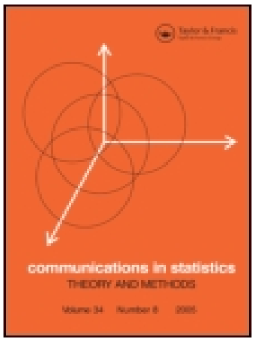

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Bayesian inference and diagnostics in zero-inflated generalized power series regression model

Gladys D. Cacsire Barriga & Dipak K. Dey

To cite this article: Gladys D. Cacsire Barriga & Dipak K. Dey (2016) Bayesian inference and diagnostics in zero-inflated generalized power series regression model, Communications in Statistics - Theory and Methods, 45:22, 6553-6568, DOI: 10.1080/03610926.2014.919397

To link to this article: https://doi.org/10.1080/03610926.2014.919397

Accepted author version posted online: 14 Jul 2016.
Published online: 14 Jul 2016.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2016, VOL. 45, NO. 22, 6553–6568
http://dx.doi.org/10.1080/03610926.2014.919397

Taylor & Francis
Taylor & Francis Group

Bayesian inference and diagnostics in zero-inflated generalized power series regression model

Gladys D. Cacsire Barriga[a] and Dipak K. Dey[b]

[a]Univesidade Estadual Paulista "Julio de Mesquita filho"–FEB, Av. Eng. Luiz Edmundo C. Coube, Bauru, SP, Brazil;
[b]Department of Statistics, University of Connecticut, Storrs, CT, USA

ABSTRACT
The paper provides a Bayesian analysis for the zero-inflated regression models based on the generalized power series distribution. The approach is based on Markov chain Monte Carlo methods. The residual analysis is discussed and case-deletion influence diagnostics are developed for the joint posterior distribution, based on the ψ-divergence, which includes several divergence measures such as the Kullback–Leibler, J-distance, L_1 norm, and χ^2-square in zero-inflated general power series models. The methodology is reflected in a data set collected by wildlife biologists in a state park in California.

ARTICLE HISTORY
Received 15 April 2013
Accepted 25 April 2014

KEYWORDS
Bayesian analysis; Count data; Divergence measures; Generalized power series model; Parameter estimation; Regression model; Zero-inflated model.

MATHEMATICS SUBJECT CLASSIFICATION
97k80

1. Introduction

Zero-inflated count data are commonly encountered in many disciplines including medicine (Böhning et al., 1999), public health (Zhou and Tu, 2000), environmental sciences (Agarwal et al., 2002), agriculture (Hall, 2000), and manufacturing applications (Lambert, 1992). Zero-inflation, a frequent manifestation of overdispersion, means that the incidence of zero counts is generally greater than anticipated. This is of interest because the incidence of zero counts frequently has significance. For instance, Ridout et al. (2001) point out that in assessing lesions on plants, a plant may have no lesions either because it is resistant to disease or it has not been exposed to it. The derivation of the zero-inflated model is derived by mixing a distribution degenerate at zero with such distribution baselines as Poisson, negative binomial, and binomial, among others.

The work on the zero-inflated Poisson (ZIP) model, described in Lambert (1992), has been studied and considered for this type of problem. The data suggest, however, that there is overdispersion not addressed by that model. In such cases, we may consider the zero inflated negative binomial (ZINB) model, which mixes a distribution degenerate at zero with a baseline negative binomial distribution over the ZIP model. Overdispersion can result from excessive zeros or other causes. Regardless, the overdispersion arises from an excess of variability. In some cases, the ZIP model may be inappropriate for use with such data since the Poisson baseline model does not accommodate overdispersion that results from zero-inflation and

CONTACT Gladys D. Cacsire Barriga glad@feb.unesp.br Universidade Estadual Paulista "Julio de Mesquita filho"-FEB, Av. Engenheiro Edmundo Carrijo Coube, 14-01, Bauru-SP, 17033-360, Brazil.

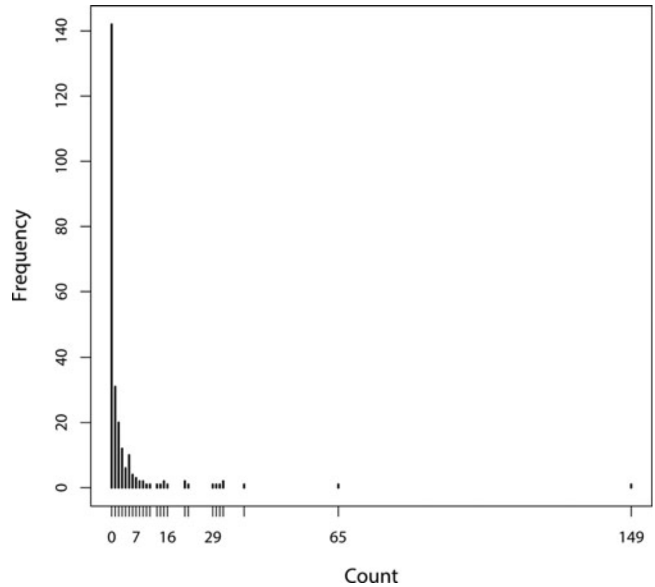

Figure 1. Distribution of number fish caught.

it is well known that negative binomial models are more flexible than their simpler Poisson counterparts in accommodating overdispersion (Lawless, 1987). Ridout et al. (2001) discuss the ZINB model and provide a score test for testing ZIP regression models against ZINB alternatives. Mwalili et al. (2008), on the other hand, demonstrate how the ZINB regression model can be enhanced to correct for misclassifications.

This paper addresses three objectives: (1) to modify the generalized power series (GPS) model introduced by Cordeiro et al. (2009) to the zero-inflated generalized power series (ZIGPS) model; (2) to develop Bayesian inference based on Markov chain Monte Carlo (MCMC) methods for ZIGPS model; and (3) to develop case-deletion diagnostics to detect influential observations in the joint posterior distributions of parameters of ZIGPS regression model.

The research examines data collected by wildlife biologists at a state park in California from groups of people who visited the park. Each group was questioned regarding the number of fish they caught, the number of people in the group, including the number of children, and whether they had brought a camper to the park. Thus a person who fished and did not catch any fish and a person who did not fish were categorized in the same group and consequently accounted for the excess of zeros. Of the 250 groups surveyed, the rate of zeros was 57% as shown in Fig. 1. The fact noted that only one group caught 149 fish could indicate an influential observation.

The presence of influential observations in data analysis is a well-accepted methodological problem, and thus the development of diagnostic measures to detect them is of interest to researchers. Influential observations in a given data set can have a strong impact on statistical inference. Accordingly, these observations are a significant aspect of the data and require careful examination, a common way of assessing the influence of an observation on model fit is case deletion. A common Bayesian diagnostic measure, the Kullback–Leibler divergence (K.L divergence) is based on case deletion and entails a measure of discrepancy between posterior distributions with and without a particular observation. Considerable research has been done for developing case influence diagnostics using K–L divergence under various parametric models. Pettit (1986) suggests its use to detect influential observations. In their review of Bayesian diagnostics, Carlin and Polson (1991) propose an approach using K–L divergence as an expected utility function to define the influence of a set of observations in a parametric modeling framework, considering the normal linear model and mixed models. Weiss and Cook (1992) first applied K–L divergence to assess the divergence between posteriors in the context of case deletion in generalized linear models. Arellano-Valle et al. (2000) used K–L divergence to investigate the influence of a given subset of observations on the posterior distributions of the location and scale parameters of elliptical regression models. Recently, Cho et al. (2009) used K–L divergence to detect influential observations in the joint posterior distributions of parameters of survival models without cure rate. The paper develops case-deletion influence diagnostics for the joint posterior distributions of the parameters of the ZIGPS regression model, based on the ψ-divergence measure (Peng and Dey, 1995; Weiss, 1996). The ψ-divergence measure includes several divergence measures, such as the K–L, J-distance, L_1 norm, and χ^2-square divergence measures.

The paper is organized as follows. In Sec. 2, a brief overview of ZIGPS regression models is presented, providing some of their properties. In Sec. 3, the inference procedure for the proposed model is described. Criteria for model comparison, as well as Bayesian residual and case influence diagnostics, are presented in Sec. 4. An application to a data set is developed in Sec. 5. Finally, Sec. 6 concludes the paper with general remarks.

2. The zero-inflated generalized power series distribution

The discrete random variable Y is said to have a generalized power series (GPS) distribution (Cordeiro et al., 2009) with mean $\mu > 0$ and dispersion parameter $\phi \geq 0$ if the probability mass function (pmf) can be given as

$$P(Y = y) = \frac{a(y, \phi)c(\mu, \phi)^y}{A(\mu, \phi)}, y \in S, \tag{1}$$

where S is a subset of positive integers, $a(y, \phi) > 0$ is positive, and the analytic functions $c(\mu, \phi)$ and $A(\mu, \phi) = \sum_{y \in S} a(y, \phi)c(\mu, \phi)^y$ (of the mean parameter μ and the dispersion parameter ϕ) are positive, finite, and twice differentiable functions. Some distributions of importance belonging to this class are the binomial, Poisson, negative binomial, and generalized Poisson distributions. For example, if k is a positive integer, $a(y, \phi) = \binom{k}{y}$ and $A(\mu, \phi) = (1 + \frac{\mu}{k-\mu})^k$, $c(\mu, \phi) = \frac{\mu}{k-\mu}$, and $S = \{0, 1 \dots, k\}$ then (1) defines the binomial distribution. Examples the remaining distributions in this work, where $S = \{0, 1 \dots, \}$ are as follows:

$$\begin{aligned}
a(y, \phi) &= 1/y!,\ \ c(\mu, \phi) = \mu \text{ and } A(\mu, \phi) = e^{\mu} : \ \text{ Poisson},\\
a(y, \phi) &= \frac{\Gamma(\phi^{-1} + y)}{y!\Gamma(\phi^{-1})};\ \ c(\mu, \phi)\\
&= \frac{\mu}{\phi^{-1} + \mu} \text{and } A(\mu, \phi) = \left(1 - \frac{\mu}{\mu + \phi^{-1}}\right)^{-1/\phi} \ :\ \text{negative binomial},\\
a(y, \phi) &= \frac{(1 + \phi y)^{y-1}}{y!},\\
c(\mu, \phi) &= \frac{\mu e^{-\mu\phi(1+\mu\phi)^{-1}}}{1 + \mu\phi} \text{ and } A(\mu, \phi) = e^{\mu(1+\mu\phi)^{-1}} : \text{generalized Poisson}.
\end{aligned}$$

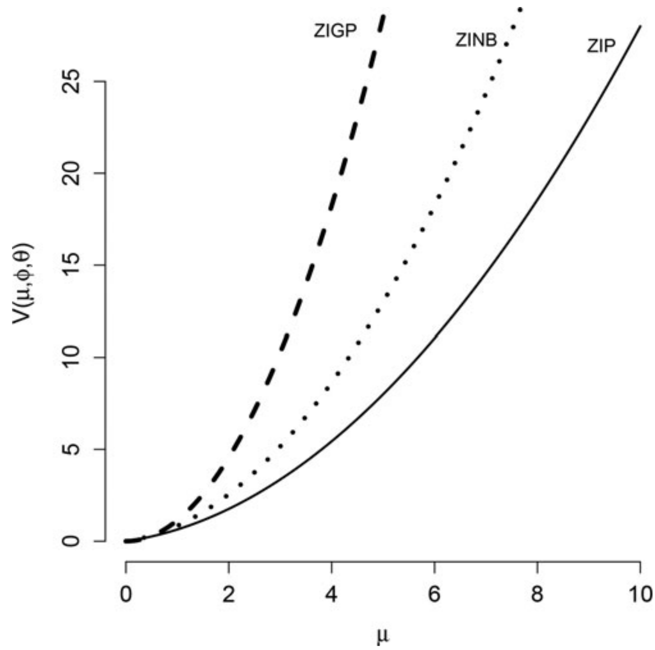

Figure 2. Variance functions for the ZIP, ZINB, and ZIGP models.

The zero-inflated generalized power series (ZIGPS) distribution is the result of mixing a GPS distribution (1) and a degenerate distribution at zero (Johnson et al., 2005). Then the random variable Y is ZIPSG distributed if its pmf is given by,

$$f(y;\mu,\phi,\theta)=\begin{cases}\theta+(1-\theta)\frac{a(0,\phi)}{A(\mu,\phi)}, & y=0\\ \\ (1-\theta)\frac{a(y,\phi)c(\mu,\phi)^y}{A(\mu,\phi)}, & y=1,2,\ldots,\end{cases} \tag{2}$$

where θ is the zero-inflated or zero-deflated parameter, which can take negative values. Note that a zero-inflated model has $0\leq\theta<1$ and a zero-deflated model arises when $-a(0,\phi)/[A(\mu,\phi)-a(0,\phi)]\leq\theta<0$. The model in (2) is referred to as ZIGPS model, which is denoted by ZIGPS(μ,ϕ,θ). The ZIGPS models include the ZIP, ZINB, and zero-inflated generalized Poisson (ZIGP) models, among others. For $\theta=0$, it reduces to a GPS model. The mean and variance of the ZIGPS model are, respectively,

$$E(Y)=(1-\theta)\mu \quad \text{and} \quad Var(Y)=\theta(1-\theta)\mu^2+(1-\theta)\frac{c(\mu,\phi)}{c'(\mu,\phi)},$$

where the primes denote differentiation with respect to μ. Plots of the variance function for ZIGPS distributions in (2) with $\theta=0.6$, $\phi=0.5$, and $\mu\in(0,10)$ are given in Fig. 2, where we observe that the variance of ZINB and ZIGP models is greater than ZIP model.

3. Bayesian inference

A Bayesian methodology for determining inference is developed for the ZIGPS model. The approach is based on MCMC methods. Suppose that $y_1,\ldots,y_n$ are independent random variables from the ZIGPS(μ,ϕ,θ) model. Let $\delta_i=1$ if $y_i=0$ and $\delta_i=0$, if $y_i=1,2,\ldots,$ then the likelihood function of μ, ϕ, and θ is given by

$$L(\mu, \phi, \theta|\mathcal{D}) = [\theta + (1-\theta)a(0,\phi)/A(\mu,\phi)]^{n_0} (1-\theta)^{n-n_0} \prod_{i=1}^{n} \left(\frac{a(y_i,\phi)c(\mu,\phi)^{y_i}}{A(\mu,\phi)} \right)^{1-\delta_i}, \quad (3)$$

where $n_0 = \sum_{i=1}^{n} \delta_i$ and $\mathcal{D} = \{n, \boldsymbol{y}, \boldsymbol{\delta}\}$, with $\boldsymbol{y} = (y_1, \dots, y_n)$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$. The ith element of the set of observations that are zeros come from two different groups, the degenerated distribution at zero or $f(0; \mu, \phi, \theta)$. Suppose we define a latent variable Δ that indicates this event. Let Δ_i be a ith latent variable given as

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

where $w.p.$ is the abbreviation for "with probability," $i = 1, \dots, n_0$, and $P(\mu, \phi, \theta) = \theta\{\theta + \frac{(1-\theta)a(0,\phi)}{A(\mu,\phi)}\}^{-1}$. Then the likelihood function based on the augmented data $\mathcal{D}^* = (\mathcal{D}, \boldsymbol{\Delta})$, where $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_{n_0})$ is given by

$$L(\mu, \phi, \theta|\mathcal{D}^*) = \theta^T (1-\theta)^{n-T} \left[\frac{a(0,\phi)}{A(\mu,\phi)} \right]^{n_0-T} \prod_{i=1}^{n} \left(\frac{a(y_i,\phi)c(\mu,\phi)^{y_i}}{A(\mu,\phi)} \right)^{1-\delta_i}, \quad (4)$$

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

Combining the prior distribution and the likelihood function in Eq. (4), the joint posterior distribution for μ, ϕ, and θ is given by

$$\pi(\mu, \phi, \theta|\mathcal{D}^*) = \theta^{T+a-1} (1-\theta)^{n-T+b-1} \left[\frac{a(0,\phi)}{A(\mu,\phi)} \right]^{n_0-T} \prod_{i=1}^{n} \left(\frac{a(y_i,\phi)c(\mu,\phi)^{y_i}}{A(\mu,\phi)} \right)^{1-\delta_i} \pi(\mu, \phi). (5)$$

Distribution (5) is analytically intractable but MCMC methods such as Gibbs sampler and Metropolis–Hasting algorithm can be used to draw samples, from which features of marginal posterior distribution of interest 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 μ, ϕ, and θ are given by

$$\begin{aligned} &\pi(\theta|\mu, \phi, \mathcal{D}^*) \sim \text{Beta}(T+1, n-T+b) \\ &\pi(\mu, \phi|\theta, \mathcal{D}^*) \sim \left[\frac{a(0,\phi)}{A(\mu,\phi|)} \right]^{n_0-T} \prod_{i=1}^{n} \left(\frac{a(y_i,\phi)c(\mu,\phi)^{y_i}}{A(\mu,\phi)} \right)^{1-\delta_i} \pi(\mu, \phi). \end{aligned} \quad (6)$$

The Metropolis–Hastings algorithm is needed to simulate samples of μ and ϕ for all models except the ZIP model.

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

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

A ZIGPS 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 (2) with parameters $\theta = \theta_i$, $\mu = \mu_i$ and ϕ. We then relate the θ_i to the covariates $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip_1})$ by the logistic link and μ_i to the covariates $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ip_2})$, by the logarithmic link, respectively, that is,

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \boldsymbol{x}_i^\top \boldsymbol{\alpha}, \quad \text{and } \log(\mu_i) = \boldsymbol{z}_i^\top \boldsymbol{\beta}, \tag{7}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p_1})^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p_2})^\top$ are unknown parameters associated with the covariates $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, respectively. Then, the likelihood function of $\boldsymbol{\vartheta} = (\phi, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is given by

$$L(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}) = \prod_{i=1}^{n}\left(\theta_i + (1-\theta_i)\frac{a(0,\phi)}{A(\mu_i,\phi)}\right)^{\delta_i}\left((1-\theta_i)\frac{a(y_i,\phi)g(\mu_i,\phi)^{y_i}}{A(\mu_i,\phi)}\right)^{1-\delta_i}, \tag{8}$$

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

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 the previous experiment, we assume prior independence among the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ϕ, that is, $\pi(\boldsymbol{\vartheta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\phi)$, where $\boldsymbol{\alpha} \sim N_{p_1}(\boldsymbol{0}, \Omega_1)$, $\boldsymbol{\beta} \sim N_{p_2}(\boldsymbol{0}, \Omega_2)$, and $\phi \sim Gamma(a, , b)$ with $N_k(0, \Omega)$ and $Gamma(a, b)$ denoting the (k)-variate normal distribution and the Gamma distribution with mean a/b, respectively. Here all the hyperparameters are specified in order to express non informative priors

Combining the prior distribution and the likelihood function in (8), the joint posterior distribution for $\boldsymbol{\vartheta}$ is obtained as $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}) \propto L(\boldsymbol{\vartheta}; \boldsymbol{\mathcal{D}})\pi(\boldsymbol{\vartheta})$. This joint posterior density is analytically intractable. Accordingly, we based our inference 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. We begin by making a change of variables to $\boldsymbol{\xi} = (\log(\phi), \boldsymbol{\alpha}, \boldsymbol{\beta})$. This transforms the parameter space to $\mathcal{R}^{p_1+p_2+1}$ (which is necessary to work with Gaussian proposal densities). In light of the Jacobian of this transformation, the joint posterior or target density is now

$$\pi(\boldsymbol{\xi}|\boldsymbol{\mathcal{D}}) \propto L(\boldsymbol{\xi}; \boldsymbol{\mathcal{D}})\pi(\boldsymbol{\xi}) \exp\{\xi_1\}.$$

To implement the Metropolis–Hastings algorithm, we proceed as follows:

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

This computational program is available from the authors' request.

3.1. Predictive distribution

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

$$f(y_{ip}|\mathcal{D}) = \int f(y_{ip}|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}|\mathcal{D})d\boldsymbol{\vartheta}, \tag{9}$$

where $f(y_{ip}|\boldsymbol{\vartheta})$ is ZIGPS model in (2) with θ_i and μ_i given in (7). Computing (9) proceeds using composition sampling; given samples $\boldsymbol{\vartheta}^{(1)}, \ldots, \boldsymbol{\vartheta}^{(Q)}$ from the posterior distribution $\pi(\boldsymbol{\vartheta}|\mathcal{D})$, we sample y_{ip} from $f(y_{ip}|\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^{(j)})$ for $i = 1, \ldots, n$ and $j = 1, \ldots, Q$. The samples $y_{ip}^{(1)}, y_{ip}^{(2)}, \ldots, y_{ip}^{(Q)}$ from the posterior predictive distribution of the ith subject, via $f(y_{ip}|\mathcal{D})$.

4. Diagnostic methods

Generally, 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 to motivate his assessment of influence analysis. He suggests 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, a 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 et al. (1992) and Geisser and Eddy (1979). Let $\mathcal{D} = \{n, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}, \boldsymbol{\delta}\}$ be the full data and $\mathcal{D}^{(-i)} = \{n-1, \boldsymbol{y}^{(-i)}, \boldsymbol{X}^{(-i)}, \boldsymbol{W}^{(-i)}, \boldsymbol{\delta}^{(-i)}\}$ denote the data with the ith observation deleted. In our model, for an observed zero, $\delta_i = 1$, we have from Sec. 2 that $f(y_i|\boldsymbol{\vartheta}) = \theta + (1-\theta)\frac{a(0,\phi)}{A(\mu,\phi)}$ and, for $\delta_i = 0$, $f(y_i|\boldsymbol{\vartheta}) = (1-\theta_i)\frac{a(y_i,\phi)g(\mu_i,\phi)^{y_i}}{A(\mu_i,\phi)}$. We denote the posterior density of $\boldsymbol{\vartheta}$ given $\mathcal{D}^{(-i)}$ by $\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})$, $i = 1, \ldots, n$. For the ith observation, CPO_i can be written as

$$CPO_i = \left\{\int_{\vartheta} \frac{\pi(\boldsymbol{\vartheta}|\mathcal{D})}{f(y_i|\boldsymbol{\vartheta})} d\boldsymbol{\vartheta}\right\}^{-1}. \tag{10}$$

A Monte Carlo estimate of CPO_i can be obtained by using a single MCMC sample from the posterior distribution $\pi(\boldsymbol{\vartheta}|\mathcal{D})$. Let $\boldsymbol{\vartheta}^{(1)}, \ldots, \boldsymbol{\vartheta}^{(Q)}$ be a sample of size Q of $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ after the burn-in. A Monte Carlo approximation of CPO_i (Dey et al., 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

$$f(y_i|\boldsymbol{\vartheta}^{(q)}) = \begin{cases} \theta_i^{(q)} + \dfrac{(1-\theta_i^{(q)})a(0,\phi^{(q)})}{A(\mu_i^{(q)},\phi^{(q)})}, & \text{for } \delta_1 = 0 \\ \dfrac{(1-\theta_i^{(q)})a(y_i,\phi_i^{(q)})g(\mu_i^{(q)},\phi^{(q)})^{y_i}}{A(\mu_i^{(q)},\phi^{(q)})}, & \text{for } \delta_i = 1 \end{cases}$$

with $\theta_i^{(q)} = \exp\left\{\boldsymbol{x}_i^\top \boldsymbol{\alpha}^{(q)}\right\}\left(1+\exp\left\{\boldsymbol{x}_i^\top \boldsymbol{\alpha}^{(q)}\right\}\right)^{-1}$, and $\mu_i^{(q)} = \exp\{\boldsymbol{z}_i^\top \boldsymbol{\beta}^{(q)}\}$. For model comparison, we use the log pseudo marginal likelihood (LPML) defined by $LPML = \sum_{i=1}^{n} \log(\widehat{CPO}_i)$. The larger is the value of $LPML$, the better is the fit of the model.

Other criteria as, the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002), the expected Akaike information criterion (EAIC; Brooks (2002)), and the expected Bayesian (or Schwarz) information criterion (EBIC; Carlin and Louis (2001)) can also be used. These criteria are based on the posterior mean of the deviance, which can be approximated by $\overline{d} = \sum_{q=1}^{Q} d(\boldsymbol{\vartheta}_q)/Q$, where $d(\boldsymbol{\vartheta}) = -2\sum_{i=1}^{n} \log\left[f(y_i|\boldsymbol{\vartheta})\right]$. The DIC can be estimated using the MCMC output by $\widehat{DIC} = \overline{d} + \widehat{\rho_d} = 2\overline{d} - \widehat{d}$, with ρ_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

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

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

4.2. Bayesian residual

The Bayesian standardized residual (Gelfand et al., 1992), r_i based in the conditional predictive ordinate distribution is defined by

$$r_i = \frac{y_i - E(y_i|\mathcal{D}^{(-i)})}{\sqrt{Var(y_i|\mathcal{D}^{(-i)})}}, \tag{11}$$

where $E(y_i|\mathcal{D}^{(-i)})$ and $Var(y_i|\mathcal{D}^{(-i)})$ are mean and variance, respectively, with respect to distribution of $y_i|\boldsymbol{\mathcal{D}}^{(i)}$. Large $|r_i|$'s cast doubt upon the model but retaining the sign of r_i allows patters 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 by using a single MCMC sample from the posterior distribution $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})$. Let $\boldsymbol{\vartheta}^{(1)}, \ldots, \boldsymbol{\vartheta}^{(Q)}$ be a sample of size Q of posterior distribution, $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})$, for ZIGPS model. A Monte Carlo approximation of $E(y_i|\mathcal{D}^{(-i)})$ and $Var(y_i|\mathcal{D}^{(-i)})$ are given, respectively, as

$$\widehat{E}(y_i|\mathcal{D}^{(-i)}) = \widehat{CPO}_i\,\frac{1}{Q}\sum_{j=1}^{Q}\frac{E(y_i|\boldsymbol{\vartheta}^{(j)})}{f(y_i|\boldsymbol{\vartheta}^{(j)})}, \tag{12}$$

and

$$\widehat{Var}(y_i|\mathcal{D}^{(-i)}) = \widehat{CPO}_i\,\frac{1}{Q}\sum_{j=1}^{Q}\frac{E(y_i^2|\boldsymbol{\vartheta}^{(j)})}{f(y_i|\boldsymbol{\vartheta}^{(j)})} - \left[\widehat{E}(y_i|\mathcal{D}^{(-i)})\right]^2. \tag{13}$$

The quantities $E(y_i|\boldsymbol{\vartheta})$ and $E(y_i^2|\boldsymbol{\vartheta})$ are expressible in closed form for ZIGPS model. Hence, for a sample MCMC of size Q from parameters of ZIGPS regression model, this quantities are given by $E(y_i|\boldsymbol{\vartheta}^{(q)}) = (1-\theta_i^{(q)})\mu_i^{(q)}$ and $E(y_i^2|\boldsymbol{\vartheta}^{(q)}) = (1-\theta_i^{(q)})[(\mu_i^{(q)})^2 + \frac{c(\mu_i^{(q)},\phi^{(q)})}{c'(\mu_i^{(q)},\phi^{(q)})}]$, with $\theta_i^{(q)} = \exp \boldsymbol{x}_i^\top \boldsymbol{\alpha}^{(q)}\} \left(1 + \exp \boldsymbol{x}_i^\top \boldsymbol{\alpha}^{(q)}\}\right)^{-1}$, and $\mu_1^{(q)} = \exp\{\boldsymbol{z}_i^\top \boldsymbol{\beta}^{(q)}\},$, $i = 1, \ldots, n$, $q = 1, l-dots, Q$. Using (12) and (13), 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

Let $D_\psi(P, P_{(-i)})$ denote the ψ-divergence between P and $P_{(-i)}$, where P denotes the posterior distribution of $\boldsymbol{\vartheta}$ for full data, and $P_{(-i)}$ denotes the posterior distribution of $\boldsymbol{\vartheta}$ without the ith case. Specifically,

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

where ψ is a convex function with $\psi(1) = 0$. Several choices of ψ are given in Dey and Birmiwal (1994). For example, $\psi(z) = -\log(z)$ defines K–L divergence, $\psi(z) = (z-1)\log(z)$ gives J-distance (or 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 χ^2-square divergence.

The relationship between the CPO (10) and the ψ-divergence measure is given by

$$D_\psi(P, P_{(-i)}) = E_{\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}}\left[\psi\left(\frac{CPO_i}{f(y_i|\boldsymbol{\vartheta})}\right)\right], \tag{14}$$

where the expected value is taken with respect to the joint posterior distribution $\pi(\boldsymbol{\vartheta}|\mathcal{D})$.

In particular, the K–L divergence can be expressed by

$$\begin{aligned} D_{\text{K-L}}(P, P_{(-i)}) &= -E_{\boldsymbol{\vartheta}|\mathcal{D}}\left\{\log(CPO_i)\right\} + E_{\boldsymbol{\vartheta}|\mathcal{D}}\left\{\log\left[f(y_i|\boldsymbol{\vartheta})\right]\right\} \\ &= -\log(CPO_i) + E_{\boldsymbol{\vartheta}|\mathcal{D}}\left\{\log\left[f(y_i|\boldsymbol{\vartheta})\right]\right\}. \end{aligned} \tag{15}$$

From (14), we can be compute $D_\psi(P, P_{(-i)})$ by sampling from the posterior distribution of $\boldsymbol{\vartheta}$ via MCMC methods. Let $\boldsymbol{\vartheta}^{(1)}, \ldots, \boldsymbol{\vartheta}^{(Q)}$ be a sample of size Q of $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})$. Then, a Monte Carlo estimate of $K(P, P_{(-i)})$ is given by

$$\widehat{D_\psi}(P, P_{(-i)}) = \frac{1}{Q}\sum_{q=1}^{Q}\psi\left(\frac{\widehat{CPO}_i}{f(y_i|\boldsymbol{\vartheta}^{(q)})}\right). \tag{16}$$

From (16) a Monte Carlo estimate of K–L divergence $D_{K-L}(P, P_{(-i)})$ is given by

$$\widehat{D_{\text{K-L}}}(P, P_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q}\sum_{q=1}^{Q}\log\left[f(y_i|\boldsymbol{\vartheta}^{(q)})\right]. \tag{17}$$

The $D_\psi(P, P_{(-i)})$ can be interpreted as the ψ-divergence of the effect of deleting of ith case from the full data on the joint posterior distribution of $\boldsymbol{\vartheta}$. As pointed by Peng and Dey (1995) and (Weiss, 1996) (see also Cancho et al., 2010, 2011), it may be difficult for a practitioner to judge the cutoff point of the divergence measure so as to determine whether a small subset of

Table 1. Posterior summaries of the parameters for the ZIGPS regression model.

	ZIP model		ZINB model		ZIGP model	
Parameter	Mean	HPD (95%)	Mean	HPD (95%)	Mean	HPD (95%)
ϕ	—	—	2.754	(1.895, 3.696)	0.773	(0.527, 1.053)
α_1	1.142	(0.519, 1.863)	0.788	(− 0.446, 1.974)	0.831	(− 0.297, 2.044)
α_2	− 0.508	(− 0.791, −0.224)	− 1.240	(− 2.041, −0.558)	− 1.161	(− 1.828, −0.496)
β_1	1.596	(1.412, 1.751)	1.372	(0.842, 1.872)	1.670	(0.929, 2.425)
β_2	− 1.036	(− 1.241, −0.857)	− 1.496	(− 1.879, −1.114)	− 1.647	(− 2.221, −1.112)
β_3	0.832	(0.651, 1.026)	0.874	(0.345, 1.430)	0.826	(0.054, 1.573)

observations is influential or not. In this context, we will use the calibration proposal given by Peng and Dey (1995) and Weiss (1996).

5. Application

To illustrate our proposed modeling discussed so far, we consider the data set that were collected by wildlife biologists in a state park of California, which can be found in the web site: "http://www.ats.ucla.edu/stat/data/fish.csv", on groups of visitors that went to park, as already stated in Sec. 1. The number of observation is 250 groups that went to a park. The response variable is how many fish each group caught (count) and the independent variables are how many children were in the group (child), how many people were in the group (persons), and whether or not they brought a camper to the park (camper).

To this data, we fit some members of ZIGPS regression model described in Sec. 2 such as, the ZIP, ZINB, and ZIGP regression models with

$$\theta_i = \frac{\exp\left(\alpha_0 + \alpha_1 \text{people}_i\right)}{1 + \exp\left(\alpha_0 + \alpha_1 \text{people}_i\right)}, \text{ and } \log(\mu_i) = \beta_0 + \beta_1 \text{child}_i + \beta_2 \text{camper}_i, \ i = 1, \ldots, 250.$$

The following independent priors were considered to perform the Metropolis–Hasting algorithm: $\alpha_j \sim N(0, 10^2)$ $j = 0, 1, 2$, $\beta_k \sim N(0, 10^2)$, $k = 0, 1, 2$, and $\phi \sim Gamma(1, 0.001)$ for ZINB and ZIGP regression models. Thus our choice is to assume a minimally but informative prior. Since, our prior is proper the posterior is proper. After a burn-in, we considered 40,000 MCMC posterior samples. We monitored convergence of the Metropolis–Hasting algorithm using the method proposed by Geweke (1992), as well as trace plots. Every $20th$ sample from the 40,000 MCMC posterior samples was used to reduce the autocorrelations and yield better convergence results. The mean and 95% highest posterior density (HPD) intervals for the parameters of the ZIP, ZINB, and ZIGP regression models are shown in Table 1. Note that for these models, the covariate number of peoples has a significant effect on the reduction of the fraction of the visitors that caught no fish (fraction of zeros). Thus, the covariates child and camper have significant effect on the mean fish caught by groups of visitors that went to park.

The quality of fit and presence of possible outliers can be observed by examining the Bayesian residuals (defined in Sec. 2.2) plotted against $\hat{E}(y_i|\mathcal{D}^{(-i)})$. These plots are presented in Fig. 3. Note that the residuals for the three models considered, almost all residuals randomly distributes around zero. Also it is noted almost all that residuals for the ZIP model exhibit greater variability than the ZINB and ZIGP models. Furthermore, the observation 138 stands out from the rest of the data in the three models. In the ZIP model, the observation 89 also stands out of the rest of the data sets, so that these two observations can be considered as

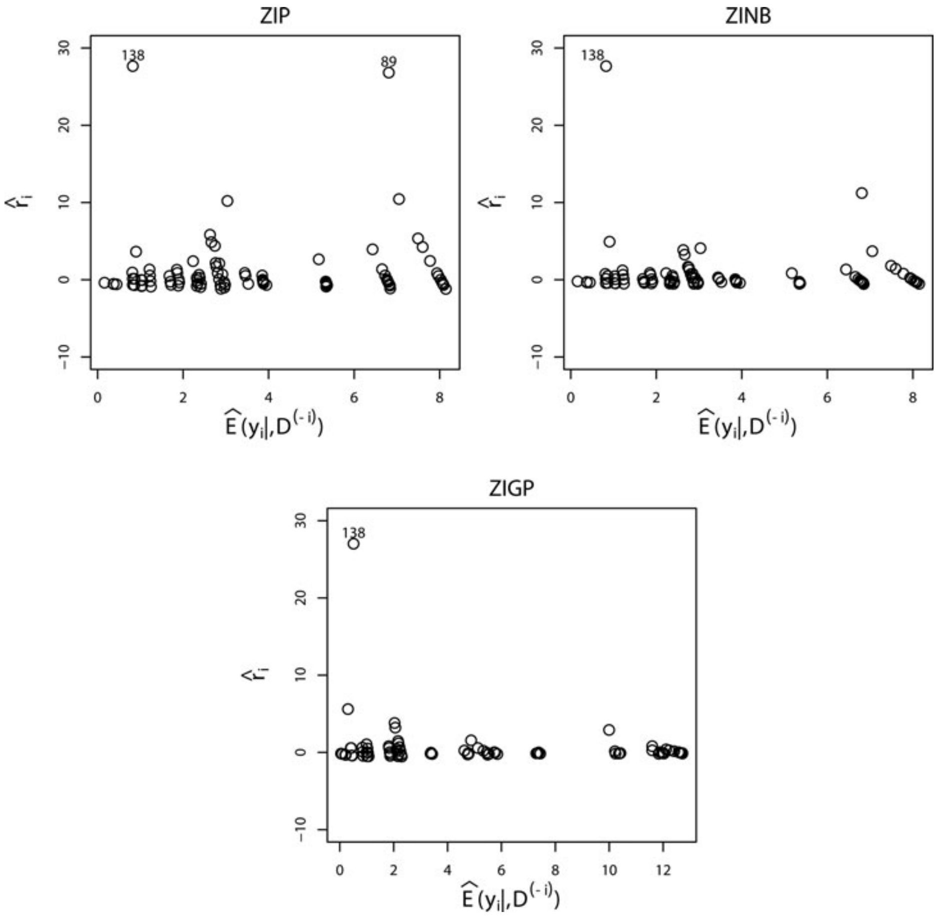

Figure 3. Plot of Bayesian residuals of the ZIGPS regression model.

possible outliers. The sum of squared residuals for the ZIP, ZINB, and ZIGP models resulted 2,008.071, 1,029.894, and 824.9694, respectively, indicating that the ZIGP model is the best fit for the data set.

The models fit is also examined using DIC, EAIC, EBIC, and LPMP criteria and their predictive probabilities of resulting counts are compared with observed counts (Table 2). All criteria favor the ZIGP model over the other considered models; however, there is little difference between the ZINB (−442.088) and ZIGP (−440.118) values computed by LPML. Figure 4 shows the probability integral transformation (PIT) histogram (Czado et al., 2009) for ZIP, ZINB, and ZIGP regression models. The PIT histogram for the ZIP indicates underdispersion while the PIT histogram of ZINB indicates overdispersion. The PIT histogram of ZIGP regression model, however, does not evidence neither under or overdispersion.

Using the samples from the posterior distributions of the parameters of the ZIGPS regression models, the ψ-divergence measures described in Sec. 4.3 are computed. The calibration proposed by Peng and Dey (1995) is used; for example, if we use K–L divergence, we can consider the ith case as an influential observation when $D_{\text{K–L}} > 0.22$. Similarly, using the J-distance, or L_1 norm, or χ^2 square divergence, an observation which $D_{\text{J}} > 0.42$ or $D_{L_1} > 0.30$, or $D_{\chi^2} > 0.36$ can be considered as influential, respectively. Table 3 shows subjects having large ψ-divergence measures values compared to the other subjects in the data

Table 2. Observed frequency distribution and predicted frequencies of ZIGPS models for the real data sets.

Count	Observed	ZIP	ZINB	ZIGP
0	142	134.506	146.246	142.291
1	31	16.849	27.252	33.814
2	20	14.614	14.716	17.141
3	12	12.612	9.775	10.734
4	6	10.970	7.168	7.171
5	10	9.333	5.684	5.281
6	4	7.649	4.619	3.914
7	3	6.232	3.631	3.221
8	2	5.219	3.070	2.580
9	2	4.949	2.664	2.226
10	1	4.970	2.211	1.771
11	1	4.608	1.871	1.541
12	0	4.288	1.760	1.333
13	1	3.659	1.519	1.123
14	1	2.966	1.311	1.008
15	2	2.310	1.189	0.905
16	1	1.602	1.094	0.775
$\geq$17	11	2.664	14.220	13.171
DIC		2,073.406	877.492	870.974
EAIC		2,078.393	884.233	877.377
EBIC		2,096.001	905.362	898.505
LPML		− 1,063.774	− 442.088	− 440.118

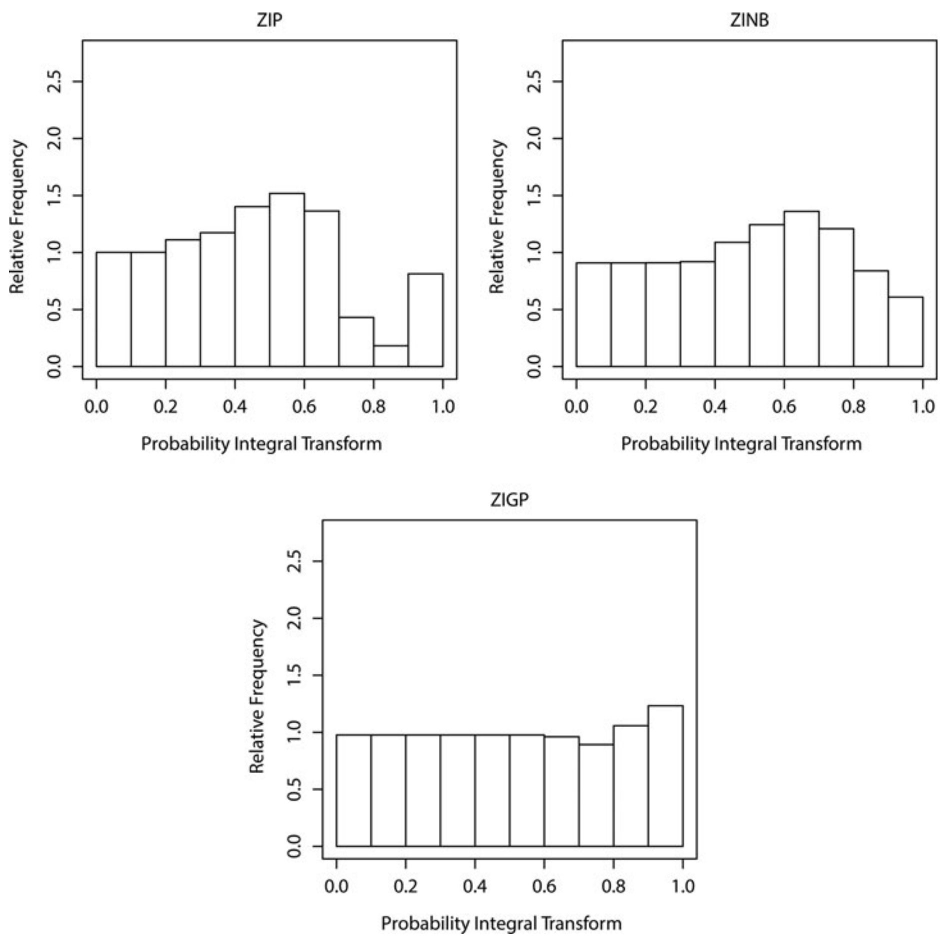

Figure 4. Probability integral transformation histogram for the fish data of the ZIP, ZINB, and ZIGP models.

Table 3. Estimates of the ψ-divergence measures for the fish data fitting the ZIGPS models.

		ZIP model				ZINB model				ZIGP model			
Group	Count	$D_{K\text{–}L}$	D_J	D_{L_1}	D_{χ^2}	$D_{K\text{–}L}$	D_J	D_{L_1}	D_{χ^2}	$D_{K\text{–}L}$	D_J	D_{L_1}	D_{χ^2}
89	149	10.68	15.18	0.95	144.30	0.90	2.17	0.52	16.59	0.20	0.45	0.52	0.83
138	31	5.98	10.70	0.90	371.25	3.16	6.41	0.82	90.74	5.10	9.60	0.82	244.19
219	5	0.22	0.46	0.27	0.67	0.22	0.45	0.26	0.64	0.45	0.95	0.26	1.86

sets. For the ZIP, ZINB, and ZIGP models, case 138 is identified as the most influential. Case 89 is also identified as likely influential observation in the ZIP and ZINB models, cases 100, 160, 186, and 207 only for the ZIP model and case 219 solely for the ZINB model. Figure 5 shows the index plot of the four L_1-divergence measure for ZIGPS model. To reveal the impact of this observation on the parameter estimates, we refitted the model under this situation.

The relative percentage changes for each parameter estimate, defined by $RC_{\vartheta_j} = |(\hat{\vartheta}_j - \hat{\vartheta}_{j(I)})/\hat{\vartheta}_j| \times 100$, where $\widehat{\vartheta}_{j(I)}$ denotes the posterior mean of ϑ_j, with $j = 1, \ldots, 6$, after the observations $I = \{89, 138, 219\}$ has been removed. Relative changes in posterior means and

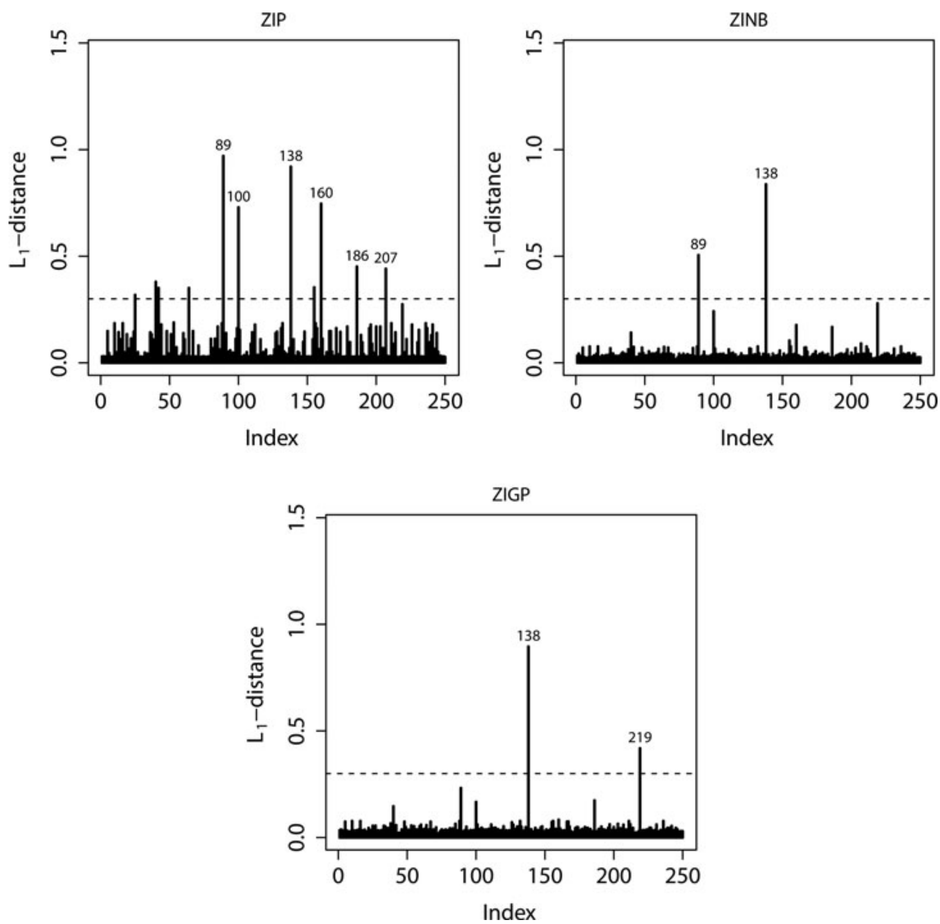

Figure 5. Index plots of ψ-divergence measures for the fish data.

Table 4. Posterior mean and relative change of the mean value with respect to the mean calculated omitting some cases for the fish data fitting the ZIGPS models.

Model	Drop	Parameter					
		ϕ	α_1	α_2	β_1	β_2	β_3
ZIP	{89}	—	10.51	18.11	1.00	21.33	31.59
		—	(0.34, 1.68)	(−0.72, −0.15)	(1.40 1.75)	(−1.03 -0.60)	(0.39, 0.78)
	{138}	—	15.85	32.87	13.03	27.12	27.67
		—	(0.61, 2.00)	(−1.00, −0.37)	(1.20, 1.57)	(−1.53, −1.11)	(0.88, 1.27)
	{219}	—	2.45	5.51	0.19	6.66	0.83
		—	(0.52, 1.86)	(−0.823, −0.23)	(1.43, 1.76)	(−1.32, −0.91)	(0.66, 1.01)
	{89, 138}	—	10.15	22.67	14.88	9.48	0.18
		—	(0.55, 1.94)	(−0.96, −0.31)	(1.18, 1.56)	(−1.34, −0.90)	(0.64, 1.05)
ZINB	{89}	8.42	0.13	2.02	5.98	9.69	14.17
		(1.67, 3.39)	(−0.45, 2.04)	(−1.96, −0.48)	(0.81, 1.79)	(−1.76, −0.99)	(0.19, 1.22)
	{138}	18.37	20.43	2.02	29.23	14.71	68.12
		(1.50, 3.03)	(−0.19, 2.12)	(−2.05, -0.67)	(0.53, 1.42)	(−2.11, −1.35)	(0.90, 1.95)
	{219}	1.38	5.58	2.82	6.27	9.16	4.01
		(1.90, 3.66)	(−0.40, 2.03)	(−2.06, −0.58)	(0.95, 2.01)	(−2.06, −1.23)	(0.27, 1.37)
	{89, 138}	26.52	20.38	0.33	31.99	4.78	46.38
		(1.30, 2.76)	(−0.17, 2.11)	(−1.92, −0.57)	(0.93, 0.48)	(−1.94, −1.20)	(0.73, 1.73)
ZIGP	{89}	4.66	10.47	4.17	7.54	8.20	18.52
		(0.50, 0.99)	(−0.42, 1.95)	(−1.76, −0.51)	(0.93, 2.19)	(−2.01, −1.02)	(0.10, 1.29)
	{138}	21.99	22.02	1.65	34.19	12.93	91.77
		(0.43, 0.80)	(−0.02, 2.02)	(−1.72, −0.56)	(0.62, 1.60)	(−2.39, −1.37)	(0.86, 2.33)
	{219}	1.42	3.61	0.26	13.77	14.69	17.80
		(0.54, 1.03)	(−0.40, 2.00)	(−1.77, −0.55)	(1.07, 2.91)	(−2.66, −1.275)	(0.05, 1.41)
	{89, 138}	22.79	19.34	1.85	37.98	6.86	78.32
		(0.41, 0.81)	(−0.02, 1.97)	(−1.71, −0.58)	(0.56, 1.49)	(−2.28, −1.30)	(0.80, 2.20)

the corresponding 95% HPD intervals for the parameters of ZIGPS regression model are displayed in Table 4. Note that there are little relative changes in posterior mean with exception in α_1 after dropping the observations {89} and {89,138}, but there are no changes of the inferences in the coefficients.

6. Final remarks

In this paper, we proposed a general class of zero-inflated model, the ZIGPS regression model, as an alternative means do model count data with an excess of zeros. This model contains the ZIP, ZINB, ZIGP, and other regression models. We use MCMC methods to obtain a Bayesian inference for the proposed model. The model can be tested for the best fitting in a straightforward way. Further, we propose a Bayesian residual and case influence diagnostic procedure based on the variational distance, J-distance, K–L divergence, and χ^2-square divergence, to study the sensitivity of the Bayesian estimates under perturbations in the model/data. Finally, we fitted our model to a real data set to show the significant potential of the methodology, and establish that the ZIGP regression model is the best fit for this data set.

References

Agarwal, D., Gelfand, A., Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environ. Ecol. Stat.* 9(4):341–355.

Arellano-Valle, R., Galea-Rojas, M., Zuazola, P.I. (2000). Bayesian sensitivity analysis in elliptical linear regression models. *J. Stat. Plann. Inference.* 86(1):175–199.

Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L., Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. Royal Stat. Soc. : Ser. A (Stat. Soc.)* 162(2):195–209.

Brooks, S.P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002). *J. Royal Stat. Soc. B.* 64:616–618.

Cancho, V., Ortega, E., Paula, G. (2010). On estimation and influence diagnostics for log-Birnbaum-Saunders student-t regression models: Full Bayesian analysis. *J. Stat. Plann. Inference.* 140(9):2486–2496.

Cancho, V., Dey, D., Lachos, V., Andrade, M. (2011). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Comput. Stat. Data Anal.* 55(1):588–602.

Carlin, B.P., Louis, T.A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*. (2nded.) Boca Raton: Chapman & Hall/CRC.

Carlin, B.P., Polson, N.G. (1991). An expected utility approach to influence diagnostics. *J. Am. Stat. Assoc.* 86(416):1013–1021.

Cho, H., Ibrahim, J.G., Sinha, D., Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics.* 65(1):116–124.

Cook, R.D. (1986). Assessment of local influence. *J. Royal Stat. Soc. Ser. B.* 48:133–169.

Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. BocaRaton, FL: Chapman & Hall/CRC.

Cordeiro, G., Andrade, M., De Castro, M. (2009). Power series generalized nonlinear models. *Comput. Stat. Data Anal.* 53(4):1155–1166.

Czado, C., Gneiting, T., Held, L. (2009). Predictive model assessment for count data. *Biometrics.* 65(4):1254–1261.

Dey, D., Birmiwal, L. (1994). Robust Bayesian analysis using divergence measures. *Stat. Probab. Lett.* 20(4):287–294.

Dey, D., Chen, M., Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics.* 53:1239–1252.

Geisser, S., Eddy, W. (1979). A predictive approach to model selection. *J. Am. Stat. Assoc.* 74(365):153–160.

Gelfand, A., Dey, D., Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods. In Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., eds. *Bayesian Statistics* 4 (pp. 147–167) Oxford: University Press.

Gelfand, A.E., Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85(410):398–409.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2004). *Bayesian Data Analysis*. Boca Raton, FL, USA: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., eds. *Bayesian Statistics 4* (pp. 169–188). Oxford: Oxford University Press.

Hall, D. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics.* 56(4):1030–1039.

Johnson, N.L., Kemp, A.W., Kotz, S. (2005). *Univariate Discrete Distributions*. 3rded. Hoboken: Wiley.

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. *Can. J. Stat. /La Revue Canadienne de Statistique*. 15(3): 209–225.

Mwalili, S., Lesaffre, E., Declerck, D. (2008). The zero-inflated negative binomial regression model with correction for misclassification: An example in caries research. *Stat. Methods Med. Res.* 17(2):123.

Peng, F., Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *Can. J. Stat.* 23(2):199–213.

Pettit, L. (1986). Diagnostics in Bayesian model choice. *Statistician.* 35:183–190.

Ridout, M., Hinde, J., Demétrio, C. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics.* 57(1): 219–223.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Royal Stat. Soc. B.* 64:583–639.

Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *J. Royal Stat. Soc. Ser. B: Methodol.* 58:739–750.

Weiss, R.E., Cook, R.D. (1992). A graphical case statistic for assessing posterior influence. *Biometrika.* 79(1):51–55.

Zhou, X., Tu, W. (2000). Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics.* 56(4):1118–1125.