of the response variable. Since $\theta \in \mathbb{R}_+$, we have that

$$\theta = -\frac{(\mu-1) - \sqrt{(\mu-1)^2 + 8\mu}}{2\mu}, \tag{2.3}$$

and, if we denote $\theta = h(\mu)$, the pmf (2.1) can be rewritten as

$$\text{P}(Y = y; \mu) = \frac{h^2(\mu)[h(\mu) + y + 2]}{[h(\mu)+1]^{y+3}}, \quad y \in \mathcal{Y}_0, \tag{2.4}$$

for $\mu \in \mathbb{R}_+$.

3 The ZMPL distribution

In addition to the interest in the case where the equidispersion assumption on the $\mathcal{P}$ distribution is violated, we may also be interested in the cases where a large/little amount of zeros is observed beyond that generated by the original process, which we already supposed to account for overdispersion. There are some typical situations where zero-modification may occur, and we list these cases in the following.

(a) Not all members of the population are affected by the process, which causes inflation of zeros to occur due to the response of unaffected subjects being zero;

(b) When zeros cannot be observed in the population (truncation at zero);

(c) The occurrence of unavoidable problems during the sampling process may lead to an increase/decrease in the probability of a zero observation being selected, hence the zero-inflation/deflation situation;

(d) A combination of (a) and (b) causes a part of the population to be zero-truncated distributed while the other part is not affected and provides the zero observations.

In this section, the $\mathcal{ZMPL}$ model is introduced as an alternative to model overdispersed count datasets when a large/little amount of zeros is observed beyond what would be expected by the $\mathcal{PL}$ distribution. Thus, let Y be a discrete

random variable defined on $\mathcal{Y}_0$. It can be stated that Y is distributed according to a $\mathcal{ZMPL}$ distribution if its pmf can be written as

$$\mathrm{P}^*(Y = y; \mu, p) = (1 - p)\delta_y + p\mathrm{P}(Y = y; \mu), \quad y \in \mathcal{Y}_0, \tag{3.1}$$

for $\mu \in \mathbb{R}_+$ and p is the zero-modification parameter. Also, δ_y is an indicator function, so that $\delta_y = 1$ if $y = 0$ and $\delta_y = 0$ otherwise. For the class of zero-modified models, the parameter p is subject to the condition (the so-called p-condition) given by

$$0 \leq p \leq \mathrm{P}^{-1}(Y > 0; \mu), \tag{3.2}$$

where $\mathrm{P}(Y > 0; \mu)$ is the probability of Y being positive under the $\mathcal{PL}$ distribution, given that its mean is μ. In this case, we have that

$$\mathrm{P}(Y > 0; \mu) = \frac{h^2(\mu) + 3h(\mu) + 1}{[h(\mu) + 1]^3}, \tag{3.3}$$

for $\mu \in \mathbb{R}_+$.

One can easily notice that (3.1) is not a mixture distribution typically chosen to model zero-inflated data, since parameter p can assume values greater than 1. However, for all values of p between 0 and the boundary $\mathrm{P}^{-1}(Y > 0; \mu)$, the equation (3.1) corresponds to a proper pmf since $\mathrm{P}^*(Y = y; \mu, p) \geq 0$ for all $y \in \mathcal{Y}_0$ and the probabilities sum to 1 over $\mathcal{Y}_0$.

The mean and the variance of $Y \sim \mathcal{ZMPL}(\mu, p)$ are given, respectively, by

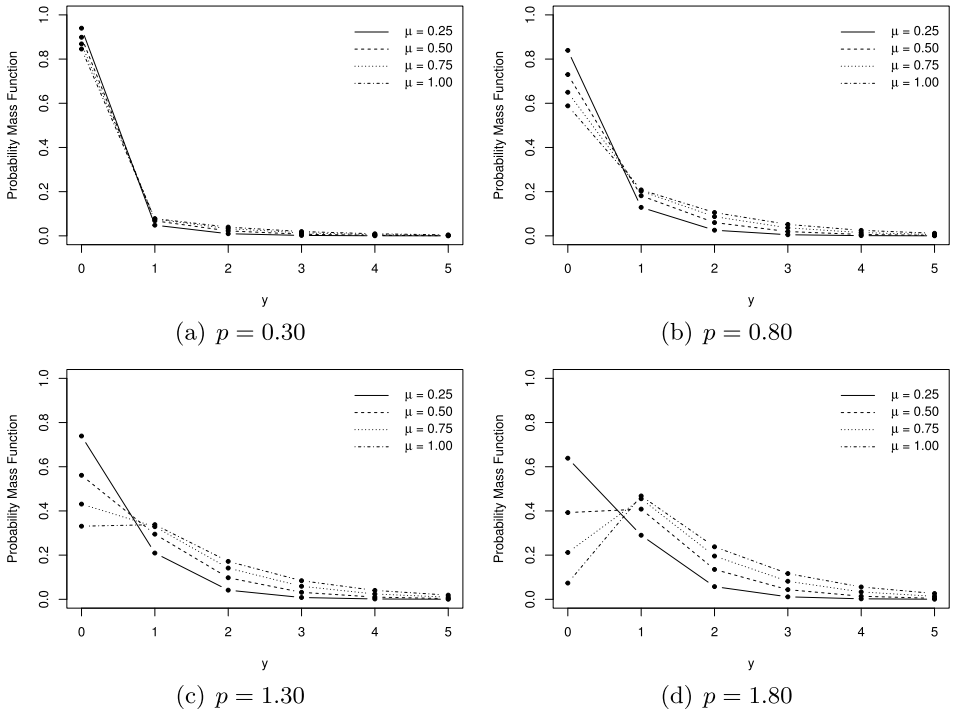
$$\mu^* = p\mu \quad \text{and} \quad (\sigma^2)^* = p[\sigma^2 + (1 - p)\mu^2],$$

where μ and σ^2 are given in equation (2.2). For the zero-modified case, the index of dispersion can be expressed as $\mathrm{ID}^* = \sigma^2\mu^{-1} + (1 - p)\mu$. The term $(1 - p)\mu$ represents the overdispersion caused by a modification on the zero frequency, regarding the $\mathcal{PL}$ distribution.

The $\mathcal{ZMPL}$ distribution may be considered an interesting alternative to the $\mathcal{ZMP}$ model, since the base distribution of the former can accommodate several levels of overdisperson, issue that the $\mathcal{P}$ distribution generally fails to deal with. Figure 1 depicts the behaviour of the $\mathcal{ZMPL}$ distribution for different values of p and for $\mu = 0.25$ (implying $p \in [0, 4.98]$), for $\mu = 0.50$ (implying $p \in [0, 2.96]$), for $\mu = 0.75$ (implying $p \in [0, 2.28]$) and for $\mu = 1.00$ (implying $p \in [0, 1.94]$). When looking at the pmf plots and the conditions regarding the zero-modification parameter, one can notice that the behaviour of pmf (3.1) is highly affected by the value of p, as can also be seen by considering the proportion of additional/missing zeros

$$\begin{aligned}\mathrm{P}^*(Y = 0; \mu, p) - \mathrm{P}(Y = 0; \mu) &= (1 - p) + p\mathrm{P}(0; \mu) - \mathrm{P}(Y = 0; \mu) \\ &= (1 - p)\mathrm{P}(Y > 0; \mu).\end{aligned} \tag{3.4}$$

The first interpretation one can take from equation (3.4) is that parameter p plays the primary role of controlling the frequency of zeros and therefore, it has



(a) $p = 0.30$ (b) $p = 0.80$

(c) $p = 1.30$ (d) $p = 1.80$

Figure 1 *Behavior of the $\mathcal{ZMPL}$ distribution for different values of μ and p.*

a natural interpretation regarding the proportions of either inflation or deflation at zero. The following statements describe the effect of parameter p on equation (3.1).

(i) If $p = 0$, then $\mathrm{P}^*(Y = 0; \mu, p) = 1$. This implies that equation (3.1) is the pmf of a degenerate distribution with all mass at zero;

(ii) If $p = 1$, then $\mathrm{P}^*(Y = 0; \mu, p) = \mathrm{P}(Y = 0; \mu)$. This implies that equation (3.1) is the pmf (3.3);

(iii) If $p \in (0, 1)$ then $(1 - p)\mathrm{P}(Y > 0; \mu) > 0$. This implies that equation (3.1) has a proportion of zeros greater than pmf (2.4), hence zero-inflation;

(iv) If $p \in [1, \mathrm{P}^{-1}(Y > 0; \mu)]$, then $(1 - p)\mathrm{P}(Y > 0; \mu) < 0$. This implies that equation (3.1) has a proportion of zeros smaller than pmf (2.4), hence zero-deflation;

(v) If $p = \mathrm{P}^{-1}(Y > 0; \mu)$, then $\mathrm{P}^*(Y = 0; \mu, p) = 0$. This implies that equation (3.1) is the zero-truncated Poisson–Lindley ($\mathcal{ZTPL}$) distribution (Ghitany, Al-Mutairi and Nadarajah, 2008), with pmf given by

$$\mathrm{P}_*(Y = y; \mu) = \frac{\mathrm{P}(Y = y; \mu)}{\mathrm{P}(Y > 0; \mu)}(1 - \delta_y), \tag{3.5}$$

where the numerator is given by (2.4) and the denominator is given by (3.3). One can notice that the reparameterization of equation (2.1) in terms of μ does not affect the general definition of the zero-truncated version of the $\mathcal{PL}$ model. See Shanker and Fesshaye (2016) for further details about the $\mathcal{ZTPL}$ distribution.

Given the value of p, one can easily identify the nature of the zero-valued observations under the phenomenon of interest. In this way, we have that the case (iii) may be appropriate in situations (a), (c) and (d) as described at the beginning of this section and the case (iv) may be appropriate in situations (c) and (d). Moreover, the index of dispersion can be investigated in terms of the modification at zero since $\mathrm{ID}^* = \mathrm{ID}$ in the standard case (ii), $\mathrm{ID}^* > \mathrm{ID}$ in the zero-inflated case (iii) and $\mathrm{ID}^* < \mathrm{ID}$ in the zero-deflated case (iv).

Now, let us rewritten pmf (3.1) as

$$\begin{aligned}\mathrm{P}^*(Y=y;\mu,p) &= \big[1-p+p\mathrm{P}(Y=0;\mu)\big]\delta_y + p\mathrm{P}(Y=y;\mu)(1-\delta_y)\\ &= \big[1-p\mathrm{P}(Y>0;\mu)\big]\delta_y\\ &\quad + p\mathrm{P}(Y=y;\mu)(1-\delta_y), \quad y\in\mathcal{Y}_0,\end{aligned}$$

for $\mu\in\mathbb{R}_+$. Taking $\omega = p\mathrm{P}(Y>0;\mu)$, we have that

$$\mathrm{P}^*(Y=y;\mu,\omega) = (1-\omega)\delta_y + \omega\mathrm{P}_*(Y=y;\mu), \quad y\in\mathcal{Y}_0, \tag{3.6}$$

where $\mathrm{P}_*(Y=y;\mu)$ is given by equation (3.5). By condition (3.2), we clearly have that $\omega\in[0,1]$.

Equation (3.6) corresponds to the hurdle version of the $\mathcal{ZMPL}$ distribution. Mullahy (1986) introduced the class of hurdle models and the relevant feature of such class is that the zero-valued observations are treated separately from the positive ones. In the main formulation, a binary probability model determines whether a zero or a nonzero outcome occurs and hence, an appropriated zero-truncated discrete distribution is chosen to describe the positive values (Saffar, Adnan and Greene, 2012). In this case, we have that the probability of $Y=0$ is $1-\omega$ and the probability of $Y>0$ is ω. The $\mathcal{ZMPL}$ distribution parameterized by ω can be denoted by $\mathcal{ZMPL}(\mu,\omega)$. Such representation can be visualized as a superposition of two random processes, that is, one that produces positive observations from the $\mathcal{ZTPL}$ distribution and another one that produces only zero-valued observations with probability $1-\omega$. Therefore, model (3.6) cannot be considered a 2-component mixture distribution.

By the hurdle representation of zero-modified models, only the positive observations are required to estimate parameter μ. Conceição et al. (2017) well discusses this fact for the class of zero-modified Power Series distributions, and here we extend this result by asserting that the zero-truncated version of the $\mathcal{ZMPL}$ distribution is equivalent to the $\mathcal{ZTPL}$ distribution and they have the same parameter μ. This can be easily checked using equation (3.6) because if we exclude the value zero from $\mathcal{Y}_0$ and divide the right-hand side of (3.6) by the probability of Y

being positive (ω), then we will get that $Y \sim \mathcal{ZTPL}(\mu)$. Besides, such a representation allows us to obtain a closed-form solution for the MLE of parameter ω, which is given by the proportion of nonzeros in the dataset. Also, it can be easily seen that, for any fixed $\mu \in \mathbb{R}_+$, the function $\omega \mathrm{P}^{-1}(Y > 0; \mu)$ maps from $[0, 1]$ to $[0, \mathrm{P}^{-1}(Y > 0; \mu)]$ bijectively and therefore, the invariance principle ensures that parameter p can be estimated using such function. Indeed, inference procedures about parameter p are required since we are often interested in identifying the kind of zero-modification (inflation or deflation) is present in a given dataset.

4 The ZMPL regression model

Let us suppose that we have a collection $(Y_1, \ldots, Y_n)$ of independent discrete random variables such that $Y_i | \boldsymbol{x}_i^\top, \boldsymbol{z}_i^\top, \boldsymbol{\beta} \sim \mathcal{ZMPL}(\mu_i, \omega_i)$, $i = 1, \ldots, n$. In this case, a regression model for count data based on the $\mathcal{ZMPL}$ distribution can be derived by rewriting equation (3.6) as

$$\mathrm{P}^*\big(Y_i = y_i; \boldsymbol{x}_i^\top, \boldsymbol{z}_i^\top, \boldsymbol{\beta}\big) = (1 - \omega_i)\delta_{y_i} + \omega_i \mathrm{P}_*(Y_i = y_i; \mu_i), \quad y_i \in \mathcal{Y}_0, \tag{4.1}$$

where $\boldsymbol{x}_i^\top = (1, x_{i1}, \ldots, x_{iq_1})$ and $\boldsymbol{z}_i^\top = (1, z_{i1}, \ldots, z_{iq_2})$ are related, respectively, to μ_i and ω_i and can include, for example, *dummy* variables, cross-level interactions and polynomials. The $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is the full vector of fixed effects, being $\boldsymbol{\beta}_1^\top = (\beta_{10}, \ldots, \beta_{1q_1})$ and $\boldsymbol{\beta}_2^\top = (\beta_{20}, \ldots, \beta_{2q_2})$. Here, q_1 (q_2) denotes the number of covariates considered on the systematic component of a linear predictor for parameter μ (ω).

The full design matrices of model (4.1) can be written as $\mathbf{X} = (\mathbf{1}_n, \mathbf{X}_{n \times q_1})$ and $\mathbf{Z} = (\mathbf{1}_n, \mathbf{Z}_{n \times q_2})$, where $\mathbf{1}_n$ is the intercept column and the submatrices $\mathbf{X}_{n \times q_1}$ and $\mathbf{Z}_{n \times q_2}$ are defined in such a way that the vector $(x_{i1}, \ldots, x_{iq_1})$ is the ith row of $\mathbf{X}_{n \times q_1}$ and the vector $(z_{i1}, \ldots, z_{iq_2})$ is the ith row of $\mathbf{Z}_{n \times q_2}$. To complete model definition, one have to specify two monotonic, invertible and twice differentiable link functions, say g_1 and g_2, in which $\mu_i = g_1^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_1)$ and $\omega_i = g_2^{-1}(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2)$ are well defined on $\mathbb{R}_+$ and $(0, 1)$, respectively. For this purpose, one can choose any suitable mappings g_1 and g_2 such that $g_1^{-1} : \mathbb{R} \to \mathbb{R}_+$ and $g_2^{-1} : \mathbb{R} \to (0, 1)$. The logarithm link function, $\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_1$, is the natural choice for g_1. For $g_2(\omega_i)$, the most usual choice is the *logit* link function,

$$\operatorname{logit}(\omega_i) = \log\left(\frac{\omega_i}{1 - \omega_i}\right) = \boldsymbol{z}_i^\top \boldsymbol{\beta}_2. \tag{4.2}$$

The *probit* link function,

$$\Phi^{-1}(\omega_i) = \boldsymbol{z}_i^\top \boldsymbol{\beta}_2, \tag{4.3}$$

is also appropriate for the requested purpose. Another possible choice for g_2 is

$$\log\big[-\log(1 - \omega_i)\big] = \boldsymbol{z}_i^\top \boldsymbol{\beta}_2, \tag{4.4}$$

which corresponds to the *complementary log–log* link function. One can notice that these link functions exclude the limit cases (i) and (v). Unlike the *logit* and *probit*, the *complementary log–log* transformation provides an asymmetric specification that can be useful when the probability of an outcome is very small/large. Further, a more sophisticated approach considering the power and the reversal power link functions was proposed by Bazán et al. (2017) and can be applied in our context to provide more flexible relationships between the linear predictor and the parameter ω.

The $\mathcal{ZMPL}$ regression model has $q_1 + q_2 + 2$ unknown parameters to be estimated, the components of the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The link functions (4.2), (4.3) and (4.4) for parameter ω exclude the limit cases (i) and (v). Besides, it is worthwhile to mention that identifiability problems could occur if the same covariate were used to model both mean (μ) and the zero-modification parameter (p) if we had considered a regression model derived from (3.6). Also, to ensure that the regression coefficients are identifiable, it is essential the covariates (within linear predictors) to be linearly independent. Unlike traditional approaches, the proposed model can be fitted to both zero-inflated/deflated datasets. In this case, given a set of covariates, the probability of a zero-valued count being observed for the ith individual is given by $1 - \omega_i$. The adopted parameterization makes model (4.1) separable into two parts due to orthogonality between parameters in the structural form of μ and ω. It also avoids non-identifiability problems as well as the use of the EM algorithm, typically used to fit mixture models. Regardless of the model framework, in this paper, we propose a fully Bayesian approach for estimation and inference procedures. The next section is dedicated to present the details of such an approach.

5 Inference

Let Y be a discrete random variable taking values on $\mathcal{Y}_0$. Suppose that a random experiment is carried out n times independently and, subject to $\boldsymbol{x}_i^\top$ and $\boldsymbol{z}_i^\top$ for each i, a vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ of observed values from Y is obtained. Considering model formulation (4.1), if $Y_i|\boldsymbol{x}_i^\top, \boldsymbol{z}_i^\top, \boldsymbol{\beta} \sim \mathcal{ZMPL}(\mu_i, \omega_i)$ holds for all i, then the likelihood function of the vector $\boldsymbol{\beta}$ can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}; \boldsymbol{y}, \mathbf{X}, \mathbf{Z}) &= \prod_{i=1}^{n} \omega_i \left(\frac{1-\omega_i}{\omega_i}\right)^{\delta_{y_i}} \left[\frac{\mathrm{P}(Y_i = y_i; \mu_i)}{\mathrm{P}(Y_i > 0; \mu_i)}\right]^{1-\delta_{y_i}} \\ &= \prod_{i=1}^{n} \left\{ g_2^{-1}(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2) \left[\frac{1 - g_2^{-1}(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2)}{g_2^{-1}(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2)}\right]^{\delta_{y_i}} \right. \\ &\quad \left. \times \left\{ \frac{\mathrm{P}[Y_i = y_i; g_1^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_1)]}{\mathrm{P}[Y_i > 0; g_1^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_1)]} \right\}^{1-\delta_{y_i}} \right\},\end{aligned}$$

and the correspondent log-likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\beta};\boldsymbol{y},\mathbf{X},\mathbf{Z}) &= \sum_{i=1}^{n}(1-\delta_{y_i})\log\left\{\frac{\mathrm{P}[Y_i=y_i;\, g_1^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_1)]}{\mathrm{P}[Y_i>0;\, g_1^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_1)]}\right\} \\ &\quad + \sum_{i=1}^{n}\left\{\log[g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)] - \delta_{y_i}\log\left[\frac{g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)}{1-g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)}\right]\right\} \\ &= \ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\mathbf{X}) + \ell_2(\boldsymbol{\beta}_2;\boldsymbol{y},\mathbf{Z}). \end{aligned} \tag{5.1}$$

For the $\mathcal{ZMPL}$ regression model, we will consider the log-linearity of the mean, that is, $g_1(\mu_i) = \log(\mu_i) = \boldsymbol{x}_i^\top\boldsymbol{\beta}_1$. The choice of g_2 is left open and the notation $\omega_i = g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)$ will be used when necessary. From equation (5.1), one can easily notice that the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal and that $\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\mathbf{X})$ depends only on the positive values on $\boldsymbol{y}$. In this way, the log-likelihood function of $\boldsymbol{\beta}_1$ takes the form

$$\begin{aligned}\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\mathbf{X}) &= \sum_{k\in\mathcal{K}_1}\log[h(\exp\{\boldsymbol{x}_k^\top\boldsymbol{\beta}_1\}) + y_k + 2] \\ &\quad + 2\sum_{k\in\mathcal{K}_1}\log[h(\exp\{\boldsymbol{x}_k^\top\boldsymbol{\beta}_1\})] \\ &\quad - \sum_{k\in\mathcal{K}_1}\log[h^2(\exp\{\boldsymbol{x}_k^\top\boldsymbol{\beta}_1\}) + 3h(\exp\{\boldsymbol{x}_k^\top\boldsymbol{\beta}_1\}) + 1] \\ &\quad - \sum_{k\in\mathcal{K}_1} y_k\log[h(\exp\{\boldsymbol{x}_k^\top\boldsymbol{\beta}_1\}) + 1], \end{aligned} \tag{5.2}$$

where $\mathcal{K}_1 = \{i : y_i > 0,\, y_i \in \boldsymbol{y}\}$ is the finite set of indexes regarding the positive observations of $\boldsymbol{y}$. Adopting this set-up is equivalent to assuming that each positive element of $\boldsymbol{y}$ comes from a $\mathcal{ZTPL}$ distribution. We are extending the fact that a loss of efficiency in an estimation procedure based on the zero-truncated Poisson model occurs whether additional information about the kind of zero-modification is available (Dietz and Böhning, 2000). Now, the log-likelihood function of $\boldsymbol{\beta}_2$ can be written as

$$\ell_2(\boldsymbol{\beta}_2;\boldsymbol{y},\mathbf{Z}) = \sum_{i=1}^{n}\log[g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)] - \sum_{k\in\mathcal{K}_2}\log\left[\frac{g_2^{-1}(\boldsymbol{z}_k^\top\boldsymbol{\beta}_2)}{1-g_2^{-1}(\boldsymbol{z}_k^\top\boldsymbol{\beta}_2)}\right], \tag{5.3}$$

where $\mathcal{K}_2 = \{i : y_i = 0,\, y_i \in \boldsymbol{y}\}$ is the finite set of indexes regarding the zero-valued observations of $\boldsymbol{y}$.

There are no closed-form solutions for the MLEs of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and therefore, nonlinear optimization algorithms or direct numerical search on the surface of log-likelihoods functions may be applied in order to obtain point estimates in the classical approach. By the maximum likelihood theory, a consistent estimator for the

covariance matrix of $\widehat{\boldsymbol{\beta}}_r$, $r = 1, 2$, is given by the inverse of the Fisher information $\mathcal{I}_r = \mathbb{E}_Y[\mathcal{J}_r]$, where

$$\mathcal{J}_1 = -\frac{\partial^2 \ell_1(\boldsymbol{\beta}_1; \boldsymbol{y}, \mathbf{X})}{\partial \boldsymbol{\beta}_1 \, \partial \boldsymbol{\beta}_1^\top} \quad \text{and} \quad \mathcal{J}_2 = -\frac{\partial^2 \ell_2(\boldsymbol{\beta}_2; \boldsymbol{y}, \mathbf{Z})}{\partial \boldsymbol{\beta}_2 \, \partial \boldsymbol{\beta}_2^\top},$$

are the observed information matrices. In our context, however, the computation of the expected value respect to Y is unfeasible and therefore, a numerical approximation for the covariance matrices can be obtained by evaluating $\mathcal{J}_r^{-1}$ at $\boldsymbol{\beta}_r = \widehat{\boldsymbol{\beta}}_r$ and using the observed vector $\boldsymbol{y}$.

5.1 *Prior* distributions

A Bayesian analysis starts by choosing suitable *prior* distributions for the set of unknown parameters. The *g-prior* (Zellner, 1986) is a common choice among Bayesian users of the multiple linear regression model, mainly due to the fact of providing a closed-form *posterior* distribution for the regression coefficients. The *g-prior* is classified as an objective *prior* method which uses the inverse of the Fisher information matrix up to a scalar variance factor ($\tau \in \mathbb{R}_+$) to obtain the *prior* covariance structure of the multivariate Normal distribution. Such specification is indeed quite attractive since the Fisher information plays a major role in the determination of large sample covariance in both Bayesian and classical inference.

The problem of eliciting conjugate *priors* for generalized linear models (GLMs) was addressed by Chen and Ibrahim (2003). Their approach can be considered as a generalization of the original *g-prior* method, but its application is restricted for the class of GLMs since the proposed *prior* does not have closed-form for non-normal exponential families. As an alternative, Gupta and Ibrahim (2009) have proposed the information matrix *prior* as a way to assess the *prior* correlation structure between the regression coefficients, not including the intercept since the design matrix is centered in order to ensure that β_0 is orthogonal to the other coefficients. This method uses the Fisher information similarly to a precision matrix whose elements are shrunken by the factor τ, which is considered fixed ($\tau \geq 1$). Based on such approach, we will consider, for the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, two multivariate Normal *prior* distributions of the form

$$\boldsymbol{\beta}_1 \sim \mathcal{MVN}[\boldsymbol{\beta}_1^0, \tau_1 \Sigma_1^0] \quad \text{and} \quad \boldsymbol{\beta}_2 \sim \mathcal{MVN}[\boldsymbol{\beta}_2^0, \tau_2 \Sigma_2^0], \tag{5.4}$$

where Σ_r^0 refers to $\mathcal{J}_r^{-1}$ evaluated numerically at $\boldsymbol{\beta}_r^0$, and τ_r is assumed known. The vectors $\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^0$ can be chosen arbitrarily if no specialized information is available. It is worthwhile to mention that we are not considering centered design matrices in our approach. Hence, we are able to include β_{10} in the proposed *g-prior* but, in this case, the intercept is *a priori* correlated with the other coefficients $(\beta_{11}, \ldots, \beta_{1q_1})$. The same applies for β_{20} and $(\beta_{21}, \ldots, \beta_{2q_2})$.

5.2 *Posterior* distributions and estimation

After *prior* specifications, the following step in a Bayesian analysis consists in the obtaining of computable *posterior* densities for the unknown model parameters. For the $\mathcal{ZMPL}$ regression model (4.1), the unnormalized joint *posterior* distribution of the vector $\boldsymbol{\beta}$ can be expressed as

$$\pi(\boldsymbol{\beta};\, \boldsymbol{y}, \mathbf{X}, \mathbf{Z}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1;\, \boldsymbol{y}, \mathbf{X}) + \ell_2(\boldsymbol{\beta}_2;\, \boldsymbol{y}, \mathbf{Z})\}\pi_1(\boldsymbol{\beta}_1)\pi_2(\boldsymbol{\beta}_2).$$

However, since $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal, we have that

$$\pi_1(\boldsymbol{\beta}_1;\, \boldsymbol{y}, \mathbf{X}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1;\, \boldsymbol{y}, \mathbf{X})\}\pi_1(\boldsymbol{\beta}_1) \tag{5.5}$$

and

$$\pi_2(\boldsymbol{\beta}_2;\, \boldsymbol{y}, \mathbf{Z}) \propto \exp\{\ell_2(\boldsymbol{\beta}_2;\, \boldsymbol{y}, \mathbf{Z})\}\pi_2(\boldsymbol{\beta}_2), \tag{5.6}$$

where ℓ_1 and ℓ_2 are given by (5.2) and (5.3), respectively. Obviously, the use of the proper *priors* given in (5.4) avoid the *posterior* distributions given in (5.5) to be improper.

From the Bayesian point of view, the parametric inference is based on the marginal *posterior* distributions, which can be obtained by integrating the joint *posterior* distributions given in (5.5). These densities have unknown forms mainly due to the complexity of the respective likelihood functions. In this case, Bayesian estimates for each element of $\boldsymbol{\beta}_r$ can be obtained by applying iterative procedures within a broad class of MCMC methods. Here we will consider the well-known Random-walk Metropolis (RwM) algorithm. Through this procedure, $q_r + 1$ chains can be generated for $\boldsymbol{\beta}_r$. The dimensionality issue will depend on how much covariates will be taken under consideration to describe the parameters of the $\mathcal{ZMPL}$ model. For the *posterior* distributions in (5.5), we will consider multivariate Normal specifications for the proposal (candidate-generating) densities in the algorithm. These distributions will be used as the main terms in the transition *kernels* when computing the acceptance probabilities. Hence, at any state $k > 0$, the MCMC simulation is performed by proposing a candidate $\boldsymbol{\psi}_r$ for $\boldsymbol{\beta}_r$ as

$$\boldsymbol{\psi}_r | \boldsymbol{\beta}_r^* \sim \mathcal{MVN}[\boldsymbol{\beta}_r^*,\, \nu_r \mathcal{S}_r^*],$$

where $\boldsymbol{\beta}_r^* = \nu_r \boldsymbol{\beta}_r^{(k-1)} + (1 - \nu_r)\boldsymbol{\beta}_r^0$ and $\nu_r = \tau_r(\tau_r + 1)^{-1}$. One can notice that transitions depends on the acceptance of pseudo-random vectors generated with mean given by the mixture between the actual state of the chains and the *priors* specification, which are shrunk by the factor $1 - \nu_r$. In addition, at any state $k > 0$, the covariance matrix of the candidate vector $\boldsymbol{\psi}_r$ can be approximated numerically by evaluating $\mathcal{S}_r^* = \mathcal{H}_r^{-1}$ at $\boldsymbol{\beta}_r = \boldsymbol{\beta}_r^*$, where

$$\mathcal{H}_1 = -\frac{\partial^2 \log[\pi_1(\boldsymbol{\beta}_1;\, \boldsymbol{y}, \mathbf{X})]}{\partial \boldsymbol{\beta}_1\, \partial \boldsymbol{\beta}_1^\top} \quad \text{and} \quad \mathcal{H}_2 = -\frac{\partial^2 \log[\pi_2(\boldsymbol{\beta}_2;\, \boldsymbol{y}, \mathbf{Z})]}{\partial \boldsymbol{\beta}_2\, \partial \boldsymbol{\beta}_2^\top}.$$

Algorithm 1 Random-walk Metropolis

procedure RwM($N, \boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}, \boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \tau_1, \tau_2$)
  Set $k \leftarrow 1$
  Set $\nu_1 \leftarrow \tau_1(1+\tau_1)^{-1}$ and $\nu_2 \leftarrow \tau_2(1+\tau_2)^{-1}$
  while $k \leq N$ **do**
    Set $\boldsymbol{\beta}_1^{(k)} \leftarrow \boldsymbol{\beta}_1^{(k-1)}$ and $\boldsymbol{\beta}_2^{(k)} \leftarrow \boldsymbol{\beta}_2^{(k-1)}$
    Set $\boldsymbol{\beta}_1^* \leftarrow \nu_1\boldsymbol{\beta}_1^{(k)} + (1-\nu_1)\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^* \leftarrow \nu_2\boldsymbol{\beta}_2^{(k)} + (1-\nu_2)\boldsymbol{\beta}_2^0$
    Generate $\boldsymbol{\psi}_1 \sim \mathcal{MVN}[\boldsymbol{\beta}_1^*, \nu\mathcal{S}_1^*]$ and $\boldsymbol{\psi}_2 \sim \mathcal{MVN}[\boldsymbol{\beta}_2^*, \nu\mathcal{S}_2^*]$
    Set $\alpha_1 \leftarrow \log[\pi_1(\boldsymbol{\psi}_1; \boldsymbol{y}, \mathbf{X})] - \log[\pi_1(\boldsymbol{\beta}_1^{(k)}; \boldsymbol{y}, \mathbf{X})]$
    Set $\alpha_2 \leftarrow \log[\pi_2(\boldsymbol{\psi}_2; \boldsymbol{y}, \mathbf{Z})] - \log[\pi_2(\boldsymbol{\beta}_2^{(k)}; \boldsymbol{y}, \mathbf{Z})]$
    Generate $u_1, u_2 \sim \mathcal{U}(0,1)$
    if $\log(u_1) < \alpha_1$ **then**
      Set $\boldsymbol{\beta}_1^{(k)} \leftarrow \boldsymbol{\psi}_1$
    end if
    if $\log(u_2) < \alpha_2$ **then**
      Set $\boldsymbol{\beta}_2^{(k)} \leftarrow \boldsymbol{\psi}_2$
    end if
    Set $k \leftarrow k+1$
  end while
  return $\{\boldsymbol{\beta}^{(k)}\}, k = 1, \ldots, N$
end procedure

Algorithm 1 can be used to generate chains for the regression coefficients using the RwM algorithm. To run the algorithm, initial conditions $\boldsymbol{\beta}_1^{(0)}$ and $\boldsymbol{\beta}_2^{(0)}$ are needed. For a specific asymptotic Gaussian environment, Roberts, Gelman and Gilks (1997) have shown that the optimal acceptance rate should be around 45% for 1-dimensional problems and asymptotically approaches to 23.40% in higher dimensional problems (> 4). Here we are considering acceptance rates varying between 23.40% and 40% as quite reasonable since the proposed model will generally have at least two parameters to be estimated. Indeed, the higher the values of τ_1 and τ_2 the smaller the acceptance rates in the RwM algorithm, which results in smaller variability of the estimates. This procedure generates a sample of size N for each parameter. The convergence of the chains can be monitored by the Gelman–Rubin (Gelman and Rubin, 1992) and Geweke (Geweke, 1992) diagnostics. After convergence, some of the generated samples can be discarded as burn-in. The procedure to decrease the correlation between generated values is the usual approach of getting thinned steps. The final sample is supposed to have size M. A summary of the *posterior* distributions can be obtained through the MCMC estimates.

In the next section, we discuss the results of a Monte Carlo simulation study that was conducted to assess the performance of the proposed Bayesian methodology.

Algorithm 2 Sequential-search

procedure SEQSEA($\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}$)
  Generate $x, u \sim \mathcal{U}(0, 1)$
  Set $\mu \leftarrow \exp\{\beta_{10} + \beta_{11}x\}$ and $\omega \leftarrow [1 + \exp\{-(\beta_{20} + \beta_{21}x)\}]^{-1}$
  Set $k \leftarrow (1 - \omega)$ and $y \leftarrow 0$
  while $u > k$ **do**
    Set $y \leftarrow y + 1$ and $k \leftarrow k + \omega \mathrm{P}_*(Y = y; \mu)$
  end while
  return y
end procedure

In Section 7, the usefulness and the competitivity of the proposed regression model are illustrated by using a real dataset. All computations were performed using the R environment (R Development Core Team, 2017). The executable *scripts* can be made available by the authors upon justified request.

6 Simulation study

The primary empirical properties of an estimation procedure can be evaluated through Monte Carlo simulations. We have performed an intensive simulation study aiming to validate the proposed Bayesian approach. For comparison purposes, the performance of the MLEs was also assessed. The simulation process was performed by generating 500 pseudo-random samples of sizes $n = 50, 100, 200$ and 500 of a variable Y following a $\mathcal{ZMPL}$ distribution under the regression framework (4.1). For the whole process, it was considered a $n \times 2$ design matrix $\mathbf{X} = (\mathbf{1}_n, \mathbf{X}_{n\times 1})$ in which $\mathbf{X}_{n\times 1}$ is a vector containing n generated values from an Uniform distribution on the unit interval. Here, we have fixed $\mathbf{Z} = \mathbf{X}$. Moreover, we have assigned different values for the vectors $\boldsymbol{\beta}_1^\top = (\beta_{10}, \beta_{11})$ and $\boldsymbol{\beta}_2^\top = (\beta_{20}, \beta_{21})$ in order to generate both zero-inflated and zero-deflated samples. We have considered two scenarios for each kind of zero-modification and these cases are treated separately in the following subsections. The logarithm link function was considered for g_1. For g_2, we have considered the link function (4.2).

Algorithm 2 can be used to generate a single pseudo-random realization from the $\mathcal{ZMPL}$ distribution in the regression framework with covariate $\mathcal{U}(0, 1)$ for μ and ω. The extension for the use of more covariates is straightforward. The process to generate a pseudo-random sample of size n consists of running the algorithm as often as necessary, say n^* times ($n^* \geq n$). The sequential-search is a black-box type of algorithm (see Hörmann, Leydold and Derflinger (2004)) and works with any computable probability vector. The main advantage of such a procedure is its simplicity. On the other hand, sequential-search algorithms may be slow as the *while-loop* may have to be repeated very often.

For the $\mathcal{ZMPL}$ distribution, the expected number of iterations (NI), that is, the expected number of comparisons in the *while* condition, is given by

$$\mathbb{E}(\mathrm{NI}) = \mu^* + 1 = \frac{\omega[h(\mu)+1]^2[h(\mu)+2]}{h(\mu)[h^2(\mu)+3h(\mu)+1]} + 1,$$

where $h(\mu)$ is given by equation (2.3).

To apply the proposed Bayesian approach in each case, we have considered the RwM algorithm for MCMC sampling. For each generated sample, a chain with $N = 50{,}000$ values was generated for each parameter, considering a burn-in of 20% of the size of the chain. Using trace plots and Geweke's diagnostic, the convergence of the chains was monitored, and their stationarity was revealed. To obtain pseudo-independent samples from the *posterior* distributions given in (5.5), a value generated out of 10 was considered, resulting in chains of size $M = 4000$ for each parameter. The *priors* were chosen to ensure that parameter p provides zero-inflation or zero-deflation depending on the case. We classify these *priors* as being "vague" since the only information we have taken into account is the kind of zero-modification present on the generated sample. We fixed $\tau_1 = \tau_2 = 5.0$, which have provided acceptance rates ranging between 30 and 35%. The *posterior* mean was considered as the Bayesian point estimator and its performance was assessed by evaluating its bias (B), its mean squared error (MSE) and its mean absolute percentage error (MAPE). Also, the coverage probability (CP) of the 95% Bayesian credible intervals (BCIs) was estimated for each parameter. Using the generated samples and letting $\gamma = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the measures of interest were obtained by

$$\mathrm{B}(\widehat{\gamma}) = \frac{1}{M}\sum_{j=1}^{M}(\widehat{\gamma}_j - \gamma), \qquad \mathrm{MSE}(\widehat{\gamma}) = \frac{1}{M}\sum_{j=1}^{M}(\widehat{\gamma}_j - \gamma)^2 \tag{6.1}$$

and

$$\mathrm{MAPE}(\widehat{\gamma}) = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{\widehat{\gamma}_j - \gamma}{\gamma}\right|. \tag{6.2}$$

The variance of $\widehat{\gamma}$ was estimated as the difference between the MSE and the square of the bias. Moreover, the CP of the BCIs was estimated as follow

$$\mathrm{CP}_{\%}(\gamma) = \frac{100}{M}\sum_{j=1}^{M}\delta_j(\gamma), \tag{6.3}$$

where $\delta_j(\gamma)$ assumes 1 if the jth BCI contains the real value γ and 0 otherwise. Also, we have estimated the below noncoverage probability (BNCP) and the above noncoverage probability (ANCP) of the BCIs. These measures are computed analogously to CP. The BNCP and ANCP may be useful to determine asymmetrical

features since they provide the probabilities of finding the real value of parameter γ on the tails of the generated *posterior* distribution.

For the classical approach, we have considered the Newton–Raphson optimization method to obtain numerical estimates, since no closed-form solution is available for the MLE of vector $\boldsymbol{\beta}$. The estimates were obtained using several initial values to guarantee convergence to the global maximum. Again, assuming $\gamma = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the bias, the MSE and the MAPE of $\widehat{\gamma}_{\mathrm{MLE}}$ were estimated as previously stated in equations (6.1) and (6.2). Besides, we were also interested in the computation of the coverage and noncoverage probabilities of the asymptotic confidence interval (ACI) of γ. The large sample approximation for the $100(1-\alpha)\%$ two-sided confidence interval of γ is given by

$$\widehat{\gamma}_{\mathrm{MLE}} \pm z_{(1-\alpha/2)}\widehat{\mathrm{SE}}(\widehat{\gamma}_{\mathrm{MLE}}),$$

where $z_{(1-\alpha/2)}$ is the upper $(\alpha/2)$th percentile of the standard Normal distribution. The standard error (SE) is estimated as the squared root of the variance of $\widehat{\gamma}_{\mathrm{MLE}}$. Finally, the CP of the ACIs is estimated using (6.3) and the noncoverage probabilities were computed analogously.

The computed measures are presented in Tables 1–3 (general summaries) and Tables 2–4 (coverage and noncoverage probabilities). We have noticed that, as expected, the parameter estimates became more accurate with increasing sample sizes since the estimated biases and mean squared errors have decreased considerably as n increased. In general, the MLEs were found more biased regarding the *posterior* mean. Although high MAPE values were obtained for some parameters, this does not compromise accuracy in estimation. For example, on Table 1 (Scenario 1), for $n = 200$, we have obtained a estimated MAPE value of approximately 55% for β_{11}, the Bayesian approach. Taking into account that the real value of this parameter is 0.50, we have that the estimates for β_{11} were ranging mostly between 0.23 and 0.78, which do not represent such a huge impact on the estimated mean (μ). Moreover, we have observed that the estimated CPs of the BCIs are converging to the nominal level of 95% and the *posterior* distributions became more symmetric with increasing sample sizes. At this point, one can see that slightly better results were obtained for the ACIs, but the use of asymptotic results plays against the classical approach since they are valid only for large n $(n \to \infty)$. Thus, considering the predefined structure, our simulation study has provided several indications about the suitability of the proposed Bayesian approach to estimate the parameters of the $\mathcal{ZMPL}$ regression model.

Regarding the comparison between the estimation procedures, it is well-known that the Bayesian approach has advantages when specialized *prior* information is available for the phenomenon under investigation. One of the main concerns of this paper is to provide the necessary tools for users interested in the application of the proposed model and who have this kind of information accessible. Nonetheless, our methodology can be applied using "vague" and noninformative *priors*. When

Table 1 *Summary of Bayesian and classical estimation procedures for zero-inflated samples*

		Bias		Variance		MSE		MAPE (%)	
n	Par.	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
					Scenario 1				
50	β_{10}	−0.0496	−0.0514	0.1201	0.1813	0.1225	0.1839	18.4958	19.0244
	β_{11}	0.0327	0.0148	0.5590	0.6374	0.5601	0.6376	119.3572	120.8523
	β_{20}	0.1516	0.0879	0.5012	0.7762	0.5242	0.7839	55.3431	55.7580
	β_{21}	−0.3017	−0.1640	1.7303	3.3837	1.8213	3.4106	68.5309	70.0988
100	β_{10}	−0.0254	−0.0343	0.0778	0.1357	0.0785	0.1369	14.5455	15.1311
	β_{11}	0.0193	0.0240	0.2704	0.3506	0.2708	0.3512	82.2737	83.4865
	β_{20}	0.0366	−0.0147	0.2353	0.5232	0.2366	0.5234	38.4661	40.5110
	β_{21}	−0.0511	0.0429	0.7113	2.3246	0.7139	2.3265	45.6203	48.1084
200	β_{10}	−0.0229	−0.0304	0.0299	0.0935	0.0305	0.0944	9.1091	9.9292
	β_{11}	0.0249	0.0294	0.1182	0.2163	0.1188	0.2172	54.5856	57.8233
	β_{20}	0.0342	−0.0059	0.0993	0.4401	0.1004	0.4401	24.9519	27.4913
	β_{21}	−0.0494	0.0357	0.2999	2.3711	0.3023	2.3724	29.5489	33.7693
500	β_{10}	−0.0065	−0.0194	0.0098	0.0780	0.0098	0.0784	5.2412	6.0263
	β_{11}	0.0013	0.0144	0.0337	0.1350	0.0337	0.1352	29.2489	32.1616
	β_{20}	0.0059	−0.0198	0.0335	0.3524	0.0353	0.3527	14.7641	17.4109
	β_{21}	−0.0033	0.0577	0.1043	2.0047	0.1043	2.0080	16.8862	20.9636
					Scenario 2				
50	β_{10}	−0.0559	−0.0763	0.2565	0.3049	0.2596	0.3108	15.0565	15.4028
	β_{11}	0.0345	0.0284	1.2664	1.3063	1.2676	1.3071	54.2357	54.2181
	β_{20}	0.0140	−0.0185	0.4487	0.9193	0.4489	0.9196	104.6621	108.5400
	β_{21}	−0.2539	−0.1335	2.0746	2.9093	2.1391	2.9271	114.3879	114.9491
100	β_{10}	−0.0274	−0.0525	0.1300	0.1613	0.1308	0.1641	11.0662	11.0420
	β_{11}	0.0047	0.0321	0.4326	0.4518	0.4326	0.4528	34.5772	34.1142
	β_{20}	0.0030	−0.0412	0.2639	0.8869	0.2639	0.8886	81.9453	87.8518
	β_{21}	−0.1044	0.0113	0.8780	2.0218	0.8889	2.0220	74.6983	79.8030
200	β_{10}	−0.0332	−0.0526	0.0433	0.3165	0.0445	0.3193	6.7164	7.6425
	β_{11}	0.0358	0.0535	0.1627	0.4814	0.1640	0.4843	21.3950	23.3487
	β_{20}	−0.0061	−0.0393	0.0955	0.7542	0.0955	0.7557	48.9634	56.3251
	β_{21}	−0.0299	0.0217	0.3616	1.5927	0.3625	1.5932	47.6516	53.7506
500	β_{10}	−0.0116	−0.0309	0.0149	0.2812	0.0151	0.2821	3.9067	4.8219
	β_{11}	0.0051	0.0225	0.0519	0.3654	0.0520	0.3659	12.1947	13.7427
	β_{20}	−0.0035	−0.0573	0.0372	1.4334	0.0372	1.4367	31.1360	40.9934
	β_{21}	−0.0051	0.0658	0.1269	2.6745	0.1269	2.6788	28.5447	34.8024

"vague" *priors* are considered, Bayesian procedures and MLEs present similar results as those obtained in this section. On the other hand, when noninformative *priors* are selected, it can be theoretically proved that both approaches present equivalent results. Therefore, a comparison between Bayesian and classical approaches will only show great Bayesian advantage if the *priors* are specified by

Table 2 *Coverage and noncoverage probabilities of the BCIs and ACIs using zero-inflated samples*

		BNCP		CP		ANCP		BNCP		CP		ANCP	
n	Par.	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
		Scenario 1						Scenario 2					
50	β_{10}	2.60	1.40	93.40	95.80	4.00	2.80	4.40	2.40	90.20	93.60	5.40	4.00
	β_{11}	4.80	2.40	91.40	95.40	3.80	2.20	5.20	4.00	90.60	92.40	4.20	3.60
	β_{20}	6.60	2.40	91.00	95.40	2.40	2.20	5.80	2.80	90.80	95.40	3.40	1.80
	β_{21}	3.40	2.80	90.80	94.40	5.80	2.80	4.00	3.20	90.20	94.80	5.80	2.00
100	β_{10}	3.00	1.80	91.80	94.40	5.20	3.80	4.80	2.20	91.00	94.20	4.20	3.60
	β_{11}	5.60	4.20	90.40	93.20	4.00	2.60	4.60	2.80	92.00	95.60	3.40	1.60
	β_{20}	3.20	2.00	94.20	96.00	2.60	2.00	4.00	2.20	92.40	95.60	3.60	2.20
	β_{21}	3.80	2.40	92.80	96.60	3.40	1.00	3.00	3.60	92.60	94.80	4.40	1.60
200	β_{10}	3.80	1.60	91.00	94.60	5.20	3.80	3.40	2.40	91.60	94.00	5.00	3.60
	β_{11}	6.00	4.40	89.20	91.20	4.80	4.40	4.80	3.00	92.60	94.80	2.60	2.20
	β_{20}	4.60	3.00	92.80	94.40	2.60	2.60	3.80	2.40	93.40	95.60	2.80	2.00
	β_{21}	3.00	1.80	93.00	95.60	4.00	2.60	3.20	2.80	92.20	94.40	4.60	2.80
500	β_{10}	4.60	3.60	92.20	93.80	3.20	2.60	2.40	1.80	92.60	95.20	5.00	3.00
	β_{11}	3.20	2.80	93.00	94.40	3.80	2.80	4.20	3.00	92.40	94.80	3.40	2.20
	β_{20}	3.40	2.00	92.00	94.80	4.60	3.20	3.80	3.00	93.20	94.80	3.00	2.20
	β_{21}	4.20	4.00	91.80	93.80	4.00	2.20	3.80	2.60	92.20	94.60	4.00	2.80

aggregating relevant information about the parameters to be estimated. Further, with small samples and good *prior* information, the Bayesian procedure is indeed more advantageous.

6.1 Zero-inflated artificial data

For the zero-inflated case, the samples were generated from the $\mathcal{ZMPL}$ distribution by considering that $p_i \in (0, 1)$ for all i. Here, the regression coefficients were chosen by taking into account that zero-inflated samples has, naturally, proportion of zeros greater than expected and therefore, the variable Y was generated with mean (μ) not even close to zero. Then, for the first scenario we have considered $\boldsymbol{\beta}_1^\top = (1.5, 0.5)$ and $\boldsymbol{\beta}_2^\top = (1.0, -1.5)$ to perform the simulation. In the following, the procedure was repeated by considering $\boldsymbol{\beta}_1^\top = (2.5, 1.5)$ and $\boldsymbol{\beta}_2^\top = (-0.5, -1.0)$. For these scenarios, the "vague" *prior* set-up is $(\boldsymbol{\beta}_1^0)^\top = (1.0, 0.0)$ and $(\boldsymbol{\beta}_2^0)^\top = (0.0, 1.0)$. Figure 2 depicts the Bayesian estimates obtained for parameter p using zero-inflated samples with $n = 500$ for each scenario. The real values are represented by the straight blue lines and the 95% BCIs are represented by the red dashed lines. The filled black dots represent the estimated values for each generated observation.

Table 3 *Summary of Bayesian and classical estimation procedures for zero-deflated samples*

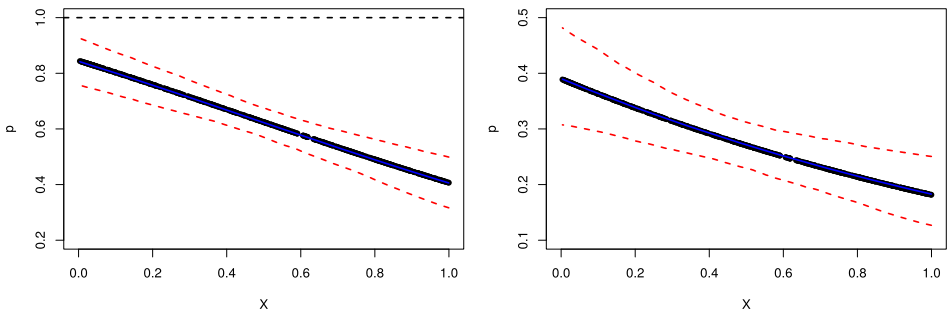
		Bias		Variance		MSE		MAPE (%)	
n	Par.	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
					Scenario 1				
50	β_{10}	−0.1971	−0.1667	0.6135	0.7634	0.6524	0.7912	60.4995	59.4275
	β_{11}	0.1060	0.0973	1.9266	2.1244	1.9378	2.1338	216.5709	214.5939
	β_{20}	0.0941	0.0674	0.5204	0.6096	0.5292	0.6142	111.2615	110.3212
	β_{21}	−0.0105	−0.0356	2.1566	3.1154	2.1567	3.1167	116.0536	115.4830
100	β_{10}	−0.0773	−0.0813	0.2853	0.3727	0.2913	0.3794	41.2579	41.6971
	β_{11}	0.0226	0.0476	0.7260	0.8623	0.7265	0.8646	132.8333	133.6173
	β_{20}	−0.0188	−0.0356	0.2673	0.4108	0.2676	0.4121	79.8809	81.0349
	β_{21}	0.1173	0.1383	0.9410	2.1766	0.9548	2.1957	75.4413	78.0565
200	β_{10}	−0.0564	−0.0695	0.1131	0.3314	0.1163	0.3362	27.0471	29.4275
	β_{11}	0.0432	0.0645	0.3103	0.6435	0.3122	0.6476	88.5523	94.8071
	β_{20}	0.0058	−0.0218	0.0965	0.3417	0.0966	0.3421	49.6113	54.0982
	β_{21}	0.0297	0.0929	0.3488	2.5299	0.3497	2.5386	47.2081	53.7624
500	β_{10}	−0.0284	−0.0433	0.0394	0.1844	0.0402	0.1863	15.8450	17.6385
	β_{11}	0.0226	0.0434	0.1000	0.3317	0.1006	0.3336	49.7987	54.7585
	β_{20}	0.0026	−0.0195	0.0352	0.2622	0.0352	0.2626	29.4331	33.6326
	β_{21}	0.0155	0.0761	0.1299	2.3141	0.1301	2.3199	27.4785	34.1051
					Scenario 2				
50	β_{10}	−0.5417	−0.3344	2.6364	2.1542	2.9298	2.2660	55.7422	49.1413
	β_{11}	0.1932	−0.0237	8.4866	8.4168	8.5239	8.4173	414.1736	397.8847
	β_{20}	0.1733	0.1010	0.5235	0.6591	0.5535	0.6693	57.8831	56.3176
	β_{21}	−0.2063	−0.1175	1.7803	3.3037	1.8228	3.3176	211.1536	212.7205
100	β_{10}	−0.2281	−0.1943	0.9002	0.9583	0.9522	0.9961	34.3945	33.8649
	β_{11}	0.0857	0.0934	2.1968	2.3735	2.2041	2.3822	222.6073	225.7564
	β_{20}	0.0325	−0.0097	0.2832	0.4358	0.2842	0.4359	41.4213	41.6895
	β_{21}	0.0074	0.0881	0.8323	2.4468	0.8324	2.4546	144.7771	150.7059
200	β_{10}	−0.1305	−0.1385	0.3088	0.9207	0.3258	0.9398	21.3299	22.6907
	β_{11}	0.0703	0.0971	0.9024	1.7183	0.9074	1.7277	149.3391	156.1515
	β_{20}	0.0219	−0.0181	0.1065	0.4449	0.1069	0.4452	25.6228	28.0340
	β_{21}	−0.0049	0.0909	0.3166	3.7195	0.3166	3.7277	90.1610	105.7611
500	β_{10}	−0.0438	−0.0675	0.0839	0.8042	0.0858	0.8087	11.5690	13.2514
	β_{11}	0.0073	0.0469	0.2345	1.1821	0.2346	1.1843	76.6633	84.2876
	β_{20}	0.0066	−0.0207	0.0373	0.3408	0.0373	0.3412	15.3647	17.9984
	β_{21}	0.0010	0.0765	0.1105	3.0407	0.1105	3.0465	52.7069	68.8181

6.2 Zero-deflated artificial data

For the zero-deflated case, the samples were generated from the $\mathcal{ZMPL}$ distribution by considering that $p_i \in [1, \mathrm{P}^{-1}(Y > 0; \mu_i)]$ for all i. Here, the regression coefficients were chosen by taking into account that zero-deflated samples has, naturally, proportion of zeros smaller than expected and therefore, the variable Y

Table 4 *Coverage and noncoverage probabilities of the BCIs and ACIs using zero-deflated samples*

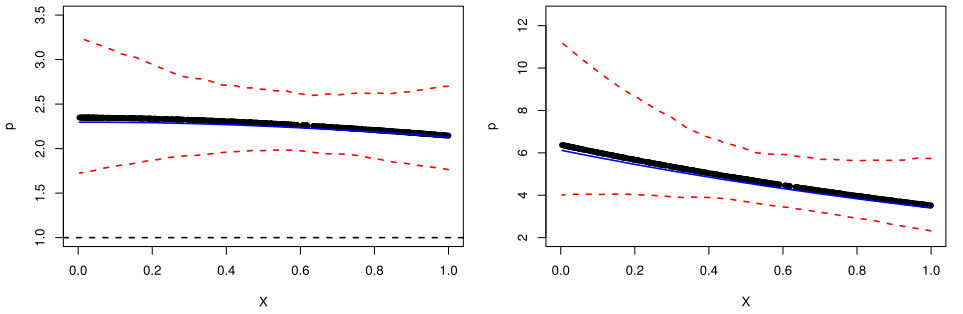
		BNCP		CP		ANCP		BNCP		CP		ANCP	
n	Par.	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
		Scenario 1						Scenario 2					
50	β_{10}	2.80	1.80	92.40	96.20	4.80	2.00	3.00	2.40	91.80	97.60	5.20	0.00
	β_{11}	4.00	2.80	92.80	95.20	3.20	2.00	4.00	1.60	91.80	97.80	4.20	0.60
	β_{20}	6.40	1.80	89.60	96.00	4.00	2.20	6.40	2.00	91.40	96.20	2.20	1.80
	β_{21}	3.40	1.40	91.40	95.80	5.20	2.80	3.00	1.60	90.40	95.80	6.60	2.60
100	β_{10}	3.40	2.40	92.80	95.60	3.80	2.00	2.80	2.60	91.00	96.60	6.20	0.80
	β_{11}	4.00	2.80	92.60	95.00	3.40	2.20	6.20	2.20	90.00	95.40	3.80	2.40
	β_{20}	3.20	1.60	92.60	95.00	4.20	3.40	4.60	1.40	91.20	94.80	4.20	3.80
	β_{21}	5.60	3.20	92.00	95.60	2.40	1.20	4.00	3.40	91.60	94.00	4.40	2.60
200	β_{10}	2.40	2.00	94.40	96.20	3.20	1.80	4.00	2.60	90.20	94.00	5.80	3.40
	β_{11}	3.40	2.20	93.80	95.60	2.80	2.20	5.80	3.40	90.40	95.00	3.80	1.60
	β_{20}	3.40	2.00	93.00	94.40	3.60	3.60	4.40	2.60	92.00	94.00	3.60	3.40
	β_{21}	2.40	1.80	93.60	95.20	4.00	3.00	3.20	3.00	92.40	93.80	4.40	3.20
500	β_{10}	3.00	2.80	93.80	94.40	3.20	2.80	2.60	2.40	93.40	95.40	4.00	2.20
	β_{11}	3.80	3.20	93.40	95.20	2.80	1.60	3.80	3.40	93.20	94.60	3.00	2.00
	β_{20}	2.60	1.40	93.20	95.80	4.20	2.80	2.60	1.80	93.20	95.20	4.20	3.00
	β_{21}	4.40	2.80	91.80	93.80	3.80	3.40	5.20	3.20	92.00	94.80	2.80	2.00



(a) Scenario 1 (b) Scenario 2

Figure 2 *Bayesian estimates for parameter p using zero-inflated samples.*

was generated with mean (μ) close to zero. Then, for the third scenario we have considered $\boldsymbol{\beta}_1^\top = (-1.0, 0.5)$ and $\boldsymbol{\beta}_2^\top = (0.5, 1.0)$ to perform the simulation. In the following, the procedure was repeated by considering $\boldsymbol{\beta}_1^\top = (-2.0, 0.5)$ and $\boldsymbol{\beta}_2^\top = (1.0, -0.5)$. For these scenarios, the "vague" *prior* set-up is $(\boldsymbol{\beta}_1^0)^\top = (0.0, -1.0)$ and $(\boldsymbol{\beta}_2^0)^\top = (2.0, 0.0)$. Figure 3 depicts the Bayesian estimates obtained for parameter p using zero-deflated samples with $n = 500$ for each scenario. Such representation has the same characteristics of Figure 2.



(a) Scenario 1 (b) Scenario 2

Figure 3 *Bayesian estimates for parameter p using zero-deflated samples.*

7 Takeover bids data analysis

In this section, the $\mathcal{ZMPL}$ regression model is considered for the analysis of a real dataset obtained from (Jaggia and Thosar, 1993). The sample consists of 126 U.S. firms that were targets of tender offers between 1978–1985 and which were taken over within a period of 52 weeks. In this study, the response variable is the number of additional bids after the initial bid received by the target firms. Also, a set of explanatory variables regarding target management actions and firm-specific characteristics was observed. The authors have analyzed the data by fitting a Poisson regression model, and they have verified that the white knight is one of the covariates associated with additional bids. The white knight is a management action of inviting a friendly third part to enter the bidding. The authors pointed out that, when inviting a friendly bidder, the management is indicating that may cede at least some control of the firm, and therefore, the entry (or potential entry) of at least one additional bidder is expected to spur the auction process.

Let us characterize the number of bids as the response variable (Y) and the white knight as a covariate (X). The variable X was coded as 0 (no additional bidder) and 1 (additional bidder). From the observed dataset, there exists evidence that Y is overdispersed since its mean is 1.74 and its variance is 2.05. Also, the range of Y is 10, and its coefficient of variation is approximately 118%. In this study, 75 out of 126 firms have at least one additional bidder invited for the process. The average number of bids was 1.18 when no additional bidder was invited and 2.12 otherwise. The absolute frequency of zeros is 9 (about 7% of the entire sample), which naively indicates a zero-deflation. Such characteristic is evidenced when fitting model (3.6) using the full dataset. This procedure was performed strictly for descriptive purposes. The model was fitted using a simpler version of Algorithm 1, since covariate X was not used. From its *posterior* summary, we have estimated equation (3.4) as $(1 - 1.87)[1 - f(0; 0.93)] \times 100 \approx -43\%$, suggesting that exist approximately 7 missing zeros, reinforcing our suspect that the sample is zero-deflated. Moreover, we had noticed that when the sample was observed, 6 firms

Table 5 Posterior *descriptive summary of the fitted $\mathcal{ZMPL}$ regression model*

					95% BCI		95% HPDI	
Par.	Mean	Median	SD	ESS	Lower	Upper	Lower	Upper
β_{10}	−0.8367	−0.8332	0.2563	2368.97	−1.3684	−0.3446	−1.3620	−0.3391
β_{11}	0.9354	0.9319	0.2972	2293.12	0.3665	1.5421	0.3806	1.5529
β_{20}	2.0608	2.0458	0.4055	2406.81	1.3147	2.9033	1.3266	2.9131
β_{21}	0.8104	0.8140	0.5965	2400.04	−0.3397	1.9806	−0.3064	2.0073

had received no further bids when no additional bidder was invited. As would be expected, the firm which has received the highest number of additional bids (10) have at least one additional bidder invited.

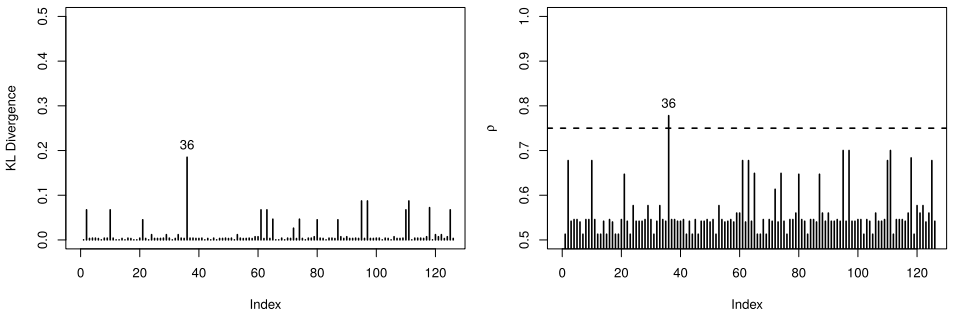
To fit the $\mathcal{ZMPL}$ regression model with X as a covariate, we have adopted a similar procedure to that one used in the previous section. As X is a dummy variable, we have fixed 0 (no additional bidder) as the *baseline* for estimation purposes. The logarithm link function was considered to relate μ_i with the linear predictor $\beta_{10} + \beta_{11} x_i$. To relate parameter ω_i with $\beta_{20} + \beta_{21} x_i$, we choose the link function given by (4.2). In this framework, parameter β_{11} represents the indirect effect of the invitation of at least one additional bidder on the mean (μ^*) and parameter β_{21} indicates the direct effect of such invitation on the probability of zeros ($1 - \omega$). We have considered the RwM algorithm, generating a chain of size $N = 50{,}000$ for each parameter whereby the first 10,000 values were discarded as burn-in. The stationarity of each chain was checked through the Geweke criterion for diagnostic of convergence. To obtain the pseudo-independent samples, we have considered one value out of every 10 generated one, resulting in chains of size $M = 4000$ for each parameter.

Table 5 presents the mean, the median and the standard deviation obtained from the *posterior* distribution of $\boldsymbol{\beta}$. To obtain the full descriptive summary, we have arbitrarily selected the *priors* $(\boldsymbol{\beta}_1^0)^\intercal = (0.0, -1.0)$, $(\boldsymbol{\beta}_2^0)^\intercal = (2.0, 0.0)$ and $\tau_1 = \tau_2 = 5.0$. In this framework, the acceptance rates in the RwM algorithm were at approximately 35%. In addition, we have calculated the number of effectively pseudo-independent draws (effective sample size—ESS) from the *posterior* distribution. The 95% BCIs were estimated empirically from the generated samples and the 95% highest *posterior* density intervals (HPDIs) were also computed.

A sensitivity analysis was performed to evaluate the behaviour of the Bayesian estimators under distinct *prior* specifications. We have established three scenarios where the model parameters were estimated considering different choices for $\boldsymbol{\beta}_r^0$ and τ_r, $r = 1, 2$. The results are displayed on Table 6. The Bayesian estimates for the parameters μ^*, p and n_0 (expected number of zeros) are also presented. Here, parameters μ^* and p^* were estimated as functions of the predictive $\mathcal{ZMPL}$ model, that is, $\widehat{\mu}^* = n^{-1} \sum_{i=1}^n \widehat{\mu}_i^*$ and $\widehat{p} = n^{-1} \sum_{i=1}^n \widehat{\omega}_i \mathrm{P}^{-1}(Y > 0; \widehat{\mu}_i)$. Since $\boldsymbol{\beta}_1$

Table 6 *Sensitivity analysis to evaluate the effect of different* prior *specifications*

Priors	Par.	Mean	$\widehat{\mu}^*$	$\widehat{p}$	$\widehat{n}_0$
	$\tau_1 = \tau_2 = 2.0$				
$(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$	β_{10}	−0.64	1.63	2.15	11
	β_{11}	0.56			
$(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$	β_{20}	2.04			
	β_{21}	0.57			
$(\boldsymbol{\beta}_1^p)^\intercal = (4.0, -4.5)$	β_{10}	−0.07	1.81	1.83	8
	β_{11}	0.09			
$(\boldsymbol{\beta}_2^p)^\intercal = (4.0, -1.0)$	β_{20}	2.25			
	β_{21}	0.93			
$(\boldsymbol{\beta}_1^p)^\intercal = (8.5, -9.5)$	β_{10}	−1.05	1.75	2.54	9
	β_{11}	0.98			
$(\boldsymbol{\beta}_2^p)^\intercal = (10.0, -7.0)$	β_{20}	2.09			
	β_{21}	1.10			
	$\tau_1 = \tau_2 = 5.0$				
$(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$	β_{10}	−0.84	1.84	2.22	10
	β_{11}	0.94			
$(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$	β_{20}	2.06			
	β_{21}	0.81			
$(\boldsymbol{\beta}_1^p)^\intercal = (4.0, -4.5)$	β_{10}	−0.58	1.85	2.03	8
	β_{11}	0.72			
$(\boldsymbol{\beta}_2^p)^\intercal = (4.0, -1.0)$	β_{20}	2.15			
	β_{21}	1.10			
$(\boldsymbol{\beta}_1^p)^\intercal = (8.5, -9.5)$	β_{10}	−1.08	1.95	2.47	8
	β_{11}	1.18			
$(\boldsymbol{\beta}_2^p)^\intercal = (10.0, -7.0)$	β_{20}	2.09			
	β_{21}	1.17			
	$\tau_1 = \tau_2 = 8.0$				
$(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$	β_{10}	−0.92	1.93	2.27	9
	β_{11}	1.07			
$(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$	β_{20}	2.06			
	β_{21}	0.93			
$(\boldsymbol{\beta}_1^p)^\intercal = (4.0, -4.5)$	β_{10}	−0.75	1.92	2.13	8
	β_{11}	0.93			
$(\boldsymbol{\beta}_2^p)^\intercal = (4.0, -1.0)$	β_{20}	2.13			
	β_{21}	1.16			
$(\boldsymbol{\beta}_1^p)^\intercal = (8.5, -9.5)$	β_{10}	−1.08	2.03	2.45	8
	β_{11}	1.23			
$(\boldsymbol{\beta}_2^p)^\intercal = (10.0, -7.0)$	β_{20}	2.09			
	β_{21}	1.19			



(a) Kullback-Leibler divergence (b) Calibration (ρ)

Figure 4 *Sensitivity analysis to detect influential points.*

and $\boldsymbol{\beta}_2$ are orthogonal, the results can be directly combined by taking $\tau_1 \neq \tau_2$. Obviously, we cannot decide on the *prior* distributions based on the *posterior* results, but we can investigate whether the *prior* specifications are influential. In this way, we have observed that the estimator of β_{20} is less sensitive regarding the *prior* choice, with variance smaller than 0.005 between estimates. For $\tau_j = 2.0$, the estimators of β_{10} and β_{11} are more sensitive, implying higher variability between estimates of parameter p. In addition, when large values are selected for τ_r, one can notice that the lower the *prior* choices for β_{11} and β_{21} the larger the estimates of μ^*. Amidst these features, it is worthwhile to mention that, as τ_r increases, the *prior* specification tends to have lower impact on the final fit $(\widehat{\mu}^*, \widehat{p}, \widehat{n}_0)$ and, in general, such impact can be considered negligible, even for quite distinct *priors*.

The analysis to verify the existence of influential points is presented in Figure 4. Figure 4(a) depicts the Kullback–Leibler (KL) divergence (see Appendix A), that used to evaluate the effect of each observation on the parameter estimates. Conservatively, we consider an observation whose distance has a calibration exceeding 0.75 as an influential point. Based on Figure 4(b), we have observed the existence of one influential point (36), corresponding to the firm with 10 additional bids. As a way to access the influence of this observation, the estimation process was repeated considering the removal of such firm of the sample. The *posterior* summary for this case and the variation percentage regarding the *posterior* summary obtained from the full dataset is presented on Table 7. When analyzing the parameter estimates, it can be observed that the removal of observation 36 impacts reasonably the model fit and the main variation is observed when estimating parameter β_{11}.

For comparison purposes, identical Bayesian procedures were adopted to fit the $\mathcal{P}$, the $\mathcal{PL}$ and the $\mathcal{ZMP}$ regression models. To estimate the fixed dispersion parameter (ϕ) of the $\mathcal{NB}$ model, we have considered a noninformative Inverse–Gamma *prior* distribution with hyperparameters $a = b = 1.0$. For each fitted model, we have estimated the measures presented in Appendix B. The model

Table 7 Posterior *descriptive summary—sample without influential observation*

Par.	Mean	Median	SD	ESS	95% BCI		95% HPDI	
					Lower	Upper	Lower	Upper
β_{10}	−0.8395	−0.8367	0.2549	2597.71	−1.3524	−0.3481	−1.3524	−0.3479
	(−0.33%)	(−0.42%)	(−0.55%)	(9.66%)				
β_{11}	0.8516	0.8493	0.2979	2394.79	0.2723	1.4447	0.2615	1.4326
	(−8.96%)	(−8.86%)	(0.24%)	(4.43%)				
β_{20}	2.0622	2.0477	0.4029	2420.74	1.3267	2.8911	1.2674	2.8193
	(0.07%)	(0.09%)	(−0.64%)	(0.58%)				
β_{21}	0.8002	0.7951	0.5903	2848.08	−0.3396	1.9810	−0.2880	2.0199
	(−1.26%)	(−2.32%)	(−1.04%)	(18.67%)				

Table 8 *Comparison criteria for the fitted models*

Model	Full dataset				Without observation 36			
	DIC	EAIC	EBIC	LMPL	DIC	EAIC	EBIC	LMPL
$\mathcal{P}$	1011.60	409.04	414.71	−195.53	959.15	388.06	393.72	−185.45
$\mathcal{NB}$	988.68	402.04	405.71	−195.87	943.12	383.81	387.47	−185.97
$\mathcal{PL}$	1104.06	446.03	451.70	−291.00	1075.44	434.58	440.24	−213.58
$\mathcal{ZMP}$	941.45	385.37	387.04	−183.65	884.90	362.77	364.42	−172.56
$\mathcal{ZMPL}$	920.02	376.81	378.49	−178.43	882.32	361.71	363.37	−171.40

comparison procedure is summarized in Table 8. One can notice that the zero-modified models have performed considerably better with $\mathcal{ZMPL}$ outperforming all. These results are highlighting that the proposed model is highly competitive with well-established models in the literature.

Figure 5 presents the marginal *posterior* densities of parameters of the $\mathcal{ZMPL}$ regression model. These densities provided the summary displayed on Table 7. The assumption of normality for the generated chains is quite reasonable even in the presence of slightly heavy tails on some of the estimated densities. Besides, there exists evidence of symmetry since *posterior* mean and median are very close to each other. For each parameter, the effective sample size was estimated greater than $M/2$, which can be considered an indication of good mixing of the generated chains, without any computational waste.

From the results displayed in Table 7, one can make some inferences and take some conclusions. Firstly, we have observed that the BCI/HPDI of the parameter β_{11} does not contain the value zero, which constitute the white knight as a relevant covariate to describe the average number of bids. On the other hand, the probability of not receive at least one additional bid is $1-[1+\exp\{-2.0622\}]^{-1} \approx 0.113$ if no additional bidder is invited and $1-[1+\exp\{-(2.0622+0.8002)\}]^{-1} \approx 0.054$ otherwise. However, as parameter ω is not affected by individual white knights,

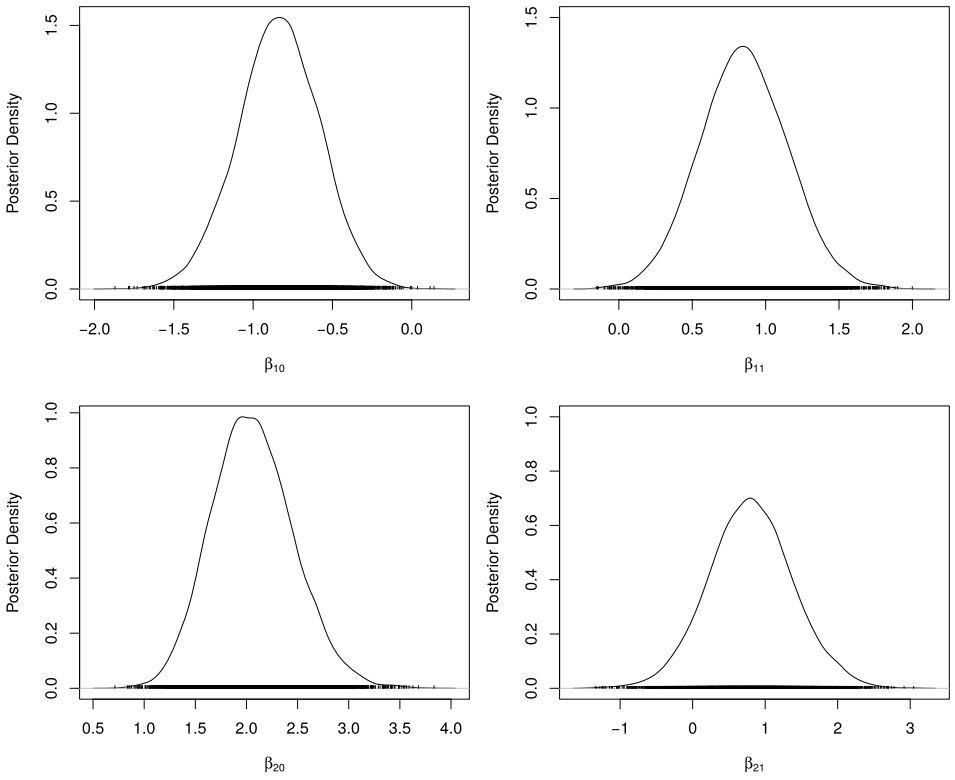


Figure 5 *Estimated* posterior *densities of vectors* $\boldsymbol{\beta}_1$ *and* $\boldsymbol{\beta}_2$.

one can re-estimate it not depending on individuals covariate. Hence, the final predictive $\mathcal{ZMPL}$ model is given by

$$\widehat{\mu}_i^* = \frac{\widehat{\omega}\widehat{\mu}_i}{\mathrm{P}(Y > 0; \widehat{\mu}_i)},$$

where $\widehat{\mu}_i = \exp\{-0.8395 + 0.8516 x_i\}$ and $\widehat{\omega} = 0.9252$. Also, since parameter μ_i was estimated using only positive observations, if at least one additional bid is being offered, its expected value will be 1.312 provided that no additional bidder is invited and 1.808 otherwise.

Table 9 presents the final *posterior* summary of the fitted models. One can notice that the estimates for n_0, obtained from the $\mathcal{P}$, $\mathcal{NB}$ and $\mathcal{PL}$ models are much larger than the real one while those provided by zero-modified models are very close (or exactly equal) to 9. Through these measures, one can better understand how the fitted models are adhering to the data since the nature of the observed counts should be well described regarding its frequency and the average number of nonzero observations. The goodness-of-fit can be evaluated by the χ^2 measure obtained from the observed and expected frequencies. To compute such statistic, we

Table 9 *Bayesian estimates for the extra parameters and goodness-of-fit evaluation*

	Full dataset				Without observation 36			
Model	$\widehat{\mu}^*$	$\widehat{p}$	$\widehat{n}_0$	χ^2	$\widehat{\mu}^*$	$\widehat{p}$	$\widehat{n}_0$	χ^2
$\mathcal{P}$	1.65	1.00	24	26.19 (<0.001)	1.58	1.00	26	24.65 (<0.001)
$\mathcal{NB}$	1.67	1.00	25	27.50 (<0.001)	1.62	1.00	25	25.75 (<0.001)
$\mathcal{PL}$	1.60	1.00	45	69.67 (<0.001)	1.56	1.00	46	70.59 (<0.001)
$\mathcal{ZMP}$	1.88	1.41	10	8.75 (0.119)	1.78	1.43	10	4.31 (0.366)
$\mathcal{ZMPL}$	1.84	2.22	10	0.61 (0.986)	1.77	2.27	9	0.65 (0.957)

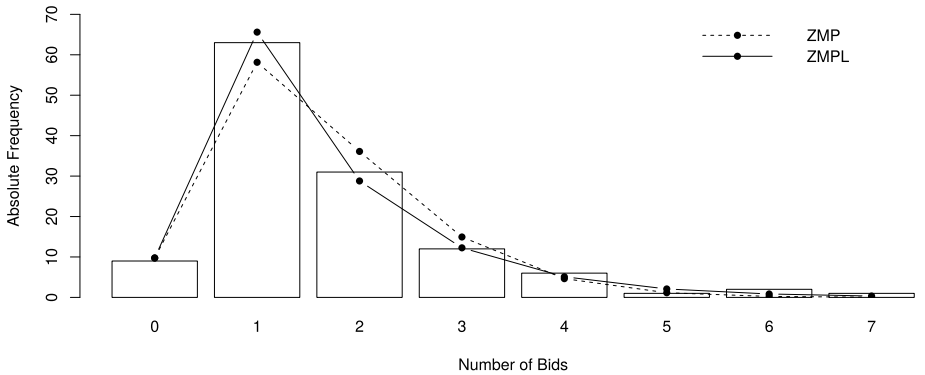


Figure 6 *Expected frequencies estimated under zero-modified models.*

have grouped cells with frequencies lower or equal than 5, resulting in 5 d.f. (full dataset) and 4 d.f. (when removing observation 36). One can notice that $\mathcal{ZMP}$ model provide reliable fits, but it is quite clear that the proposed model adheres much better on the considered datasets (p-values greater than 0.95).

Figure 6 depicts the expected frequencies estimated through predictive $\mathcal{ZMP}$ and $\mathcal{ZMPL}$ models, considering the removal of the influential point. The results highlight the better adherence of the fitted models when observation 36 is discarded in the estimation procedure. Besides, using the χ^2 statistic and the comparison criteria, one can notice that the proposed model provides a more realistic fit for the considered dataset and inferences about parameter p allow us to classify the observed sample as being zero-deflated ($\widehat{p} > 1$). In other words, we have that, by the proposed model, it would be expected that more firms would not have received additional bids after the initial bid.

8 Concluding remarks

The excess or deficit of zeros-valued observations is an issue often encountered in real applications involving count data. In such a way, this paper aimed to introduce the $\mathcal{ZMPL}$ regression model as an alternative for the analysis of overdispersed datasets exhibiting zero-inflation/deflation in the presence of covariates. By using the hurdle version of the $\mathcal{ZMPL}$ distribution, it was possible to write separable likelihood functions for the parameter vectors, which led us to less complicated Bayesian procedures based on the *g-prior* method. Also, we have shown that the mean of the $\mathcal{ZMPL}$ model can be estimated using only the positive observations.

An intensive Monte Carlo simulation study was performed in order to evaluate the empirical properties of the Bayesian estimators and MLEs, and the obtained results highlighted the suitability of the adopted methodology. Due to the "vague" nature of the *prior* distributions, similar results were achieved, but the Bayesian approach remains an excellent option since does not depend on asymptotic results for inference and has the advantage of incorporating specific information about parameters when available.

The proposed model was considered for the analysis of a real dataset obtained from an economic study with legal implications, where the response variable is the number of additional bids after the initial bid received by 126 U.S. firms. The response variable was identified as being overdispersed and zero-deflated, which justifies the use of the $\mathcal{ZMPL}$ model. A sensitivity analysis was conducted using the Kullback–Leibler divergence, and one firm was identified as locally influent. The main inferential conclusion one can take from the fitted model is that the white knight is statistically relevant to describe the average number of additional bids. In addition, when looking at the χ^2 statistic and the *posterior* based comparison criteria, we have noticed that the proposed model had presented a better fit when compared with its competitors, and therefore, the $\mathcal{ZMPL}$ regression model can be considered an excellent addition to the set of models that can be used when analyzing overdispersed and zero-modified count data.

Appendix A: Influential points

The identification of influential observations is one of the essential steps in any statistical analysis. Usually, the presence of influential points impacts the inferential procedures and the subsequent conclusions considerably. In this way, this subsection is dedicated to presenting some case deletion Bayesian diagnostic measures that can be useful to quantify the influence of each observation in a given dataset.

The computation of divergence measures between *posterior* distributions is a very useful way to quantify influence. According Csiszár (1967), the φ-divergence measure between two densities f and g for $\boldsymbol{\theta} \in \mathcal{D}$ is defined by

$$d_\varphi = \int_{\mathcal{D}} g(\boldsymbol{\theta}) \varphi\left[\frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\right] \mathrm{d}\boldsymbol{\theta}, \tag{A.1}$$

where φ is a smooth convex, lower semicontinuous function such that $\varphi(1) = 0$. Some popular divergence measures can be obtained by choosing specific functions for φ (Peng and Dey, 1995). We are interested in the well-known Kullback–Leibler divergence that can be obtained by setting $\varphi(t) = -\log(t)$.

Suppose that we are studying a discrete random variable Y whose distribution is indexed by a parametric vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Aiming to estimate such vector, we have observed n independent values of Y, hence obtaining the full observed vector $\boldsymbol{y}$. Now, let $\boldsymbol{y}_{(-i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ be a vector obtained after removal of the ith observation from $\boldsymbol{y}$. Given a *prior* distribution $\pi(\boldsymbol{\theta})$, the full *posterior* density of $\boldsymbol{\theta}$ can be expressed as

$$\pi(\boldsymbol{\theta}; \boldsymbol{y}) = \frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}},$$

where $\mathcal{L}$ stands for the likelihood function of $\boldsymbol{\theta}$. Conversely, using the vector without the ith observation, the *posterior* distribution of $\boldsymbol{\theta}$ can be written as

$$\pi[\boldsymbol{\theta}; \boldsymbol{y}_{(-i)}] = \frac{\mathcal{L}[\boldsymbol{\theta}; \boldsymbol{y}_{(-i)}]\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \mathcal{L}[\boldsymbol{\theta}; \boldsymbol{y}_{(-i)}]\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}.$$

Now, taking $f(\boldsymbol{\theta}) = \pi[\boldsymbol{\theta}; \boldsymbol{y}_{(-i)}]$ and $g(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}; \boldsymbol{y})$, equation (A.1) becomes

$$d_\varphi = \mathbb{E}_{\boldsymbol{\theta}}\left\{\varphi\left[\frac{\mathbb{E}_{\boldsymbol{\theta}}^{-1}[\mathrm{P}^{-1}(Y_i = y_i; \boldsymbol{\theta}); \boldsymbol{y}]}{\mathrm{P}(Y_i = y_i; \boldsymbol{\theta})}\right]; \boldsymbol{y}\right\}, \tag{A.2}$$

where $\mathbb{E}_{\boldsymbol{\theta}}^{-1}[\mathrm{P}^{-1}(Y_i = y_i; \boldsymbol{\theta}); \boldsymbol{y}]$ is the conditional predictive ordinate (CPO) statistic (Geisser, 1993) for the ith observation. Given a sample $\{\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^M\}$ from the *posterior* distribution $\pi(\boldsymbol{\theta}; \boldsymbol{y})$, a Monte Carlo estimator for the CPO_i is given by

$$\widehat{\mathrm{CPO}}_i = \left[\frac{1}{M}\sum_{k=1}^{M} \mathrm{P}^{-1}\big(Y_i = y_i; \boldsymbol{\theta}^{(k)}\big)\right]^{-1}, \tag{A.3}$$

and hence, one can estimate the local influence of a particular observation y_i on the *posterior* distribution of $\boldsymbol{\theta}$ as

$$\widehat{d}_\varphi = \frac{1}{M}\sum_{k=1}^{M} \varphi\left[\frac{\widehat{\mathrm{CPO}}_i}{\mathrm{P}(Y_i = y_i; \boldsymbol{\theta}^{(k)})}\right].$$

From equation (A.2), one can notice that, if $\pi[\boldsymbol{\theta}; \boldsymbol{y}_{(-i)}] = \pi(\boldsymbol{\theta}; \boldsymbol{y})$, then there is no divergence caused by the observation y_i. In practice, however, it may be quite difficult to define a threshold value for the divergence measure in order to decide about the magnitude of the influence. A measure of calibration for the Kullback–Leibler divergence was proposed by McCulloch (1989). The idea is based on the typical toy binary experiment of tossing a coin once and observing its upper face is observed. This experiment can be described by $\mathrm{P}(Y = y; \rho) = \rho^y(1-\rho)^{1-y}$, where $\rho \in [0, 1]$ is the probability of success. Regardless what success means, if

the coin is unbiased, then $\mathrm{P}(Y = y; \rho) = 0.50$, $y \in \{0, 1\}$. Thus, the φ-divergence between a (possibly) biased and an unbiased coin is given by

$$d_\varphi(\rho) = \frac{\varphi(2\rho) + \varphi[2(1-\rho)]}{2},$$

from which one may conclude that the divergence between two *posteriors* distributions can be associated with the biasedness of a coin (Peng and Dey, 1995). By analogy, this implies that predict unobserved responses through $\pi[\boldsymbol{\theta}; \boldsymbol{y}_{(-i)}]$ instead of $\pi(\boldsymbol{\theta}; \boldsymbol{y})$ is equivalent to describe a not observed event as having probability ρ_i, when the correct probability is 0.50. For the Kullback–Leibler divergence, we have

$$d_\varphi(\rho_i) = -\frac{1}{2}\log[4\rho_i(1-\rho_i)].$$

The function $d_\varphi(\rho)$ is symmetric about $\rho = 0.50$ and increases as ρ moves away from 0.50. Also, $\inf_{\rho\in(0,1)} d_\varphi(\rho) = 0$, which is attained at $\rho = 0.50$ since $d_\varphi(0.50) = \varphi(1) = 0$. Therefore, a general measure of calibration based on a φ-divergence can be obtained by solving equation $2d_\varphi(\rho) - \varphi(2\rho) - \varphi[2(1-\rho)] = 0$. A Monte Carlo estimator for the calibration measure (ρ) associated with the Kullback–Leibler divergence is given by

$$\widehat{\rho}_i = \frac{1}{2}\big[1 + \sqrt{1 - e^{-2\widehat{d}_i}}\big],$$

where the local influence of each y_i can be estimated by

$$\widehat{d}_i = \frac{1}{M}\sum_{k=1}^{M} \log\big[\mathrm{P}\big(Y_i = y_i; \boldsymbol{\theta}^{(k)}\big)\big] - \log(\widehat{\mathrm{CPO}}_i).$$

One can notice that $\rho_i \in [\frac{1}{2}, 1]$ and therefore, for $\rho_i \gg 0.50$, the ith observation can be considered as an influential point. For example, if $\rho_i \geq 0.80$ is considered a significative bias, then the ith observation will be classified as being influential if $\widehat{d}_i > 0.223$ ($d_\varphi(0.80) \approx 0.223$).

Appendix B: Model comparison

There are several methods for Bayesian model selection that are useful to compare competing models fitted to the same dataset. One of the most used criteria is the deviance information criterion (DIC), which was proposed to work simultaneously as a measure of fit and complexity of the model. To define such a measure, suppose again that we are studying a discrete random variable Y whose distribution is indexed by a parametric vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and let $\boldsymbol{y}$ as a vector of n independent observations obtained from Y. The DIC criterion is given by

$$\mathrm{DIC} = \bar{\mathrm{D}}(\boldsymbol{\theta}) + \rho_{\mathrm{D}} = 2\bar{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\tilde{\boldsymbol{\theta}}),$$

where $\bar{\mathrm{D}}(\boldsymbol{\theta}) = -2\mathbb{E}[\ell(\boldsymbol{\theta}; \boldsymbol{y})]$ is the *posterior* expectation of the *deviance* and ℓ is the log-likelihood function of $\boldsymbol{\theta}$. In this case, the *deviance* is evaluated at some estimate $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ (e.g., the *posterior* conditional mean). In connection with a measure of model complexity, the criterion considers the measure $\rho_{\mathrm{D}} = \bar{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\tilde{\boldsymbol{\theta}})$, which correspond to the effective number of parameters in the model. One can notice that the computation of $\bar{\mathrm{D}}(\boldsymbol{\theta})$ is a complex numerical problem. In this case, a Monte Carlo estimator for such a measure is given by

$$\bar{\mathrm{D}} = -\frac{2}{M}\sum_{k=1}^{M} \ell(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}),$$

and hence, the DIC can be approximated by

$$\widehat{\mathrm{DIC}} = 2\bar{\mathrm{D}} - \mathrm{D}(\tilde{\boldsymbol{\theta}}).$$

Using the estimate $\bar{\mathrm{D}}$, one can define other measures that can be considered when comparing models. The expected Akaike information criterion (EAIC) and the expected Bayesian information criterion (EBIC) can be estimated as

$$\widehat{\mathrm{EAIC}} = \bar{\mathrm{D}} + 2q \quad \text{and} \quad \widehat{\mathrm{EBIC}} = \bar{\mathrm{D}} + q\log(n),$$

where q is the total number of estimated model parameters. See Carlin and Louis (2010) and Brooks (2002) for further details on these comparison criteria.

Another widely used criterion is derived from the CPO measure, which is based on the cross-validation criterion to compare models. For the ith individual, the CPO can be estimated using equation (A.3). A summary statistic of the estimated CPO's is the log-marginal pseudo-likelihood (LMPL) given by the sum of the logarithms of $\widehat{\mathrm{CPO}}_i$'s. Regarding model comparison, we have that the lower the value of DIC, EAIC and EBIC the better the fit. On the other hand, for the latter criterion, we have that the larger the LMPL, the better the fit. One can notice that, for all the presented criteria, the computation of the likelihood function is a crucial step to estimate such measures.

Acknowledgments

The research of Wesley Bertoli is partially supported by the Federal University of Technology—Paraná and by a doctoral grant from Fundação Araucária (CP 18/2015). The research of Francisco Louzada is supported by the Brazilian organizations Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

Angers, J. F. and Biswas, A. (2003). A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis* **42**, 37–46. MR1963008

Bahn, G. D. and Massenburg, R. (2008). Deal with excess zeros in the discrete dependent variable, the number of homicide in Chicago census tract. In *Joint Statistical Meetings of the American Statistical Association*, 3905–3912.

Bazán, J. L., Torres-Avilés, F., Suzuki, A. K. and Louzada, F. (2017). Power and reversal power links for binary regressions: An application for motor insurance policyholders. *Applied Stochastic Models in Business and Industry* **33**, 22–34. MR3615258

Bertoli, W., Conceição, K. S., Andrade, M. G. and Louzada, F. (2018). On the zero-modified Poisson–Shanker regression model and its application to fetal deaths notification data. *Computational Statistics* **33**, 807–836. MR3809623

Beuf, K. D., Schrijver, J. D., Thas, O., Criekinge, W. V., Irizarry, R. A. and Clement, L. (2012). Improved base-calling and quality scores for 454 sequencing based on a hurdle Poisson model. *BMC Bioinformatics* **13**, 303.

Bohara, A. K. and Krieg, R. G. (1996). A zero-inflated Poisson model of migration frequency. *International Regional Science Review* **19**, 211–222.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde.

Bulmer, M. G. (1974). On fitting the Poisson–Lognormal distribution to species-abundance data. *Biometrics* 101–110.

Carlin, B. P. and Louis, T. A. (2010). *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC.

Chen, M. H. and Ibrahim, J. G. (2003). Conjugate priors for generalized linear models. *Statistica Sinica* 461–476. MR1977737

Cohen, A. C. (1960). An extension of a truncated Poisson distribution. *Biometrics* **16**, 446–450. MR0121900

Conceição, K. S., Andrade, M. G. and Louzada, F. (2013). Zero-modified Poisson model: Bayesian approach, influence diagnostics, and an application to a Brazilian leptospirosis notification data. *Biometrical Journal* **55**, 661–678. MR3097373

Conceição, K. S., Louzada, F., Andrade, M. G. and Helou, E. S. (2017). Zero-modified power series distribution and its hurdle distribution version. *Journal of Statistical Computation and Simulation* **87**, 1842–1862. MR3635033

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* **2**, 299–318. MR0219345

Dietz, E. and Böhning, D. (2000). On estimation of the Poisson parameter in zero-modified Poisson models. *Computational Statistics & Data Analysis* **34**, 441–459.

Geisser, S. (1993). *Predictive Inference: An Introduction*. *Chapman & Hall/CRC Monographs on Statistics & Applied Probability.* Boca Raton: CRC Press. MR1252174

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 457–472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics* **4**, 169–193.

Ghitany, M. E., Al-Mutairi, D. K. and Nadarajah, S. (2008). Zero-truncated Poisson–Lindley distribution and its application. *Mathematics and Computers in Simulation* **79**, 279–287. MR2477530

Ghosh, S. K., Mukhopadhyay, P. and Lu, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference* **136**, 1360–1375. MR2253768

Gupta, M. and Ibrahim, J. G. (2009). An information matrix prior for Bayesian analysis in generalized linear models with high dimensional data. *Statistica Sinica* **19**, 1641–1663. MR2589202

Gurmu, S. and Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics* **14**, 469–477.

Heilbron, D. C. and Gibson, D. R. (1990). Shared needle use and health beliefs concerning AIDS: Regression modeling of zero-heavy count data. Poster session. In *Proceedings of the Sixth International Conference on AIDS*. *San Francisco, CA*.

Hörmann, W., Leydold, J. and Derflinger, G. (2004). *Automatic Nonuniform Random Variate Generation*. Berlin: Springer. MR2143197

Hu, M. C., Pavlicova, M. and Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. *The American Journal of Drug and Alcohol Abuse* **37**, 367–375.

Jaggia, S. and Thosar, S. (1993). Multiple bids as a consequence of target management resistance: A count data approach. *Review of Quantitative Finance and Accounting* **3**, 447–457.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review* **73**, 35–58.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society, Series B (Methodological)* 102–107. MR0095550

McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association* **84**, 473–478.

McDowell, A. (2003). From the help desk: Hurdle models. *Stata Journal* **3**, 178–184.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley. MR1789474

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* **33**, 341–365. MR0867980

Ngatchou-Wandji, J. and Paris, C. (2011). On the zero-inflated count models with application to modelling annual trends in incidences of some occupational allergic diseases in France. *Journal of Data Science* **9**, 639–659. MR2838666

Oliveira, M., Einbeck, J., Higueras, M., Ainsbury, E., Puig, P. and Rothkamm, K. (2016). Zero-inflated regression models for radiation-induced chromosome aberration data: A comparative study. *Biometrical Journal* **58**, 259–279. MR3499114

Peng, F. and Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics* **23**, 199–213.

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ridout, M., Demétrio, C. G. B. and Hinde, J. (1998). Models for count data with many zeros. In *Proceedings of the XIXth International Biometric Conference, Vol. 19*, 179–192.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7**, 110–120. MR1428751

Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communications in Statistics Theory and Methods* **32**, 281–289. MR1958514

Saffar, S. E., Adnan, R. and Greene, W. (2012). Parameter estimation on hurdle Poisson regression model with censored data. *Jurnal Teknologi* **57**, 189–198.

Sankaran, M. (1970). The discrete Poisson–Lindley distribution. *Biometrics* **26**, 145–149.

Shaban, S. A. (1981). On the discrete Poisson–Inverse Gaussian distribution. *Biometrical Journal* **23**, 297–303. MR0629707

Shanker, R. (2016a). The discrete Poisson–Shanker distribution. *Jacobs Journal of Biostatistics* **1**, 1–7.

Shanker, R. (2016b). The discrete Poisson–Sujatha distribution. *International Journal of Probability and Statistics* **5**, 1–9.

Shanker, R. and Fesshaye, H. (2016). On zero-truncation of Poisson, Poisson–Lindley and Poisson–Sujatha distributions and their applications. *Biometrics & Biostatistics International Journal* **3**, 1–13.

Umbach, D. (1981). On inference for a mixture of a Poisson and a degenerate distribution. *Communications in Statistics Theory and Methods* **10**, 299–306. MR0612561

Wagh, Y. S. and Kamalja, K. K. (2018). Zero-inflated models and estimation in zero-inflated Poisson distribution. *Communications in Statistics - Simulation and Computation* **47**, 2248–2265. MR3843056

Zamani, H. and Ismail, N. (2010). Negative Binomial–Lindley distribution and its application. *Journal of Mathematics and Statistics* **6**, 4–9.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti* **6**, 233–243. MR0881437

Zorn, C. J. W. (1996). Evaluating zero-inflated and hurdle Poisson specifications. *Midwest Political Science Association* **18**, 1–16.

W. Bertoli
Department of Statistics
Federal University of Technology—Paraná
Curitiba, PR 80230-901
Brazil
E-mail: wbsilva@utfpr.edu.br

K. S. Conceição
M. G. Andrade
F. Louzada
Department of Applied Mathematics and Statistics
Institute of Mathematical and Computer Sciences
University of São Paulo
São Carlos, SP 13566-590
Brazil
E-mail: katiane@icmc.usp.br
marinho@icmc.usp.br
louzada@icmc.usp.br