

## Estimation and influence diagnostics for zero-inflated hyper-Poisson regression model: full Bayesian analysis

Vicente G. Cancho, Bao Yiqi, Jose A. Fiorucci, Gladys D. C. Barriga & Dipak K. Dey

To cite this article: Vicente G. Cancho, Bao Yiqi, Jose A. Fiorucci, Gladys D. C. Barriga & Dipak K. Dey (2018) Estimation and influence diagnostics for zero-inflated hyper-Poisson regression model: full Bayesian analysis, Communications in Statistics - Theory and Methods, 47:11, 2741-2759, DOI: [10.1080/03610926.2017.1342839](https://doi.org/10.1080/03610926.2017.1342839)

To link to this article: <https://doi.org/10.1080/03610926.2017.1342839>



Accepted author version posted online: 28 Jun 2017.

Published online: 11 Oct 2017.



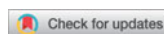
Submit your article to this journal [↗](#)



Article views: 121



View Crossmark data [↗](#)



# Estimation and influence diagnostics for zero-inflated hyper-Poisson regression model: full Bayesian analysis

Vicente G. Cancho<sup>a</sup>, Bao Yiqi<sup>a,b</sup>, Jose A. Fiorucci<sup>a,b</sup>, Gladys D. C. Barriga<sup>a</sup>, and Dipak K. Dey<sup>c</sup>

<sup>a</sup>Sciences institute of mathematics and computers, University of São Paulo, Brazil; <sup>b</sup>Department of Statistics, Federal University of São Carlos, São Carlos, Brazil; <sup>c</sup>Department of Statistics, University of Connecticut, Storrs, USA; <sup>d</sup>Production Engineering Department, São Paulo State University, Bauru, Brazil

## ABSTRACT

The purpose of this paper is to develop a Bayesian analysis for the zero-inflated hyper-Poisson model. Markov chain Monte Carlo methods are used to develop a Bayesian procedure for the model and the Bayes estimators are compared by simulation with the maximum-likelihood estimators. Regression modeling and model selection are also discussed and case deletion influence diagnostics are developed for the joint posterior distribution based on the functional Bregman divergence, which includes  $\psi$ -divergence and several others, divergence measures, such as the Itakura–Saito, Kullback–Leibler, and  $\chi^2$  divergence measures. Performance of our approach is illustrated in artificial, real apple cultivation experiment data, related to apple cultivation.

## ARTICLE HISTORY

Received 7 December 2016  
Accepted 6 June 2017

## KEYWORDS

Bayesian inference;  
hyper-Poisson distribution;  
Kullback–Leibler divergence;  
zero-inflated models.

## MATHEMATICS SUBJECT CLASSIFICATION

62C10

## 1. Introduction

Count data with excess of zeros (or zero-inflated) are commonly encountered in many disciplines, including medicine (Böhning et al. 1999), public health (Zhou and Tu 2000), environmental sciences (Agarwal, Gelfand, and Citron-Pousty 2002), agriculture (Hall 2000), and manufacturing applications (Lambert 1992). Zero-inflation (ZI), a frequent manifestation of overdispersion, means that the incidence of zero counts is generally greater than expected by the proposed model. This is of interest since the incidence of zero counts has practical implication. For instance, Ridout, Hinde, and Demétrio (2001) observe that in counting disease lesions on plants a plant may have no lesions either because it is resistant to the disease, or simply because no disease spores have landed on it. The basic idea behind the derivation of the ZI model is to mix a distribution degenerate at zero with distributions (baseline distributions) such as Poisson, negative binomial and binomial, among others.

In most cases the zero-inflated Poisson (ZIP) model, described in Lambert (1992), has been studied but this raises problem in that the data may suggest additional overdispersion. In such cases, the zero-inflated negative binomial (ZINB) model which mixes a distribution degenerate at zero with a baseline negative binomial distribution maybe considered over the ZIP model. Overdispersion can result from excess zeros or other causes. In any case, the result is excess variability. In some cases, the ZIP model may not be suitable for such

data, since the baseline Poisson model does not accommodate the remaining dispersion unaccounted for in ZIP models and negative binomial (NB) models are more flexible than their simpler Poisson counterparts in accommodating overdispersion (Lawless 1987). The ZINB model has been discussed in Ridout, Hinde, and Demétrio (2001), which provides a score test for testing ZIP regression models against ZINB alternatives, and Mwalili, Lesaffre, and Declerck (2008) illustrates how the ZINB regression model can be corrected for misclassification.

The hyper-Poisson (hP) distribution was first introduced by Bardwell and Crow (1964) for modeling count data characterized by over or underdispersion (Bardwell and Crow 1964; Sáez-Castillo and Conde-Sánchez 2013), and has been evaluated in the context of generalized linear models (GLM) (Sáez-Castillo and Conde-Sánchez 2013). Recently, Sáez-Castillo and Conde-Sánchez (2015) worked with a zero-inflated hyper-Poisson (ZIhP), which is the result of mixing an hP distribution with a degenerate distribution at zero. They included covariates in the mean parameter of hP distribution and focused on the frequentist estimation of the regression coefficients of the parameter for large samples. To simplify interpretation of the results comparison between models did not include covariates in the proportion of structural zeros and was fixed in the simulation study. However, it is often difficult to obtain large samples due to high cost and limited access. Furthermore, in many applications of zero-inflated models, the proportion of structural zeros plays a key role. For this reason, it is important to associate covariates with the proportion of structural zeros.

Accordingly, this study develops a Bayesian inference for the ZIhP regression model based on Markov chain Monte Carlo (MCMC) methods and the Bayes estimators is compared by simulation with the maximum-likelihood estimators. Covariates are included in all parameters of the model, which focuses on the Bayesian estimation of regression coefficients for large samples but also small samples. The simulation indicated that Bayesian estimators of regression coefficients of the proportion of structural zeros are more accurate than classical estimators for ZIhP regression model particularly for small samples.

After fitting a model, it is important to check its assumptions and conduct sensitivity studies to detect influential or extreme observations, which may distort analysis results, leading to the so called diagnostic methods. Following the pioneering work by Cook (1986), case-deletion and local influence diagnostic methods have been widely applied to many regression models. To the best of our knowledge, however, there are no studies on Bayesian inference or influence diagnostics related to ZIhP distribution models. The development of statistical influence diagnostic Bayesian tools for ZIhP distribution models represents a significant contribution to this field. A further study objective is to develop diagnostic measures for proposed model from a Bayesian perspective based on functional Bergman divergence (Goh and Dey 2014).

This paper is organized as follows. Section 2 provides a brief sketch of ZIhP regression models and its properties. Section 3 provides implements Bayesian inference for ZIhP models. Measures of model selection and Bayesian diagnostics based on functional Bergman divergence are given in Sec. 4. Section 5 describes a simulation study conducted to assess the performance of the Bayesian estimator and compare it with the classical estimator. The methodology is illustrated in Sec. 6, in which ZIhP models are compared with several zero-inflated models according to their Bayesian inference and case diagnostics. Finally, concluding remarks are offered in Sec. 7.

### 2. The ZIhP distribution

hP distribution is a generalization of Poisson distribution. Its probability mass function (pmf) can be given by

$$P(Y = y; \phi, \lambda) = \frac{1}{{}_1F_1(1; \phi; \lambda)} \frac{\lambda^y}{(\phi)_y}; y = 0, 1, \dots \tag{1}$$

where  $\phi, \lambda > 0$  and  $(a)_r = a(a + 1) \dots (a + r - 1) = \frac{\Gamma(a+r)}{\Gamma(a)}$  for  $a > 0$  and  $r$  a positive integer and

$${}_1F_1(a; b; w) = \sum_{r=0}^{\infty} \frac{(a)_r w^r}{(b)_r r!}$$

is the confluent hypergeometric series (Slater 1960). The mean and variance of hP distribution are, respectively,

$$E(Y) = \mu = \lambda - (\phi - 1) \frac{{}_1F_1(1; \phi; \lambda) - 1}{{}_1F_1(1; \phi; \lambda)} \text{ and}$$

$$Var(Y) = \lambda + (\lambda - (\phi - 1))\mu - \mu^2 \tag{2}$$

Note that when  $\phi = 1$  in (1), the hP distribution reduces to the Poisson distribution with parameter  $\lambda$ . If  $\phi < 1$ , in 2, there is under-dispersion in relation to Poisson distribution. On the other hand, if  $\phi > 1$  there is overdispersion.

The ZIhP distribution is the result of mixing an hP distribution and a degenerate distribution at zero. Then, the random variable  $Y$  is ZIhP distributed if its pmf is given by

$$Pr(Y_i = y_i) = \begin{cases} \theta + \frac{1-\theta}{{}_1F_1(1; \phi; \lambda)}, & y_i = 0, \\ (1 - \theta) \frac{1}{{}_1F_1(1; \phi; \lambda)} \frac{\lambda^y}{(\phi)_y}, & y_i = 1, 2, \dots \end{cases} \tag{3}$$

where  $\theta$  is the zero-inflated or zero-deflated parameter with negative values. Note that a zero-inflated model has  $0 \leq \theta < 1$  and a zero-deflated model arises when  $\max\{-1, \frac{-{}_1F_1(1; \phi; \lambda)}{1 - {}_1F_1(1; \phi; \lambda)}\} \leq \theta < 0$ . The model in (3) is referred to as the ZIhP model, which is denoted by ZIhP( $\phi, \lambda, \theta$ ). The ZIhP model include the ZIP model. For  $\theta = 0$ , it reduces to hP model with parameters  $\lambda$  and  $\phi$ . The mean and variance of the ZIhP model are, respectively,  $E(Y) = (1 - \theta)\mu$  and  $Var(Y) = (1 - \theta)[\lambda + \mu(\lambda - (\phi - 1) - (1 - \theta)\mu)]$ .

### 3. Bayesian inference

A Bayesian methodology for determining inference is developed for the ZIhP model. The approach is based on MCMC methods.

Suppose that  $y_1, \dots, y_n$  are independent random variables from the ZIhP( $\lambda, \phi, \theta$ ) model. Let  $\delta_i = 1$  if  $y_i = 0$  and  $\delta_i = 0$ , if  $y_i = 1, 2, \dots$ , then the likelihood function of  $\lambda, \phi$  and  $\theta$  is given by

$$L(\lambda, \phi, \theta | \mathcal{D}) = [\theta + (1 - \theta)/{}_1F_1(1; \phi; \lambda)]^{n_0} (1 - \theta)^{n-n_0} \prod_{i=1}^n \left( \frac{1}{{}_1F_1(1; \phi; \lambda)} \frac{\lambda^{y_i}}{(\phi)_{y_i}} \right)^{1-\delta_i} \tag{4}$$

where  $n_0 = \sum_{i=1}^n \delta_i$  and  $\mathcal{D} = \{n, \mathbf{y}, \boldsymbol{\delta}\}$ , with  $\mathbf{y} = (y_1, \dots, y_n)$  and  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ . The  $i$ th element of the set of observations that are zeros derive from two different groups, the degenerated distribution at zero or  $P(Y = 0; \lambda, \phi, \theta)$ . Suppose we define a latent variable  $\Delta$  that

indicates this event. Let  $\Delta_i$  be a  $i$ th latent variable given as

$$\Delta_i = \begin{cases} 1, & \text{w.p. } P(\lambda, \phi, \theta), \\ 0, & \text{w.p. } 1 - P(\lambda, \phi, \theta) \end{cases}$$

where *w.p.* is the abbreviation for “with probability”,  $i = 1, \dots, n_0$ , and  $P(\lambda, \phi, \theta) = \theta \left\{ 1 + \frac{(1-\theta)}{{}_1F_1(1; \phi; \lambda)} \right\}^{-1}$ . Then the likelihood function based on the augmented data  $\mathcal{D}^* = (\mathcal{D}, \mathbf{\Delta})$  where  $\mathbf{\Delta} = (\Delta_1, \dots, \Delta_{n_0})$  is given by

$$L(\lambda, \phi, \theta | \mathcal{D}^*) = \theta^T (1 - \theta)^{n-T} \left[ \frac{1}{{}_1F_1(1; \phi; \lambda)} \right]^{n_0-T} \prod_{i=1}^n \left( \frac{1}{{}_1F_1(1; \phi; \lambda)} \frac{\lambda^{y_i}}{(\phi)_{y_i}} \right)^{1-\delta_i} \tag{5}$$

where  $T = \sum_{i=1}^{n_0} \Delta_i \sim \text{Binomial}(n_0, P(\lambda, \phi, \theta))$ . This likelihood function suggests a natural choice for the following independent priors:  $\theta \sim \text{Beta}(a, b)$  (beta distribution) and  $(\lambda, \phi) \sim \pi(\lambda, \phi)$ , with all the hyperparameters are specified.

Combining the prior distribution and the likelihood function in (5), the joint posterior distribution for  $\lambda, \phi$  and  $\theta$  is given by

$$\begin{aligned} \pi(\lambda, \phi, \theta | \mathcal{D}^*) &= \theta^{T+a-1} (1 - \theta)^{n-T+b-1} \left[ \frac{1}{{}_1F_1(1; \phi; \lambda)} \right]^{n_0-T} \prod_{i=1}^n \left( \frac{1}{{}_1F_1(1; \phi; \lambda)} \frac{\lambda^{y_i}}{(\phi)_{y_i}} \right)^{1-\delta_i} \\ &\times \pi(\lambda, \phi) \end{aligned} \tag{6}$$

Distribution (6) is analytically intractable, but MCMC methods such as the Gibbs sampler and Metropolis–Hasting algorithm can be used to draw samples, from which features of marginal posterior distribution can be inferred. In addition, MCMC sampling enables us to make inferences for any sample size without resorting to asymptotic calculations. The full conditional distributions for MCMC algorithm from the posterior distribution of  $\lambda, \phi$  and  $\theta$  are given by

$$\begin{aligned} \pi(\theta | \lambda, \phi, \mathcal{D}^*) &\sim \text{Beta}(T + a, n - T + b) \\ \pi(\lambda, \phi | \theta, \mathcal{D}^*) &\sim \left[ \frac{1}{{}_1F_1(1; \phi; \lambda)} \right]^{n_0-T} \prod_{i=1}^n \left( \frac{1}{{}_1F_1(1; \phi; \lambda)} \frac{\lambda^{y_i}}{(\phi)_{y_i}} \right)^{1-\delta_i} \pi(\lambda, \phi) \end{aligned} \tag{7}$$

The Metropolis–Hastings algorithm is needed to simulate samples of  $\lambda$  and  $\phi$ , with exception of the ZIP model.

In the case of this model, if  $\lambda \sim \text{Gamma}(c, d)$  (gamma distribution), the steps for the MCMC algorithm are as follows two steps:

- (1) Given  $(\lambda^{(j)}, \theta^{(j)})$  at the  $j$ -stage, we sample  $T^{(j+1)}$ , from binomial( $n_0, P(\lambda^{(j)}, \theta^{(j)})$ ) with  $P(\lambda^{(j)}, \theta^{(j)}) = \theta^{(j)} / [\theta^{(j)} + (1 - \theta^{(j)})e^{-\lambda^{(j)}}]$
- (2) Given  $T^{(j+1)}$ , are obtain sample from  $\theta^{(j+1)} \sim \text{Beta}(T^{(j+1)} + a, n - T^{(j+1)} + b)$ , and  $\lambda^{(j+1)} \sim \text{Gamma}(\sum_{i=1}^n (1 - \delta_i)y_i + c, n - T^{(j)} + d)$

A ZIhP regression model can be defined as follows. Let  $y_1, \dots, y_n$  be independent random variables such that each  $y_i$  for  $i = 1, \dots, n$  has a probability function (3) with parameters  $\theta = \theta_i, \lambda = \lambda_i$  and  $\phi = \phi_i$  depend on vectors of explanatory variables  $\mathbf{x}_i, \mathbf{z}_i$  and  $\mathbf{w}_i$ , respectively. Following Sáez-Castillo and Conde-Sánchez (2013), propose to relate  $\theta_i$  to covariates  $\mathbf{x}_i$  by the logistic link and  $\mu_i$ , and  $\phi_i$  to covariates  $\mathbf{z}_i$  and  $\mathbf{w}_i$ , by the logarithmic link, respectively, i.e.,

$$\log \left( \frac{\theta_i}{1 - \theta_i} \right) = \mathbf{x}_i^\top \boldsymbol{\alpha}, \quad \log(\mu_i) = \mathbf{z}_i^\top \boldsymbol{\beta}, \quad \text{and} \quad \log(\phi_i) = \mathbf{w}_i^\top \boldsymbol{\gamma} \tag{8}$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{p_1})^\top$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_2})^\top$  and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{p_3})^\top$  are unknown parameters. The unknown value  $\lambda_i$  is found by numerical inversion of equation (2). Then the likelihood function of  $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$  is given by

$$L(\boldsymbol{\vartheta}|\mathcal{D}) = \prod_{i=1}^n \left[ \theta_i + \frac{(1 - \theta_i)}{{}_1F_1(1; \phi_i; \lambda_i)} \right]^{\delta_i} \left[ (1 - \theta_i) \frac{1}{{}_1F_1(1; \phi_i; \lambda_i)} \frac{\lambda_i^{\gamma_i}}{(\phi_i)_{\gamma_i}} \right]^{1 - \delta_i} \quad (9)$$

where  $\mathcal{D} = \{n, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}\}$  with  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ ,  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^\top$  and  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^\top$  is the matrix of covariates of order  $n \times p_1$ ,  $n \times p_2$  and  $n \times p_3$ , respectively, and  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ .

To complete the Bayesian specification of the model we need only consider the prior distribution for all the unknown parameters. Since we have no prior information from historical data or from previous experiment, we assume prior independence among the parameters  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\gamma}$ , i.e.,  $\pi(\boldsymbol{\vartheta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})$ , where  $\boldsymbol{\alpha} \sim N_{p_1}(\mathbf{0}, \Omega_1)$ ,  $\boldsymbol{\beta} \sim N_{p_2}(\mathbf{0}, \Omega_2)$  and  $\boldsymbol{\gamma} \sim N_{p_3}(\mathbf{0}, \Omega_3)$  with  $N_k(0, \Omega)$  denoting the (k)-variate normal distribution. Here all the hyperparameters are specified in order to express non informative priors

Combining the prior distribution and the likelihood function in (9), the joint posterior distribution for  $\boldsymbol{\vartheta}$  is obtained as  $\pi(\boldsymbol{\vartheta}|\mathcal{D}) \propto L(\boldsymbol{\vartheta}; \mathcal{D})\pi(\boldsymbol{\vartheta})$ . This joint posterior density is analytically intractable. Accordingly, the inference is based on MCMC simulation methods. In particular, the Gibbs sampler algorithm (see Gelfand and Smith 1990) has proved to be a powerful alternative. To this end, we observed that there is no closed form expression available for any of the full conditional distributions needed to implement Gibbs sampler. Thus the Metropolis–Hastings algorithm was used instead. To implement the Metropolis–Hastings algorithm, we proceed as follows:

- (1) start with any point  $\boldsymbol{\vartheta}_{(0)}$  and stage indicator  $j = 0$ ,
- (2) generate a point  $\boldsymbol{\vartheta}'$  according to the transitional kernel  $Q(\boldsymbol{\vartheta}', \boldsymbol{\vartheta}_j) = N_{p_1+p_2+p_3}(\boldsymbol{\vartheta}_j, \tilde{\Sigma})$ , where  $\tilde{\Sigma}$  is covariance matrix of  $\boldsymbol{\theta}$  is the same at any stage,
- (3) update  $\boldsymbol{\vartheta}_{(j)}$  to  $\boldsymbol{\vartheta}_{(j+1)} = \boldsymbol{\vartheta}'$  with probability  $p_j = \min\{1, \pi(\boldsymbol{\vartheta}'|\mathcal{D})/\pi(\boldsymbol{\vartheta}_{(j)}|\mathcal{D})\}$ , or keep  $\boldsymbol{\vartheta}_{(j)}$  with probability  $1 - p_j$ ,
- (4) repeat steps (2) and (3) by increasing the stage indicator until the process reaches a stationary distribution.

Because of the high computational cost, we implement the MCMC algorithms in C language and the results are analyzed in R language (R Development Core Team 2016) through the “coda” package (Plummer et al. 2006). This computational program is available from the authors request.

### 3.1. Predictive distribution

The distribution of a future observation conditional on the observed data  $\mathcal{D}$  is given by its posterior predictive distribution (Gelman et al. 2004). For the ZIhP regression model the predictive distribution for  $i$ th individual,  $y_{ip}$  is defined as follows:

$$f(y_{ip}|\mathcal{D}) = \int f(y_{ip}|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}|\mathcal{D})d\boldsymbol{\vartheta} \quad (10)$$

where  $f(y_{ip}|\boldsymbol{\vartheta})$  is ZIhP model in (3) with  $\theta_i$ ,  $\lambda_i$  and  $\phi_i$  given in (8). Computing (10) proceeds using composition sampling; given samples  $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(Q)}$  from the posterior distribution  $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ , we sample  $y_{ip}$  from  $f(y_{ip}|\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^{(j)})$  for  $i = 1, \dots, n$  and  $j = 1, \dots, Q$ . The

samples are  $y_{ip}^{(1)}, y_{ip}^{(2)}, \dots, y_{ip}^{(Q)}$  from the posterior predictive distribution of the  $i$ th subject via  $f(y_{ip}|\mathcal{D})$ .

#### 4. Diagnostic methods

When regression modeling is considered, to perform a sensitivity analysis is strongly advisable since it may be sensitive to the underlying model assumptions. Cook (1986) uses this idea in his assessment of influence analysis to suggest that more confidence can be put in a model which is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook and Weisberg 1982) in which the effects are studied by completely removing cases from the analysis. This reasoning will form the basis for our Bayesian global influence methodology and in doing so it will be possible to determine which subjects might be influential for the analysis. Thus model checking and adequacy play an important role in count data modeling with excess of zeros. In this section, some model comparison criteria, Bayesian residuals and a local influence measure from a Bayesian perspective are proposed to check the underlying model and to identify the presence of outliers and/or influential observations.

##### 4.1. Model comparison criteria

There exist a variety of methodologies to compare several competing models for a given data set and to select the one that best fits the data. Here we consider one of the most used in applied Bayesian researches, which is derived from the conditional predictive ordinate (CPO) statistic. For a detailed discussion on the CPO statistic and its applications to model selection see Gelfand, Dey, and Chang (1992) and Geisser and Eddy (1979). Let  $\mathcal{D} = \{n, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}\}$  be the full data and  $\mathcal{D}^{(-i)} = \{n - 1, \mathbf{y}^{(-i)}, \mathbf{X}^{(-i)}, \mathbf{Z}^{(-i)}, \mathbf{W}^{(-i)}, \boldsymbol{\delta}^{(-i)}\}$  denote the data with the  $i$ th observation deleted. In our model, for an observed zero,  $\delta_i = 1$ , we have from Sec. 2 that  $f(y_i|\boldsymbol{\vartheta}) = \theta_i + (1 - \theta_i)/{}_1F_1(1; \phi_i; \lambda_i)$  and, for  $\delta_i = 0$ ,  $f(y_i|\boldsymbol{\vartheta}) = (1 - \theta_i)\lambda_i^{y_i}/{}_1F_1(1; \phi_i; \lambda_i)(\phi_i)^{y_i}$ . We denote the posterior density of  $\boldsymbol{\vartheta}$  given  $\mathcal{D}^{(-i)}$  by  $\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})$ ,  $i = 1, \dots, n$ . For the  $i$ th observation, CPO<sub>*i*</sub> can be written as

$$CPO_i = \int_{\boldsymbol{\vartheta}} f(y_i|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})d\boldsymbol{\vartheta} = \left\{ \int_{\boldsymbol{\vartheta}} \frac{\pi(\boldsymbol{\vartheta}|\mathcal{D})}{f(y_i|\boldsymbol{\vartheta})} d\boldsymbol{\vartheta} \right\}^{-1} \tag{11}$$

A Monte Carlo estimate of CPO<sub>*i*</sub> can be obtained using a single MCMC sample from the posterior distribution  $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ . Let  $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(Q)}$  be a sample of size Q of  $\pi(\boldsymbol{\vartheta}|\mathcal{D})$  after the burn-in. A Monte Carlo approximation of CPO<sub>*i*</sub> (Dey, Chen, and Chang 1997) is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\boldsymbol{\vartheta}^{(q)})} \right\}^{-1}$$

where

$$f(y_i|\boldsymbol{\vartheta}^{(q)}) = \begin{cases} \theta_i^{(q)} + \frac{(1-\theta_i^{(q)})}{{}_1F_1(1; \phi_i^{(q)}; \lambda_i^{(q)})}, & \text{for } \delta_i = 0 \\ \frac{(1-\theta_i^{(q)})\lambda_i^{(q)y_i}}{{}_1F_1(1; \phi_i^{(q)}; \lambda_i^{(q)})(\phi_i^{(q)})^{y_i}}, & \text{for } \delta_i = 1 \end{cases}$$

with  $\theta_i^{(q)} = \exp\{\mathbf{x}_i^\top \boldsymbol{\alpha}^{(q)}\} / (1 + \exp\{\mathbf{x}_i^\top \boldsymbol{\alpha}^{(q)}\})$ ,  $\lambda_i^{(q)}$  is solution of the non linear equation in (2) with  $\mu_i^{(q)} = \exp\{\mathbf{z}_i^\top \boldsymbol{\beta}^{(q)}\}$ , and  $\phi_i^{(q)} = \exp\{\mathbf{w}_i^\top \boldsymbol{\gamma}^{(q)}\}$ . For model comparison the log pseudo

marginal likelihood (LPML) defined by  $LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$  is used. The larger value of  $LPML$ , the better fit.

Other criteria, such as the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002), the expected Akaike information criterion (EAIC) proposed by Brooks (2002), and the expected Bayesian or Schwarz information criterion (EBIC) proposed by (Carlin and Louis 2001), can also be used. These criteria are based on the posterior mean of the deviance, which can be approximated by  $\bar{d} = \sum_{q=1}^Q d(\boldsymbol{\vartheta}^{(q)})/Q$ , where  $d(\boldsymbol{\vartheta}) = -2 \sum_{i=1}^n \log[f(y_i|\boldsymbol{\vartheta})]$ . DIC can be estimated using the MCMC output by  $\widehat{DIC} = \bar{d} + \widehat{\rho}_D = 2\bar{d} - \widehat{d}$ , with  $\rho_D$  is the effective number of parameters, which is defined as  $E\{d(\boldsymbol{\vartheta})\} - d\{E(\boldsymbol{\vartheta})\}$ , where  $d\{E(\boldsymbol{\vartheta})\}$  is the deviance evaluated at the posterior mean and is be estimated as

$$\widehat{D} = d \left( \frac{1}{Q} \sum_{q=1}^Q \boldsymbol{\phi}^{(q)}, \frac{1}{Q} \sum_{q=1}^Q \boldsymbol{\alpha}^{(q)}, \frac{1}{Q} \sum_{q=1}^Q \boldsymbol{\beta}^{(q)} \right)$$

Similarly, the EAIC and EBIC criteria can be estimated by means of  $\widehat{EAIC} = \bar{d} + 2\#(\boldsymbol{\vartheta})$  and  $\widehat{EBIC} = \bar{d} + \#(\boldsymbol{\vartheta}) \log(n)$ , where  $\#(\boldsymbol{\vartheta})$  is the number of model parameters.

### 4.2. Bayesian residual

The Bayesian standardized residual (Gelfand, Dey, and Chang 1992),  $r_i$  based in the conditional predictive ordinate distribution is defined by

$$r_i = \frac{y_i - E(y_i|\mathcal{D}^{(-i)})}{\sqrt{Var(y_i|\mathcal{D}^{(-i)})}} \tag{12}$$

where  $E(y_i|\mathcal{D}^{(-i)})$  and  $Var(y_i|\mathcal{D}^{(-i)})$  are mean and variance, respectively, with respect of distribution of  $y_i|\mathcal{D}^{(i)}$ . Large  $|r_i|$ 's cast doubt upon the model but retaining the sign of  $r_i$  allows patterns of under or over fitting to be revealed.

A Monte Carlo estimate of  $E(y_i|\mathcal{D}^{(-i)})$  and  $Var(y_i|\mathcal{D}^{(-i)})$  can be obtained using a single MCMC sample from the posterior distribution  $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ . Let  $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(Q)}$  be a sample of size  $Q$  of posterior distribution,  $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ , for ZIhP model. A Monte Carlo approximation of  $E(y_i|\mathcal{D}^{(-i)})$  and  $Var(y_i|\mathcal{D}^{(-i)})$  are given, respectively, as

$$\widehat{E}(y_i|\mathcal{D}^{(-i)}) = \widehat{CPO}_i \frac{1}{Q} \sum_{q=1}^Q \frac{E(y_i|\boldsymbol{\vartheta}^{(q)})}{f(y_i|\boldsymbol{\vartheta}^{(q)})} \tag{13}$$

and

$$\widehat{Var}(y_i|\mathcal{D}^{(-i)}) = \widehat{CPO}_i \frac{1}{Q} \sum_{q=1}^Q \frac{E(y_i^2|\boldsymbol{\vartheta}^{(q)})}{f(y_i|\boldsymbol{\vartheta}^{(q)})} - [\widehat{E}(y_i|\mathcal{D}^{(-i)})]^2 \tag{14}$$

The quantities  $E(y_i|\boldsymbol{\vartheta})$  and  $E(y_i^2|\boldsymbol{\vartheta})$  are expressible in closed form for the ZIhP model. Hence for a sample MCMC of size  $Q$  from parameters of ZIhP regression model, this quantities are given by  $E(y_i|\boldsymbol{\vartheta}^{(q)}) = (1 - \theta_i^{(q)})\mu_i^{(q)}$  and  $E(y_i^2|\boldsymbol{\vartheta}^{(q)}) = (1 - \theta_i^{(q)})[\lambda_i^{(q)} + (\lambda_i^{(q)} - (\phi_i^{(q)} - 1))\mu_i^{(q)}]$ , with  $\mu_i^{(q)} = \exp\{\mathbf{z}_i^T \boldsymbol{\beta}^{(q)}\}$  and  $\lambda_i^{(q)}$  is solution of non linear equation,  $\lambda_i^{(q)} - \mu_i^{(q)} - (\phi_i^{(q)} - 1) \frac{{}_1F_1(1; \phi_i^{(q)}; \lambda_i^{(q)}) - 1}{{}_1F_1(1; \phi_i^{(q)}; \lambda_i^{(q)})} = 0$ ,  $\theta_i^{(q)} = \exp\{\mathbf{x}_i^T \boldsymbol{\alpha}^{(q)}\} (1 + \exp\{\mathbf{x}_i^T \boldsymbol{\alpha}^{(q)}\})^{-1}$ , and  $\phi_i^{(q)} = \exp\{\boldsymbol{\omega}_i^T \boldsymbol{\gamma}^{(q)}\}$ ,  $i = 1, \dots, n$ ,  $q = 1, \dots, Q$ . Using (13) and (14), the Monte Carlo estimate the



Bayesian standardized residual,  $r_i$  is given by

$$\widehat{r}_i = \frac{y_i - \widehat{E}(y_i|\mathcal{D}^{(-i)})}{\sqrt{\widehat{Var}(y_i|\mathcal{D}^{(-i)})}}$$

### 4.3. Bayesian case influence diagnostics

In this section, a newly developed approach is used to measure influence diagnostics. Let  $\psi : (0, \infty) \rightarrow \mathbb{R}$  be a strictly convex and differentiable function with  $\psi(1) = 0$ . Then the functional Bregman divergence  $D_\psi$  is defined as

$$D_\psi(f_1, f_2) = \int \{\psi(f_1(x)) - \psi(f_2(x)) - (f_1(x) - f_2(x))\psi'(f_2(x))\}d\nu(x) \tag{15}$$

where  $\psi'$  represents the derivative of  $\psi$ . The functional Bregman divergence includes varieties of distortion functions according to the choice of the convex function. Several options of  $\psi$  are given in the literature. Goh and Dey (2014), for example, considered the following class of convex functions  $\psi_\alpha(x)$  for  $\alpha \in \mathbb{R}$

$$\psi_\alpha(x) = \begin{cases} x \log x - x + 1, & \alpha = 1 \\ -\log x + x - 1, & \alpha = 0 \\ \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha(\alpha - 1)}, & \text{o.w} \end{cases} \tag{16}$$

The functional Bregman divergence becomes well-known divergences according to an appropriate choice of  $\alpha$ . For instance, when  $\alpha = 0$ , the divergence is the Itakura-Saito (IS) divergence. If  $\alpha = 1$ , then it reduces to the Kullback–Leibler (K–L) divergence. For  $\alpha = 2$ , it becomes the  $L^2/2$  Euclidean distance. For Bayesian model diagnostics using a functional Bregman divergence, Goh and Dey (2014) proposed rate and direct comparison approaches.

#### 4.3.1. Rate comparison

Let  $P$  denote the posterior distribution of  $\boldsymbol{\vartheta}$  for full data and  $P_{(-i)}$  denote the posterior distribution of  $\boldsymbol{\vartheta}$  without the  $i$ th case and  $\Phi_\psi(P, P_{(-i)})$  be the  $\psi$ –divergence (Peng and Dey 1995; Weiss 1996) between  $P$  and  $P_{(-i)}$  with a convex function  $\psi(\cdot)$  with  $\psi(1) = 0$ .

$$\begin{aligned} D_\psi(P/P_{(-i)}, 1) &= \int_{\boldsymbol{\vartheta} \in \Theta} \left\{ \psi\left(\frac{\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\mathcal{D})}\right) - \psi(1) - \left(\frac{\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\mathcal{D})} - 1\right)\psi'(1) \right\} \pi(\boldsymbol{\vartheta}|\mathcal{D}) d\boldsymbol{\vartheta}, \\ &= \int_{\boldsymbol{\vartheta} \in \Theta} \psi\left(\frac{\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\mathcal{D})}\right) \pi(\boldsymbol{\vartheta}|\mathcal{D}) d\boldsymbol{\vartheta} - \psi(1) \\ &= \Phi_\psi(P, P_{(-i)}) \end{aligned}$$

Here, if we choose  $\psi(z) = -\log(z)$  it reduces to K–L divergence,  $\psi(z) = (z - 1) \log(z)$  gives  $J$ -distance (the symmetric version of K–L divergence),  $\psi(z) = 0.5|z - 1|$  defines the variational distance or  $L_1$  norm and  $\psi(z) = (z - 1)^2$  defines the  $\chi^2$ -square divergence.

$D_\psi(P/P_{(-i)}, 1)$  can be computed by sampling the posterior distribution of  $\boldsymbol{\vartheta}$  via MCMC methods. Let  $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(Q)}$  be a sample of size  $Q$  of  $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ . Then a Monte Carlo estimate of  $D_\psi(P/P_{(-i)}, 1)$  is given by

$$\widehat{D}_\psi(P/P_{(-i)}, 1) = \frac{1}{Q} \sum_{q=1}^Q \psi\left(\frac{\widehat{CPO}_i}{f(y_i|\boldsymbol{\vartheta}^{(q)})}\right)$$

$D_\psi(P/P_{(-i)}, 1)$  can be interpreted as the  $\psi$ -divergence of the effect of deleting the  $i$ th case from the full data on the joint posterior distribution of  $\boldsymbol{\vartheta}$ . As noted by Peng and Dey (1995) and Weiss (1996), Cancho, Ortega, and Paula (2010) Cancho et al. (2011), it may be difficult to judge the cutoff point of the divergence measure so as to determine whether a small subset of observations is influential. In this context, the calibration proposal given by Peng and Dey (1995) and Weiss (1996) will be used.

**4.3.2. Direct comparison**

Since the functional Bregman divergence measures the discrepancies between two probability densities, the posterior distributions  $P_{(-i)}$  and  $P$  can be directly compared using the functional Bregman divergence as given below:

$$D_\psi(P, P_{(-i)}) = \int_{\boldsymbol{\vartheta} \in \Theta} \{ \psi(\pi(\boldsymbol{\vartheta}|\mathcal{D})) - \psi(\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) - (\pi(\boldsymbol{\vartheta}|\mathcal{D}) - \pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) \times \psi'(\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) \} d\boldsymbol{\vartheta}. \tag{17}$$

In many cases, the above quantity cannot be obtained directly as the integral cannot be expressed in a closed form. Goh and Dey (2014) proposed Gaussian approximation and importance-weighted marginal density estimation (IWMDE) to overcome this problem. The first approximation is reasonable for studies which focuses on GLMs and the second approximation is more flexible since it can be used for any choice of convex function. In this study the IWMDE approximation is used to determine the Bregman divergence in (17).

Suppose  $q(\cdot)$  is a sampling density, then Equation (17) can be expressed as

$$\begin{aligned} D_\psi(P, P_{(-i)}) &= \int_{\boldsymbol{\vartheta} \in \Theta} \{ \psi(\pi(\boldsymbol{\vartheta}|\mathcal{D})) - \psi(\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) - (\pi(\boldsymbol{\vartheta}|\mathcal{D}) - \pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)}))\psi'(\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) \} \frac{q(\boldsymbol{\vartheta})}{q(\boldsymbol{\vartheta})} d\boldsymbol{\vartheta} \\ &= E_{q(\boldsymbol{\vartheta})} \left\{ \frac{ \psi(\pi(\boldsymbol{\vartheta}|\mathcal{D})) - \psi(\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) - (\pi(\boldsymbol{\vartheta}|\mathcal{D}) - \pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)}))\psi'(\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})) }{ q(\boldsymbol{\vartheta}) } \right\} \end{aligned}$$

Therefore, a Monte Carlo estimate of  $D_\psi(P, P_{(-i)})$  using IWMDE is given as

$$\frac{1}{Q} \sum_{q=1}^Q \left\{ \frac{ \psi(\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D})) - \psi(\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}^{(-i)})) - (\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}) - \pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}^{(-i)}))\psi'(\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}^{(-i)})) }{ \pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}) } \right\}$$

where  $\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}) = \frac{f(\mathbf{y}|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta})}{m^{IW}(\mathbf{y})}$ ,  $m^{IW}(\mathbf{y}) = [\frac{1}{Q} \sum_{q=1}^Q \frac{w(\boldsymbol{\vartheta}^{(q)})}{f(\mathbf{y}|\boldsymbol{\vartheta}^{(q)})\pi(\boldsymbol{\vartheta}^{(q)})}]^{-1}$ , and  $w(\cdot)$  is a multivariate Gaussian distribution with mean  $\hat{\boldsymbol{\mu}}$  and variance  $\hat{\boldsymbol{\Sigma}}$ , where

$$\hat{\boldsymbol{\mu}} = \frac{1}{Q} \sum_{q=1}^Q \boldsymbol{\vartheta}^{(q)} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{q=1}^Q (\boldsymbol{\vartheta}^{(q)} - \hat{\boldsymbol{\mu}})(\boldsymbol{\vartheta}^{(q)} - \hat{\boldsymbol{\mu}})^\top$$

and the estimate of the perturbed posterior is given by

$$\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}^{(-i)}) = \frac{\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}) \cdot \widehat{CPO}_i}{f(\mathbf{y}_i|\boldsymbol{\vartheta}^{(q)})}$$

Note that if  $\psi(x) = \psi_\alpha(x)$  and  $\{\boldsymbol{\vartheta}^{(q)}\}_{q=1}^Q$  are samples from a posterior density  $p(\boldsymbol{\vartheta}|\mathbf{y})$ , then a Monte Carlo estimate of the divergence is given as follows:

$$\begin{aligned} \hat{D}_{\psi_\alpha}^{IW}(P, P_{(-i)}) &= \frac{1}{N} \sum_{q=1}^Q \left\{ \frac{\delta(\boldsymbol{\vartheta}^{(q)}, \mathbf{y})^{-1} + \log[\delta(\boldsymbol{\vartheta}^{(q)}, \mathbf{y})] - 1}{\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}^{(-i)})} \right\} && \text{if } \alpha = 0 \\ \hat{D}_{\psi_\alpha}^{IW}(P, P_{(-i)}) &= \frac{1}{N} \sum_{q=1}^Q \{-\log[\delta(\boldsymbol{\vartheta}^{(q)}, \mathbf{y})]\}, && \text{if } \alpha = 1 \\ \hat{D}_{\psi_\alpha}^{IW}(P, P_{(-i)}) &= \frac{1}{N} \sum_{q=1}^Q \left\{ \frac{1 - \alpha\delta(\boldsymbol{\vartheta}^{(q)}, \mathbf{y})^{\alpha-1} + (\alpha - 1)\delta(\boldsymbol{\vartheta}^{(q)}, \mathbf{y})^\alpha}{\alpha(\alpha - 1)[\pi^{IW}(\boldsymbol{\vartheta}^{(q)}|\mathcal{D}^{(-i)})]^{1-\alpha}} \right\} && \text{o.w.,} \end{aligned}$$

where  $\delta(\boldsymbol{\vartheta}^{(q)}, \mathbf{y}) = \widehat{CPO}_i / f(y_i|\boldsymbol{\vartheta}^{(q)})$ .

### 5. Simulation studies

This section presents some simulation studies for the proposed model and compares the performance of the Bayes estimators with that of the maximum-likelihood estimators, obtained by the Newton-Raphson method (NR). The samples were generated according to a ZIhP regression model with the parameter  $\theta_i = \exp(\alpha_0 + \alpha_1x_{i1} + \alpha_2x_{i2}) / (1 + \exp(\alpha_0 + \alpha_1x_{i1} + \alpha_2x_{i2}))$ ,  $\mu_i = \exp(\beta_0 + \beta_1x_{i1} + \beta_2x_{i2})$  and  $\phi_i = \exp(\gamma_1x_{i1} + \gamma_2x_{i2})$ ,  $i = 1, \dots, n$ . Covariates  $\mathbf{x}_1$  are generated as discrete variables, with values drawn from a discrete uniform distribution on {2.2, 4.4, 8.8, 17.6} and  $\mathbf{x}_2$  as binary with the values generated from Bernoulli distribution with parameter 0.5, and the parameter values are set as  $\alpha_0 = -4.00$ ,  $\alpha_1 = -0.01$ ,  $\alpha_2 = 3.80$ ,  $\beta_0 = 1.90$ ,  $\beta_1 = 0.01$ ,  $\beta_2 = -0.38$ ,  $\gamma_1 = 0.1$  and  $\gamma_2 = 3.3$ . Thus the generated samples are close were conducted for each sample size.

The vague information on prior distributions of the parameters is considered, letting  $\alpha_j \sim N(0, 100)$   $j = 0, 1, 2$ ;  $\beta_k \sim N(0, 100)$   $k = 0, 1, 2$  and  $\phi_l \sim N(0, 100)$   $l = 1, 2$ . Thus, our choice is to assume a weakly informative prior. Inasmuch as our prior is informative the posterior is proper. After a burn-in, 2,000 MCMC posterior samples were considered, thus obtaining an effective sample of size 2,000, on which to base the posterior distribution. To monitor the convergence of the Gibbs samples, the Geweke's convergence diagnostic proposed by Geweke (1992) was used.

The results of the simulation are presented in Table 1, in which MC Mean denotes the arithmetic average of the 200 estimates given by  $\sum_{j=1}^{200} \hat{\vartheta}_{ij} / 200$ ; MC bias is the empirical mean of bias;  $\sum_{j=1}^{200} (\hat{\vartheta}_{ij} - \vartheta_i) / 200$  and MC MSE are the empirical mean-squared error  $\sum_{j=1}^{200} (\hat{\vartheta}_{ij} - \vartheta_i)^2 / 200$ , with  $\vartheta_i$  as the respective true value of the estimated parameters  $\hat{\vartheta}_i$ . In a Bayesian simulation study, it can be observed that all estimates are closed for the real values of the parameters as  $n$  is increasing, so that its biases and MSEs approach zero as  $n$  is increasing. Because we consider an a priori distribution with weakly information, the performance of the Bayesian estimates are quite similar to the estimates of maximum likelihood.

#### 5.1. Influence of outlying observations

A principal goal of the study is to demonstrate the need for robust models to address with the presence of outliers. Considering the same parameter values previously posited for two perturbation cases, four datasets of 200 were generated from the ZIhP model, and cases 50 and

**Table 1.** Simulation results for the ZIhP regression model.

n	Parameter	Bayesian			Classical		
		MC Mean	MC Bias	MC MSE	MC MLE	MC Bias	MC MSE
50	$\alpha_0$	-3.2512	0.7488	2.0222	-4.5167	-0.5167	2.6748
	$\alpha_1$	-0.3075	-0.3175	0.8941	-0.0011	-0.0111	0.0059
	$\alpha_2$	3.2956	-0.5044	1.5867	4.3523	0.5523	2.2653
	$\beta_0$	1.8942	-0.0058	0.0163	1.9076	0.0076	0.0158
	$\beta_1$	0.0089	-0.0011	0.0001	0.0097	-0.0003	0.0002
	$\beta_2$	-0.4361	-0.0561	0.1056	-0.3902	-0.0102	0.0513
	$\gamma_1$	-0.3196	-0.2196	0.2684	-0.1699	-0.0699	0.0242
100	$\gamma_2$	2.4237	-0.8763	2.6282	3.2007	-0.0993	0.5298
	$\alpha_0$	-3.9219	0.0781	0.6204	-4.3795	-0.3795	1.2337
	$\alpha_1$	-0.0386	-0.0486	0.0315	0.0083	-0.0017	0.0033
	$\alpha_2$	3.7984	-0.0016	0.4800	4.2586	0.4586	1.1705
	$\beta_0$	1.8917	-0.0083	0.0095	1.8884	-0.0116	0.0083
	$\beta_1$	0.0104	0.0004	0.0001	0.0110	0.0010	0.0001
	$\beta_2$	-0.3919	-0.0119	0.0351	-0.3544	0.0256	0.0256
200	$\gamma_1$	-0.2288	-0.1288	0.1989	-0.1524	-0.0524	0.0226
	$\gamma_2$	2.8713	-0.4287	1.0843	3.2096	-0.0904	0.5033
	$\alpha_0$	-4.0831	-0.0831	0.5592	-4.1818	-0.1818	0.8110
	$\alpha_1$	0.0060	-0.0040	0.0016	0.0112	0.0012	0.0013
	$\alpha_2$	3.9021	0.1021	0.5058	3.9413	0.1413	0.7666
	$\beta_0$	1.8996	-0.0004	0.0038	1.8870	-0.0130	0.0032
	$\beta_1$	0.0098	-0.0002	0.0001	0.0107	0.0007	0.0001
400	$\beta_2$	-0.3811	-0.0011	0.0152	-0.3960	-0.0160	0.0130
	$\gamma_1$	-0.1343	-0.0343	0.0148	-0.1231	-0.0231	0.0025
	$\gamma_2$	3.1093	-0.1907	0.5893	3.4159	0.1159	0.2598
	$\alpha_0$	-3.9751	0.0249	0.2843	-4.0074	-0.0074	0.3719
	$\alpha_1$	0.0097	-0.0003	0.0007	0.0015	-0.0085	0.0006
	$\alpha_2$	3.7600	-0.0400	0.2468	3.9023	0.1023	0.3893
	$\beta_0$	1.9021	0.0021	0.0018	1.8981	-0.0019	0.0019
	$\beta_1$	0.0098	-0.0002	0.0001	0.0104	0.0004	0.0001
	$\beta_2$	-0.3815	-0.0015	0.0051	-0.3750	0.0050	0.0044
	$\gamma_1$	-0.1175	-0.0175	0.0039	-0.1129	-0.0129	0.0009
	$\gamma_2$	3.3106	0.0106	0.1704	3.3387	0.0387	0.1028

150 were selected for perturbation. To create influential observation in the dataset, one or both of the selected cases were chosen, and the response variable perturbed as follows  $\tilde{y}_i = y_i + 4S_y$  for  $i = 50$  and  $150$ , where  $S_y$  is the standard deviations of the  $y_i$ 's. Here, we consider four kind of data sets in the study. Dataset (a): original dataset, without outliers; Dataset (b): data with perturbation at case 50; Dataset (c): data with perturbation at case 150; and Dataset (d): data with perturbation at cases 50 and 150. The MCMC computations and convergence check are made similar to those in the simulation studies. Table 2 reports the posterior mean, standard deviation of the parameters of the ZIhP model for each dataset. Note that only parameters  $\beta_2$  and  $\gamma_2$  are relatively insensitive with the perturbations. On the other hand, other parameters

**Table 2.** Mean and standard deviation (SD) of the parameters of the ZIhP model for each dataset.

Dataset names	Perturbed case	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\beta_0$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$
a	None	-4.6445 (1.0776)	-0.0313 (0.0343)	4.9152 (1.0727)	1.8542 (0.0702)	0.0162 (0.0117)	0.3646 (0.3287)	0.0233 (0.0328)	1.5590 (0.7253)
b	50	-4.6884 (1.0783)	-0.0333 (0.0346)	4.9451 (1.0706)	1.8569 (0.0716)	0.0143 (0.0112)	0.8087 (0.4203)	0.0180 (0.0262)	2.2916 (0.7231)
c	150	-4.7095 (1.0959)	-0.0300 (0.0357)	4.9392 (1.0978)	1.8592 (0.0689)	0.0150 (0.0109)	0.5173 (0.3277)	0.0232 (0.0278)	1.8048 (0.6739)
d	50, 150	-4.7929 (1.1402)	-0.0325 (0.0361)	5.0442 (1.1379)	1.8586 (0.0705)	0.0141 (0.0113)	0.7682 (0.3627)	0.0169 (0.0261)	2.2591 (0.6394)

**Table 3.** Comparison between scheme fitting by using different Bayesian criteria.

Dataset names	DIC	pd	EAIC	EBIC	B
a	859.9418	7.1449	868.7969	895.1834	− 430.3177
b	872.9769	7.1082	881.8687	908.2553	− 437.7647
c	871.9061	7.0842	880.8219	907.2085	− 436.6024
d	884.1129	7.2631	892.8498	919.2364	− 443.0937

have estimates close to each other in all selected cases. The EAIC, EBIC, DIC and LPML Bayesian criteria for each perturbed version of the original dataset are shown in Table 3. The ZIhP model before perturbation under the dataset (a) stands out as the best.

The sample from the posterior distributions of the parameters of the ZIhP model are considered to calculate the  $\psi$ -divergence measures, as described in Sec. 4.3. For each dataset, the four divergence measures, K–L, J,  $L_1$ , and  $\chi^2$ , are calculated as presented in Table 4. Note that before perturbation (dataset (a)), all selected cases are not influential according to all  $\psi$ -divergence measures. However, after perturbation (datasets (b)–(d)), the measures increase indicating that the perturbed cases are influential. To show this result, the K–L measure is plotted in Figure 1. It can be clearly observed that the measure performed well in identifying influential cases, providing greater values when compared to the others.

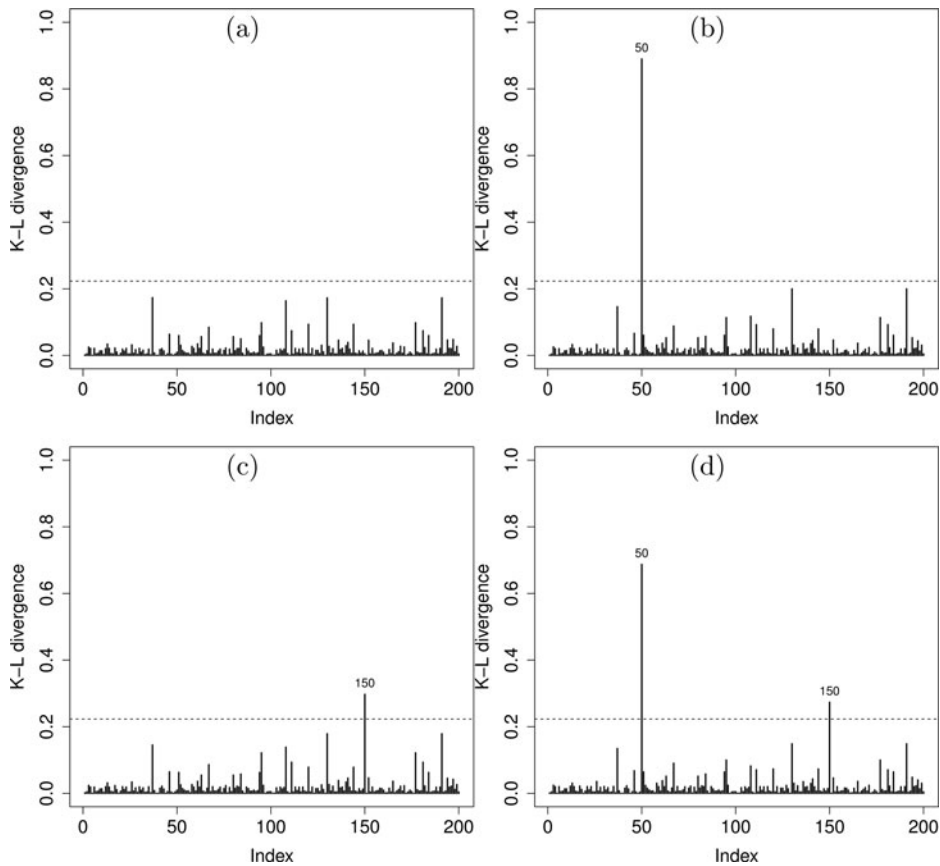
## 6. Application

This section illustrates the study's methodology with the apple cultivar data set which was extracted from Ridout, Hinde, and Demétrio (2001) and analyzed by Galea, Leiva-Sanchez, and Paula (2004) and Barriga and Louzada (2014). The data refer to 270 micropropagated shoots from the columnar apple cultivar Trajan. During the rooting period, the shoots had been produced under an 8- or 16-hour photoperiod in culture media that used one of four concentrations of the cytokinin BAP ( $\mu M$ ). Almost all the shoots rooted in the 8-hour photoperiod. In the 16-hour photoperiod, however, only about half did. The objective of this study is to analyze the influence of BAP concentration and photoperiod on rooting of the apple cultivar. The sample size corresponds to  $n = 270$ , and the percentage of zeros observed was 23.7%. The variables considered in this study include  $y_i$ : number of root in the shoot;  $x_{i1}$ : concentrations of the cytokinin BAP (2.2  $\mu M$ , 4.4  $\mu M$ , 8.8  $\mu M$ , 17.6  $\mu M$ ) and  $x_{i2}$ : photoperiod (0 = 8-hour, 1 = 16-hour), for  $i = 1, \dots, 270$ .

The ZIhP regression model described in Sec. 3 is fitted and the independent priors for the parameters as  $\alpha_j \sim N(0, 10^2)$ ,  $j = 0, 1, 2$ ;  $\beta_k \sim N(0, 10^2)$ ,  $k = 0, 1, 2$ ; and  $\gamma_l \sim N(0, 10^2)$ ,  $l = 1, 2$  are considered. Then the MCMC algorithms are implemented in R language (R Development Core Team 2016), and the likelihood function of the proposed model

**Table 4.**  $\psi$ -divergence measures for the simulated data fitting the ZIhP model.

Dataset names	Case number	$d_{KL}$	$d_J$	$d_{L_1}$	$d_{\chi^2}$
a	50	0.0097	0.0195	0.0557	0.0199
	150	0.0051	0.0103	0.0406	0.0105
b	50	0.8898	1.8359	0.5139	4.9513
	150	0.0051	0.0103	0.0403	0.0106
c	50	0.0096	0.0193	0.0554	0.0195
	150	0.2972	0.6255	0.3069	1.0293
d	50	0.6872	1.4705	0.4569	3.9499
	150	0.2734	0.5854	0.2965	1.0049



**Figure 1.** Index plots of K–L divergence measure for the datasets (a): without perturbation, (b): with perturbation at case 50, (c): with perturbation at case 150 and (d): with perturbation at cases 50 and 150.

is implemented in C language to reduce computational cost. The results were analyzed in R language through the coda package (Plummer et al. 2006). The MCMC algorithms ran 30,000 iterations, discarding the first 10,000 as burn-in and thinning to every fifth. Posterior results are then based on 4,000 iterations of the Markov chain. The Metropolis acceptance rate for these parameters ranged from 20% to 30%. The convergence was confirmed using the Geweke diagnostic. The models were compared using criteria presented in Sec. 4.1.

The posterior means, medians, standard deviations, and 95% highest posterior density (HPD) intervals are presented in Table 5. Note that the covariate of BAP concentration only

**Table 5.** Posterior summaries of the parameters for the ZIhP regression model.

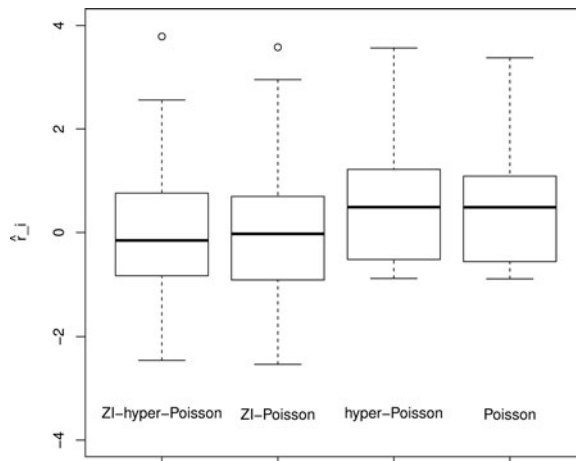
Parameter	Mean	Median	SD	HPD interval (95%)	
				LB	UB
$\alpha_0$	−4.180	−4.135	0.699	−5.507	−2.941
$\alpha_1$	0.009	0.011	0.030	−0.046	0.067
$\alpha_2$	3.78	3.707	0.674	2.481	5.008
$\beta_0$	1.928	1.929	0.054	1.820	2.018
$\beta_1$	0.005	0.005	0.005	−0.004	0.015
$\beta_2$	−0.378	−0.371	0.097	−0.569	−0.187
$\gamma_1$	−0.069	−0.069	0.058	−0.183	0.024
$\gamma_2$	3.244	3.239	0.150	2.983	3.557

**Table 6.** Bayesian criteria for the fitted models.

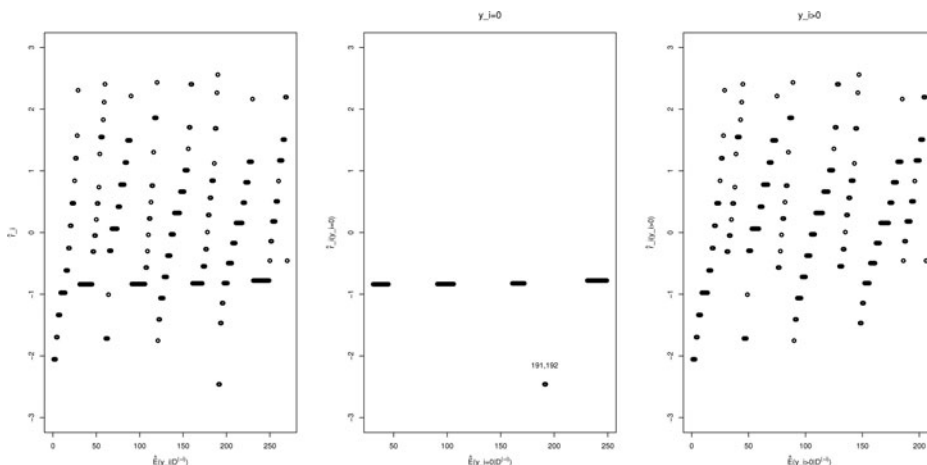
	DIC	EAIC	EBIC	LPML
ZI-hyper-Poisson	1252.82	1262.21	1290.99	− 625.85
ZI-Poisson	1273.05	1279.09	1300.68	− 637.25
hyper-Poisson	1444.27	1450.50	1468.49	− 710.26
Poisson	1440.97	1443.96	1454.76	− 721.37

has a significant effect on dispersion parameter  $\phi$ . However, the covariate of photoperiod is significant in the zero-inflated parameter  $\theta$  which denotes the proportion of zeros, the Poisson parameter  $\lambda$ , and the dispersion parameter.

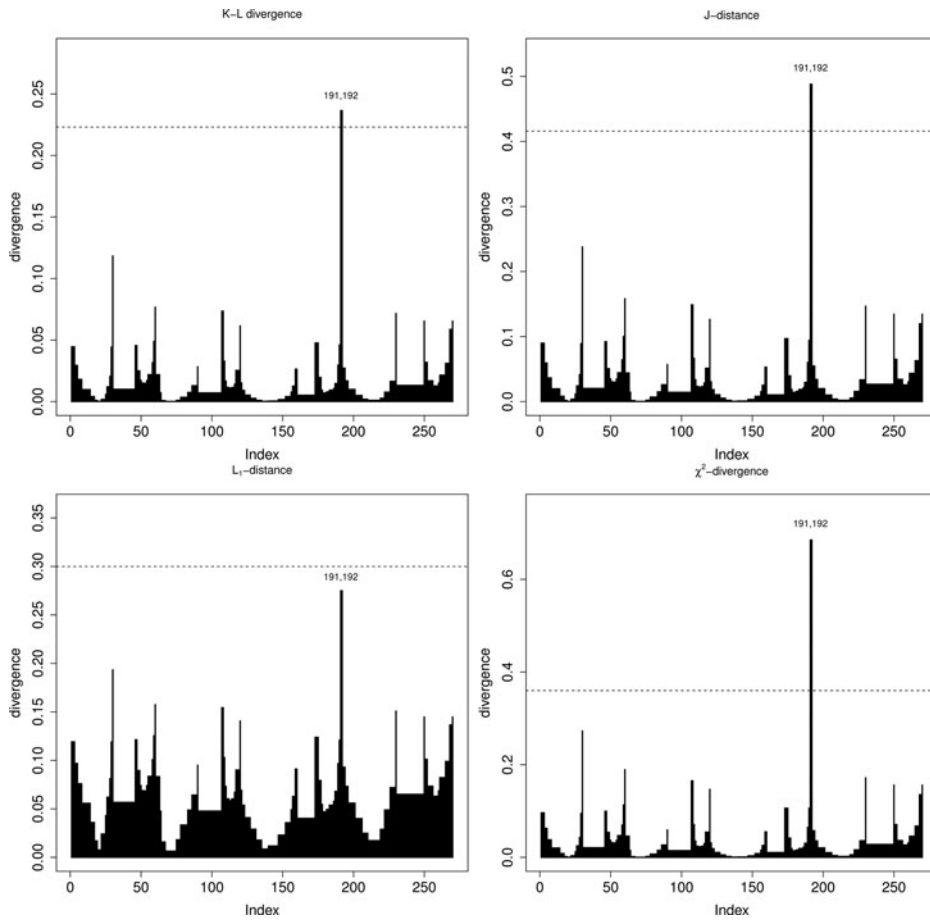
To compare the ZIhP model with its sub-models ZI-Poisson, hP and the Poisson models, we obtained the DIC, EAIC, EBIC and LPML criteria were obtained and furnished the values presented in Table 6. According to all criteria, the ZIhP model, stands out as best. The Bayesian standardized residuals were calculated for all fitted models to check the quality of the fit. Figures 2 and 3 show that the Bayesian standardized residuals of the ZIhP and ZIP



**Figure 2.** Boxplot of Bayesian standardized residuals for fitted models.



**Figure 3.** Bayesian standardized residuals for ZIhP model.



**Figure 4.** Index plots of functional Bregman divergence measures for the ZIhP regression model using the rate comparison approach.

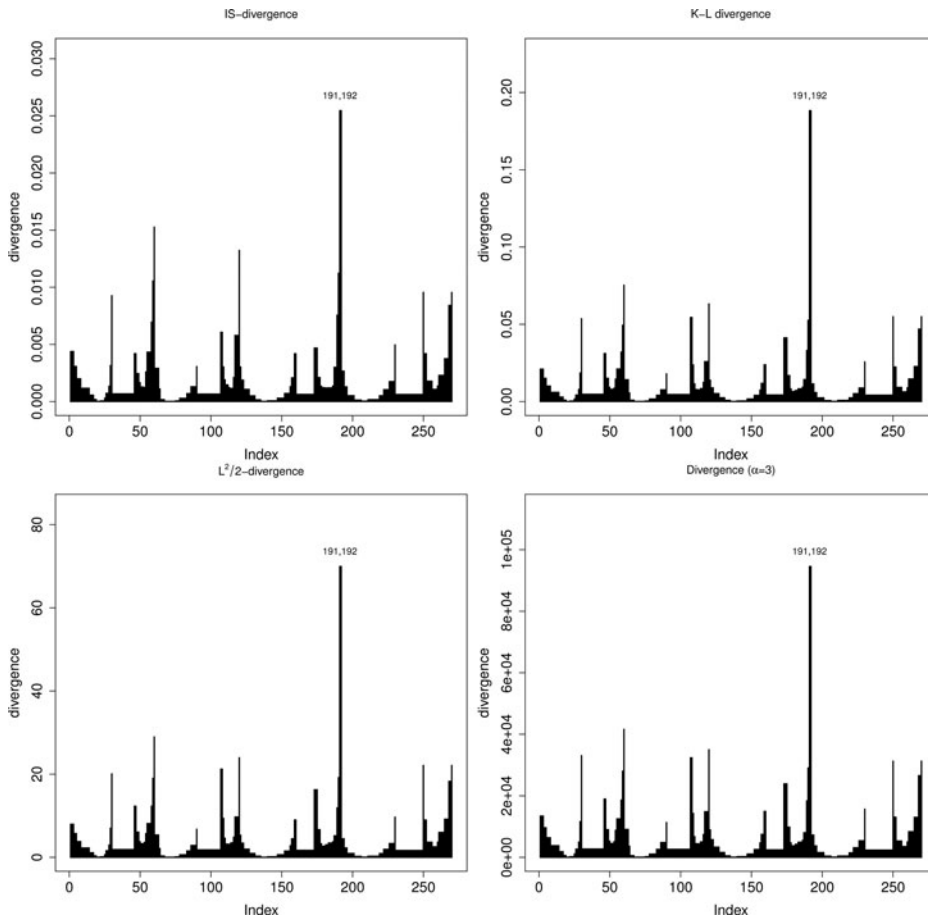
models are randomly distributed around the zero. On the other hand, the standardized residuals of hP, and Poisson models have asymmetric behavior around the zero. For the proposed model, standardized residuals for  $y_i = 0$  and  $y_i > 0$  were plotted separately. Note that observations 191 and 192 stand out from the rest of the data. Accordingly, the ZIhP regression model was selected as the study's working model.

Considering the samples of the posterior distributions of the parameters of the ZIhP regression model, the functional Bregman divergence described in Sec. 4 was computed. Figures 4 and 5 show the index plot of the four functional Bregman divergence measures using the rate and direct comparison approaches. Note that all divergence measures have similar shapes and we can that cases 191 and 192 are possible influential observations in the posteriori distribution.

To reveal the impact of this observation on the parameter estimates, the model was refitted under this circumstance. The relative changes (RC) in the percentage of each parameter estimate were calculated, which is defined by  $RC_{\vartheta_j} = |(\hat{\vartheta}_j - \hat{\vartheta}_{j(I)})/\hat{\vartheta}_j| \times 100$ , where  $\hat{\vartheta}_{j(I)}$  denotes the posterior mean of  $\vartheta_j$ , with  $j = 1, \dots, 8$ , after the observations  $I = \{191, 192\}$  have been removed.

Relative changes in posterior means and the corresponding bracketted 95% HPD intervals for the ZIhP regression model parameters are  $-24.896[-7.502; -3.449]$  for  $\alpha_0$ ,





**Figure 5.** Index plots of Functional Bregman divergence measures for the ZIhP regression model.

237.517[−0.074; 0.051] for  $\alpha_1$ , −33.242[3.258; 7.120] for  $\alpha_2$ , −0.250[1.850; 2.029] for  $\beta_0$ , 13.117[−0.005; 0.013] for  $\beta_1$ , 14.578[−0.491; −0.132] for  $\beta_2$ , 76.453[−0.090; 0.054] for  $\gamma_1$ , and 18.061[2.324; 3.067] for  $\gamma_2$ . Note that the  $\alpha$ 's parameters are more sensitive under deletion of the outstanding observations than others. The parameter  $\alpha_1$ , however, is not significant, and there are we do no inferential changes after removing the observations. The values of DIC, EAIC, EBIC, and LPML criteria for refitted models are 1231.153, 1239.669, 1268.397, and −615.4211, respectively. These are lower than the ZIhP regression model for the data without removing the detected observations. Finally, the model used in the analysis was selected in view of the significant covariates in the link functions, i.e,  $\theta_i = \frac{\exp(\alpha_0 + \alpha_2 x_{2i})}{1 + \exp(\alpha_0 + \alpha_2 x_{2i})}$ ,  $\mu_i = \exp(\beta_0 + \beta_2 x_{2i})$ , and  $\phi_i = \exp(\gamma_2 x_{2i})$ , for  $i = 1, \dots, 270$ .

The posterior summaries of the parameters for the fitted selected model are presented in Table 7. The values of DIC, EAIC, EBIC, and LPML criteria for refitted models are 1229.903, 1235.31, 1253.265, and −615.234, respectively. Barriga and Louzada (2014) proposed the ZICOM-Poisson model and fitted their model for this dataset. After removing the influential observations and non significant covariates, the DIC and LPML values for their model are 1240.2 and −620.2, respectively. Note that the study's proposed model is better than the ZICOM-Poisson model. Moreover, the influential observations were detected by divergence measures, which improves the quality of the model's fit. Table 8 shows the posterior summaries of the  $\theta$ ,  $\lambda$ , and  $\phi$  stratified by photoperiod levels. Note that posterior mean of  $\theta$  for

**Table 7.** Posterior summaries of the parameter for the fitted selected model for data set without individuals 191 and 192.

Parameter	Mean	Median	SD	HPD interval (95%)	
				LB	UB
$\alpha_0$	− 5.325	− 5.215	0.985	− 6.975	− 3.380
$\alpha_2$	5.037	4.901	0.992	3.276	6.890
$\beta_0$	1.971	1.971	0.031	1.911	2.033
$\beta_2$	− 0.350	− 0.349	0.091	− 0.532	− 0.194
$\gamma_2$	2.493	2.487	0.506	1.565	3.491

**Table 8.** Posterior summaries of the zero-inflated parameter  $\theta$ , Poisson parameter  $\lambda$  and dispersion parameter  $\phi$ , stratified by photoperiod levels and median of concentrations of the cytokinin BAP under the ZIhP regression model.

Parameter	Photoperiod	Mean	Median	SD	HPD interval (95%)	
					LB	UB
$\theta$	8-hour	0.007	0.005	0.007	0.000	0.020
	16-hour	0.429	0.430	0.048	0.340	0.521
$\lambda$	8-hour	7.182	7.177	0.225	6.753	7.629
	16-hour	16.805	15.361	6.439	7.471	29.220
$\phi$	8-hour	1.000	1.000	0.000	1.000	1.000
	16-hour	13.775	12.025	7.498	3.350	28.531

the 16-hour photoperiod is 0.429, which means that 42.9% of zeros came from the degenerate distribution and the rest from the hP distribution. Thus the hP model is inappropriate for the dataset.

## 7. Conclusion

This study proposes a Bayesian approach for the ZIhP model as an alternative for count data with an excess of zeros. In the simulation study, it was noted that the bias and mean-squared errors of the Bayesian estimators approach zero as sample size increases, while the mean-squared errors of the classical estimators retained high values for the regression coefficients of the proportion of structural zeros, even if the sample size was large.

Moreover, a case influence diagnostic Bayesian procedure based on two approaches of functional Bregman divergence is proposed to study the sensitivity of the Bayesian estimates under perturbations. Finally, the model is fitted to a real data set to illustrate the method's potential, demonstrating that the ZIhP model delivers the best fit when compared to its sub-models (ZI-Poisson, hP, and Poisson) and the ZICOM-Poisson model after removing the influential observations. Furthermore, the ZIhP model, also permits to include covariates in the equation mean, whereas the traditional ZICOM-Poisson model do that in the location parameter.

## Acknowledgments

The financial support from FAPESP and CNPq is gratefully acknowledged. We thank the associate editor and the reviewer for constructive comments.

## References

- Agarwal, D., A. Gelfand, and S. Citron-Pousty. 2002. Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* 9 (4):341–55.
- Bardwell, G. E., and E. L. Crow. 1964. A two-parameter family of hyper-poisson distributions. *Journal of the American Statistical Association* 59 (305):133–41.
- Barriga, G. D., and F. Louzada. 2014. The zero-inflated Conway–Maxwell–Poisson distribution: Bayesian inference, regression modeling and influence diagnostic. *Statistical Methodology* 21: 23–34.
- Böhning, D., E. Dietz, P. Schlattmann, L. Mendonça, and U. Kirchner. 1999. The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162 (2):195–209.
- Brooks, S. P. 2002. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *Journal of the Royal Statistical Society B* 64:616–8.
- Cancho, V., E. Ortega, and G. Paula. 2010. On estimation and influence diagnostics for log-Birnbaum–Saunders student-t regression models: Full Bayesian analysis. *Journal of Statistical Planning and Inference* 140 (9):2486–96.
- Cancho, V., D. Dey, V. Lachos, and M. Andrade. 2011. Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis* 55 (1):588–602.
- Carlin, B. P., and T. A. Louis. 2001. *Bayes and empirical Bayes methods for data analysis*. 2nd ed. Boca Raton: Chapman & Hall/CRC.
- Cook, R. D. 1986. Assessment of local influence. *Journal of the Royal Statistical Society, Series B* 48: 133–69.
- Cook, R. D., and S. Weisberg. 1982. *Residuals and influence in regression*. Boca Raton, FL: Chapman & Hall/CRC.
- Dey, D., M. Chen, and H. Chang. 1997. Bayesian approach for nonlinear random effects models. *Biometrics* 53:1239–52.
- Galea, M., V. Leiva-Sanchez, and G. Paula. 2004. Influence diagnostics in log-Birnbaum–Saunders regression models. *Journal of Applied Statistics* 31 (9):1049–64.
- Geisser, S., and W. Eddy. 1979. A predictive approach to model selection. *Journal of the American Statistical Association* 74(365):153–60.
- Gelfand, A. E., D. K. Dey, and H. Chang. 1992. Model determination using predictive distributions, with implementation via sampling-based methods. In *Bayesian Statistics 4*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, 147–67, Oxford, London: Oxford University Press.
- Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85 (410):398–409.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian data analysis*. New York, USA: Chapman & Hall/CRC.
- Geweke, J. 1992. *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. Oxford, UK: Clarendon Press.
- Goh, G., and D. K. Dey. 2014. Bayesian model diagnostics using functional Bregman divergence. *Journal of Multivariate Analysis* 124:371–83.
- Hall, D. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56 (4):1030–9.
- Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1):1–14.
- Lawless, J. 1987. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* 15 (3):209–25.
- Mwalili, S., E. Lesaffre, and D. Declerck. 2008. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical Methods in Medical Research* 17 (2):123.
- Peng, F., and D. Dey. 1995. Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics* 23 (2):199–213.
- Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. Coda: Convergence diagnosis and output analysis for MCMC. *R News* 6 (1):7–11.

- R Development, Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ridout, M., J. Hinde, and C. Demétrio. 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* 57 (1):219–23.
- Sáez-Castillo, A., and A. Conde-Sánchez. 2013. A hyper-poisson regression model for overdispersed and underdispersed count data. *Computational Statistics & Data Analysis* 61:148–57.
- Sáez-Castillo, A. J., and A. Conde-Sánchez. 2015. Detecting over-and under-dispersion in zero inflated data with the hyper-poisson regression model. *Statistical Papers* 58 (1):19–33.
- Slater, L. J. 1960. *Confluent hypergeometric functions*. University Press Cambridge.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* 64 (4):583–639.
- Weiss, R. 1996. An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (4):739–50.
- Zhou, X., and W. Tu. 2000. Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics* 56 (4):1118–25.