On the estimation and influence diagnostics for the zero–inflated Conway–Maxwell–Poisson regression model: A full Bayesian analysis

Gladys D. Cacsire Barriga[1], and Francisco Louzada[2]
Faculty of Engineering at Bauru, São Paulo State University, Brazil
Department of Applied Maths and Statistics, São Paulo State University, Brazil

Abstract

In this paper we develop a Bayesian analysis for the zero-inflated regression models based on the COM-Poisson distribution. Our approach is based on Markov chain Monte Carlo methods. We discuss model selection, as well as, develop case deletion influence diagnostics for the joint posterior distribution based on the ψ-divergence, which has several divergence measures as particular cases, such as the Kullback-Leibler (K-L), J-distance, L_1 norm and χ^2-square divergence measures. The performance of our approach is illustrated in and artificial dataset as well as in a real dataset on an apple cultivar experiment.

Keywords: Bayesian inference, COM–Poisson distribution, Kullback-Leibler, zero-inflated models.

1 Introduction

Count data with excess of zeros (or zero-inflated) are commonly encountered in many disciplines, including medicine (Bohning *et al.*, 1999), public health (Zhou & 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 expected. This is of interest since the incidence of zero counts frequently have special status from the practical point of view. For instance, Ridout *et al.* (2001) point out 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 zero-inflation (ZI) model is then to mix a distribution degenerate at zero with distributions (baseline models) as Poisson, Negative Binomial, Binomial, among others.

In most cases the zero-inflated Poisson (ZIP) model, described in Lambert (1992) seminal work, has been studied and considered for this type of problems. A problem here is that the data suggest additional overdispersion. In such case, we may consider the zero inflated negative binomial (ZINB) model, mixing a distribution degenerate at zero with a baseline negative binomial distribution, over the ZIP model, and so on. Without confusion, overdispersion can be the result of excess zeros or some other causes. In any case, the result is excess of variability. In some cases, the ZIP model may not be appropriate for such data, since the baseline (Poisson) model does not accommodate the remaining over dispersion not accounted for through zero-inflation and it is well known that negative binomial (NB) models are more flexible than their simpler Poisson counterparts in accommodating overdispersion (Lawless, 1987). The ZINB model have been discussed in Ridout *et al.* (2001), where a a score test is provided for testing ZIP regression models against ZINB alternatives. While Mwalili *et al.* (2008) illustrated how the ZINB regression model can be corrected for misclassification.

The Conway–Maxwell–Poisson distribution, usually known as COM-Poisson, was first introduced by Conway & Maxwell (1962) for modeling queues and service rates. The COM-Poisson distribution has very recently been re-introduced for modeling count data that are characterized by either over– or under–dispersion (Shmueli *et al.*, 2005; Lord *et al.*, 2008; Guikema & Goffelt, 2008). This distribution has also been evaluated in the context of a Generalized linear models (GLM) (Guikema & Goffelt, 2008). However, to the best of our knowledge, nobody has so far examined how the COM-Poisson could be used for modeling for count data with excess zeros.

In this paper we propose the zero-inflated COM-Poisson distribution, hereafter the ZICOM-Poisson distribution. Moreover, we develop a Bayesian analysis for the ZICOM-Poisson regression models based on Markov chain Monte Carlo methods.

After fitting a model, it is important to check its assumptions and conduct sensitivity studies in order to detect possible influential or extreme observations, which may cause distortions on the 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. However, to the best of our knowledge, there are no studies on Bayesian inference for the ZICOM–Poisson inflated zeros (ZICOM-Poisson) models or on influence diagnostics related this research topic. Thus, we believe that the development of statistical influence diagnostic Bayesian tools for the ZICOM-Poisson models is a significant contribution to this field. The objective here is to develop diagnostic measures from a Bayesian perspective based on the ψ-divergence (Peng & Dey, 1995; Weiss, 1996) between the posterior distributions of the parameters of the proposed model.

The paper is organized as follows. In Section 2 we give a brief sketch of ZICOM-Poisson regression models and present some of its properties. In Section 3, we carry out Bayesian inference for ZICOM-Poisson models. Some measures of model selection are discussed in Section 3. In the Section 4 we introduce Bayesian case diagnostics based on the ψ-divergence. The methodology is illustrated in Section 5, in which ZICOM-Poisson and ZIP models are compared according to their Bayesian inference and case diagnostics. Finally, some concluding remarks are presented in Section 6.

2 The ZICOM-Poisson distribution

The COM-Poisson distribution is a generalization of the Poisson distribution. Its probability mass function (pmf) can be given by

$$Pr(Y = y) = \frac{1}{Z(\lambda, \phi)} \frac{\lambda^y}{(y!)^\phi}, \;\; y = 0, 1, \ldots, \tag{1}$$

where $Z(\lambda, \phi) = \sum_{s=0}^{\infty} \lambda^s/(s!)^\phi$, $\lambda > 0$, is a centering parameter that is approximately the mean of the observations in many cases and $\phi > 0$ is defined as the shape parameter. The COM-

Poisson can model both under-dispersed ($\phi > 1$) and over-dispersed ($\phi < 1$) data, and several common pmfs are its special cases in its the original formulation. Specifically, setting $\phi = 0$ yields the geometric distribution; $\lambda < 1$ and $\phi \to \infty$ yields the Bernoulli distribution in the limit; and $\phi = 1$ yields the Poisson distribution. Such flexibility greatly expands the types of problems for which the COM-Poisson distribution can be used to model count data. The mean and variance of the COM-Poisson distribution are, respectively, given by

$$E(Y) = \frac{\partial \log Z(\lambda, \phi)}{\partial \log \lambda}, \text{ and } \text{Var}(Y) = \frac{\partial^2 \log Z(\lambda, \phi)}{\partial \log^2 \lambda}. \tag{2}$$

The COM–Poisson distribution does not have closed-form expressions for its moments in terms of the parameters λ and ϕ. Shmueli *et al.* (2005) using a asymptotic expression for $Z(\lambda, \phi)$ in Equation (1) derive an approximation for mean and variance of the COM–Poisson distribution given by $E(Y) \approx, \lambda^{1/\phi} + \frac{1}{2\phi} - \frac{1}{2}$ and $\text{Var}(Y) \approx \frac{1}{\phi}\lambda^{1/\phi}$. Shmueli *et al.* (2005) argued that care should be taken in using these approximations. In particular, they may not be accurate for $\phi > 1$ or $\phi < 10$.

The COM-Poisson has limitations in its usefulness as a basis for a GLM, as documented in Guikema & Goffelt (2008). In particular, neither λ nor ϕ provide a clear centering parameter. While λ is approximately the mean when ϕ is close to one, it differs substantially from the mean for small λ. Given that, λ would be expected to be small for over-dispersed data, this would make a COM–Poisson model based on the original COM formulation difficult to interpret and use for over-dispersed data. To circumvent this problem, Guikema & Goffelt (2008) proposed a reparameterization ($\mu = \lambda^{1/\phi}$) to provide a clear centering parameter. This new formulation of the COM–Poisson is given by

$$Pr(Y = y) = \frac{1}{S(\mu, \phi)} \left(\frac{\mu^y}{y!}\right)^{\phi}, \quad y = 0, 1, \ldots, \tag{3}$$

where $S(\lambda, \phi) = \sum_{s=0}^{\infty} \left(\frac{\mu^s}{s!}\right)^{\phi}$. The mean and variance of Y are given in terms of the new formulation as $E(Y) = \frac{1}{\phi}\frac{\partial \log S(\lambda,\phi)}{\partial \log \mu}$ and $E(Y) = \frac{1}{\phi}\frac{\partial \log S(\lambda,\phi)}{\partial \log^2 \mu}$ with asymptotic approximations $E(Y) \approx \mu + \frac{1}{2\phi} - \frac{1}{2}$ and $\text{Var}(Y) \approx \frac{\mu}{\phi}$, especially accurate once $\phi > 10$. With this new parameterization, the integral part of μ is now the mode leaving μ as a reasonable approximation of the mean for all but low values of μ. Also, allows ϕ to keep its role as a shape parameter. That is, if $\phi < 1$, the variance is greater than the mean while $\phi > 1$ leads to underdispersion.

The ZICOM-Poisson distribution is the result of mixing a COM–Poisson distribution and a degenerate distribution at zero. Then, the random variable $Y_i, (i = 1, \ldots, n)$ is ZICOM-Poisson distributed if its pmf is given by,

$$Pr(Y_i = y_i) = \begin{cases} p_i + \dfrac{(1-p_i)}{S(\mu_i, \phi_i)}, & y_i = 0, \\ \\ (1-p_i)\dfrac{1}{S(\mu_i, \phi_i)}\left(\dfrac{\mu_i^{y_i}}{y_i!}\right)^{\phi_i}, & y_i = 1, 2, \ldots, \end{cases} \tag{4}$$

where $0 \leq p_i \leq 1$, $\mu_i \geq 0$ and $\phi_i > 0$. The mean and the variance of the model defined in (4) can be approximated by $E(Y_i) \approx (1-p_i)(\mu_i - \frac{(1-\phi_i)}{2\phi_i})$, $Var(Y_i) \approx (1-p_i)(\frac{\mu_i}{\phi_i} + p_i(\mu_i - \frac{(1-\phi_i)}{2\phi_i})^2)$.

When $p_i = 0$, the random variable Y_i has a COM–Poisson distribution with parameters λ_i and ϕ_i. distribution.

In many practical applications it is common to assume that the parameters p_i, λ_i and ϕ_i depend on vectors of explanatory variables $\boldsymbol{x}_i$, $\boldsymbol{z}_i$ and $\boldsymbol{w}_i$, respectively. Following Lord *et al.* (2008), we propose relate the p_i to covariates $\boldsymbol{x}_i$ by the logistic link and λ_i and ϕ to covariates $\boldsymbol{z}_i$ and $\boldsymbol{w}_i$, by the logarithmic link respectively, i.e.,

$$\log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i^\top \boldsymbol{\alpha}, \;\; \log(\mu_i) = \boldsymbol{z}_i^\top \boldsymbol{\beta}, \text{ and } \log(\phi_i) = \boldsymbol{w}_i^\top \boldsymbol{\gamma}, \tag{5}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p_1})^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p_2})^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{p_3})^\top$ are unknown parameters.

3 Inference

For inference we adopt a full Bayesian approach. Let $\boldsymbol{\vartheta}$ denote the parameter vector of the distribution of the time to event Y in (4). From n independent observations, $(y_1, \delta_1, \boldsymbol{x}_1, \boldsymbol{z}_1, \boldsymbol{w}_1)$, $\ldots, (y_n, \delta_n, \boldsymbol{x}_n, \boldsymbol{z}_n, \boldsymbol{w}_n)$ the corresponding likelihood function is given by

$$L(\boldsymbol{\vartheta}; \boldsymbol{\mathcal{D}}) = \prod_{i=1}^{n} \left[p_i + \frac{(1-p_i)}{S(\mu_i, \phi_i)}\right]^{\delta_i} \left[\frac{1}{S(\mu_i, \phi_i)} \left(\frac{\mu_i^{y_i}}{y_i!}\right)^{\phi_i}\right]^{1-\delta_i}, \tag{6}$$

where $\delta_i = 1$ if $y_i = 0$ and $\delta_i = 0$ otherwise. The maximum likelihood estimation of the parameter vector $\boldsymbol{\vartheta}$ is carried out by direct numerical maximization of the log-likelihood function $\ell(\boldsymbol{\vartheta}; \boldsymbol{\mathcal{D}}) = \log(L(\boldsymbol{\vartheta}; \boldsymbol{\mathcal{D}}))$, which is accomplished by using extant software (R Development Core Team, 2010). A possible disadvantage of direct maximization of log-likelihood functions is that it may not converge unless good starting values are used.

Now, some inferential tools are investigated under a Bayesian viewpoint. The normal distribution is denoted by $N(\mu, \sigma^2)$. In this context we assume that $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are *a priori* independent, that is,

$$\pi(\boldsymbol{\vartheta}) = \prod_{i=1}^{p_1} \pi(\alpha_i) \prod_{i=1}^{p_2} \pi(\beta_i) \prod_{i=1}^{p_3} \pi(\gamma_i), \tag{7}$$

where $\alpha_j \sim N(0, \sigma^2_{\beta_j})$, $j = 1, \ldots, p_1$, $\beta_k \sim N(0, \sigma^2_{\beta_k})$, $k = 0, 1, \ldots, p_2$ and $\gamma_l \sim N(0, \sigma^2_{\gamma_l})$, $l = 0, 1, \ldots, p_3$. Here all the hyper-parameters are specified in order to express non-informative priors.

Combining the likelihood function (6) and the prior distribution in (7), the joint posterior distribution for $\boldsymbol{\vartheta}$ is obtain as $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}) \propto L(\boldsymbol{\vartheta}; \boldsymbol{\mathcal{D}})\pi(\boldsymbol{\vartheta})$, which is analytically intractable. So, we based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. In particular, the Gibbs sampler and Metropolis-Hastings algorithms (see Gamerman & Lopes, 2006) has proved to be a powerful alternative.

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 $Q(\boldsymbol{\xi}', \boldsymbol{\xi}_j) = N_{p_1+p_2+p_3}\left(\boldsymbol{\xi}_j, \tilde{\Sigma}\right)$, where $\tilde{\Sigma}$ is covariance matrix of $\boldsymbol{\xi}$ is same in any stage;

(3) update $\boldsymbol{\xi}_{(j)}$ to $\boldsymbol{\xi}_{(j+1)} = \boldsymbol{\xi}'$ with probability $q_j = \min\{1, \pi(\boldsymbol{\xi}'|\boldsymbol{\mathcal{D}})/\pi(\boldsymbol{\xi}_{(j)}|\boldsymbol{\mathcal{D}})\}$, or keep $\boldsymbol{\theta}_{(j)}$ with probability $1 - q_j$;

(4) repeat steps (2) and (3) by increasing the stage indicator until the process reaches a stationary distribution.

The computational program is available from the authors upon request.

4 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. One of the most used in applied works 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 & Eddy (1979). Let $\boldsymbol{\mathcal{D}}$ the full data and $\boldsymbol{\mathcal{D}}^{(-i)}$ denote the data with the i-th observation deleted. In our model, we have from Section 3 that $g(y_i|\boldsymbol{\vartheta}) = Pr(Y = y_i)$. We denote the posterior density of $\boldsymbol{\vartheta}$ given $\boldsymbol{\mathcal{D}}^{(-i)}$ by $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}^{(-i)})$, $i = 1, \ldots, n$. For the i-th observation, CPO_i can be written as

$$CPO_i = \int_{\vartheta \in \Theta} g(y_i|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}^{(-i)})d\boldsymbol{\vartheta} = \left\{ \int_{\vartheta} \frac{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})}{g(y_i|\boldsymbol{\vartheta})} d\boldsymbol{\vartheta} \right\}^{-1}. \qquad (8)$$

The CPO_i can be interpreted as the height of the marginal density of the time to event at y_i. Thus, large values of CPO_i imply a better fit of the model. For the proposed model a closed form of the CPO_i is not available. However, a Monte Carlo estimate of CPO_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 $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})$ after the burn-in. A Monte Carlo approximation of CPO_i (Ibrahim *et al.*, 2001) is given by

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

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 like, 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 & Louis (2001) can 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 $d(\boldsymbol{\vartheta}) = -2 \sum_{i=1}^{n} \log\left[g(y_i|\boldsymbol{\vartheta})\right]$. The DIC criterion 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

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

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.

5 Bayesian case influence diagnostics

Sinceregression models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. 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 & Weisberg, 1982) in which the effects are studied of 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.

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,

$$D_\psi(P, P_{(-i)}) = \int_{\vartheta\in\Theta} \psi\left(\frac{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})}\right)\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})\, d\boldsymbol{\vartheta}. \tag{9}$$

where ψ is a convex function with $\psi(1) = 0$. Several choices of ψ are given in Dey & Birmiwal (1994). For example, $\psi(z) = -\log(z)$ defines Kullback-Leibler (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 (8) and the ψ-divergence measure is described in next proposition.

Proposition 1 The ψ-divergence measure can be written as

$$D_\psi(P, P_{(-i)}) = E_{\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}}\left[\psi\left(\frac{CPO_i}{g(y_i|\boldsymbol{\vartheta})}\right)\right], \tag{10}$$

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

Proof: From Bayes's theorem the posterior distribution of $\boldsymbol{\vartheta}$ is given by

$$\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}) = \frac{\pi(\boldsymbol{\vartheta})\prod_{j\in\boldsymbol{\mathcal{D}}} g(y_j|\boldsymbol{\vartheta})}{\int_{\boldsymbol{\vartheta}\in\Theta}\pi(\boldsymbol{\vartheta})\prod_{j\in\boldsymbol{\mathcal{D}}} g(y_j|\boldsymbol{\vartheta})d\boldsymbol{\vartheta}}$$

where $\pi(\boldsymbol{\vartheta})$ and $\prod_{i=1}^n g(y_i|\boldsymbol{\vartheta})$ represents the prior distribution and the likelihood function $\boldsymbol{\vartheta}$ respectively. The ratio of the posterior distributions is given by

$$\begin{aligned}\frac{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})} &= \frac{\pi(\boldsymbol{\vartheta})\prod_{j\in D^{(i)}} g(y_i|\boldsymbol{\vartheta})}{\int_{\boldsymbol{\vartheta}\in\Theta}\pi(\boldsymbol{\vartheta})\prod_{j\in D^{(i)}} g(y_j|\boldsymbol{\vartheta})d\boldsymbol{\vartheta}} \times \frac{\int_{\boldsymbol{\vartheta}\in\Theta}\pi(\boldsymbol{\vartheta})\prod_{j\in D} g(y_j|\boldsymbol{\vartheta})d\boldsymbol{\vartheta}}{\pi(\boldsymbol{\vartheta})\prod_{j\in D} g(y_j|\boldsymbol{\vartheta})} \\ &= \frac{1}{g(y_i|\boldsymbol{\vartheta})} \times \frac{\int_{\boldsymbol{\vartheta}\in\Theta}\pi(\boldsymbol{\vartheta})\prod_{j\in D} g(y_j|\boldsymbol{\vartheta})d\boldsymbol{\vartheta}}{\int_{\boldsymbol{\vartheta}\in\Theta}\frac{1}{g(y_i|\boldsymbol{\vartheta})}\pi(\boldsymbol{\vartheta})\prod_{j\in D} g(y_j|\boldsymbol{\vartheta})d\boldsymbol{\vartheta}} \\ &= \frac{\left(\int_{\boldsymbol{\vartheta}\in\Theta}\frac{1}{g(y_i|\boldsymbol{\vartheta})}\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})d\boldsymbol{\vartheta}\right)^{-1}}{g(y_i|\boldsymbol{\vartheta})} = \frac{CPO_i}{g(y_i|\boldsymbol{\vartheta})}. \quad \square\end{aligned}$$

From proposition 1, the K-L divergence can be expressed by

$$\begin{aligned} D_{\text{K-L}}(P, P_{(-i)}) &= -E_{\boldsymbol{\vartheta}|\mathcal{D}}\left\{\log(CPO_i)\right\} + E_{\boldsymbol{\vartheta}|\mathcal{D}}\left\{\log\left[g(y_i|\boldsymbol{\vartheta})\right]\right\} \\ &= -\log(CPO_i) + E_{\boldsymbol{\vartheta}|\mathcal{D}}\left\{\log\left[g(y_i|\boldsymbol{\vartheta})\right]\right\}. \end{aligned} \tag{11}$$

From (10) we can be computed $D_\psi(P, P_{(-i)})$ 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

$$\widehat{D_\psi}(P, P_{(-i)}) = \frac{1}{Q}\sum_{q=1}^{Q} \psi\left(\frac{\widehat{CPO_i}}{g(y_i|\boldsymbol{\vartheta}^{(q)})}\right). \tag{12}$$

From (12) a Monte Carlo estimate of K-L divergence $D_{K-L}(P, P_{(-i)})$ is given by

$$\widehat{D_{\text{K-L}}}(P, P_{(-i)}) = -\log(\widehat{CPO_i}) + \frac{1}{Q}\sum_{q=1}^{Q} \log\left[g(y_i|\boldsymbol{\vartheta}^{(q)})\right]. \tag{13}$$

The $D_\psi(P, P_{(-i)})$ can be interpretes as the ψ-divergence of the effect of deleting of i-th case from the full data on the joint posterior distribution of $\boldsymbol{\vartheta}$. As pointed by Peng & 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 observations is influential or not. In this context, we will use the proposal given by Peng & Dey (1995) anda Weiss (1996) by considering as follows. Consider a biased coin, which has success probability p. Then the ψ-divergence between the biased and an unbiased coin is

$$D_\psi(f_0, f_1) = \int \psi\left(\frac{f_0(x)}{f_1(x)}\right) f_1(x)dx, \tag{14}$$

where $f_0(x) = p^x(1-p)^{1-x}$ and $f_1(x) = 0.5$, $x = 0, 1$. Now if $D_\psi(f_0, f_1) = d_\psi(p)$ then it can be easily checked that d_ψ, satisfies the following equation

$$d_\psi(p) = \frac{\psi(2p) + \psi(2(1-p))}{2} \tag{15}$$

It is not dicult to see for the divergence measures considered, that d_ψ, increases as p moves away from 0.5. In addition, $d_\psi(p)$, is symmetric about $p = 0.5$ and d_ψ, achieves its minimum at $p = 0.5$. In this point, $d_\psi(0.5) = 0$, and $f_0 = f_1$. Therefore, if we consider $p > 0.80$ (or $p \leq 0.20$) as a strong bias in a coin, then, since $d_{L_1}(0.80) = 0.30$ or $d_{\chi^2}(0.80) = 0.36$. This equation implies that ith case is considered influential when $d_{L_1}(0.80) > 0.30$ or $d_{\chi^2} > 0.36$. Thus, if we use the Kullback-Leibler divergence, we can consider an influential observation when $d_{\text{K-L}} > 0.22$. Similarly, using the J-distance, an observation which $d_{\text{J}} > 0.42$ can be considered as influential.

6 Application

In this section, we illustrate our methodology with artificial and real datasets.

6.1 Simulated data

To examine the performance of the proposed diagnostics measures, we considered simulated datasets with one or more of the generated cases perturbed.

In order to do this, we consider a sample of size 100 generated by the ZICOM–Poisson regression model with parameter $p_i = \exp(\alpha_0 + \alpha_1 x_i)/(1 + \exp(\alpha_0 + \alpha_1 x_i))$, $\mu_i = \exp(\beta_0 + \beta_1 x_i)$, $i = 1, \ldots, n$ and $\phi_i = \exp(\gamma_0 + \gamma_1 x_i)$. In our simulations we have one binary covariate x with values drawn from Bernoulli distribution with parameter 0.5. We took $\alpha_0 = -0.3$, $\alpha_1 = 0.7$, $\beta_0 = 2.0$ and $\beta_1 = 1.0$, $\gamma_0 = -1.0$ and $\gamma_1 = -1.0$. The series $S(\cdot, \cdot)$ in (4) may be computed by truncating the numerical series. In our application, after a preliminary sensitivity study on the parameter estimates, these series were truncated at the 101*st* term.

We selected cases 11 and 59 for perturbation. To create influential observation in the dataset, we choose one or two of these selected cases and perturbed the response variable as follows $\widetilde{y_i} = y_i + 3S_y$, $i = 11$; and 59, where S_y is the standard deviations of the y_i's. 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 $\gamma_l \sim N(0, 10^2)$, $l = 0, 1, 2$. After burn-in, 40,000 MCMC posterior. We monitored convergence of the Metropolis-Hasting algorithm using the method proposed by Geweke (1992), as well as trace plots. We used every tenth sample from the 40,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results. To evaluate the robustness of the model with regard to the choice of the hyper-parameters of the prior distributions, a small sensitivity study was carried out with larger standard deviations for the prior distributions. The posterior summaries of the parameters do not display much difference and do not alter the results presented in Table 1.

We fit the ZICOM–Poisson model. Table 1 shows that the posterior inferences to parameters are sensitive to the perturbation of the selected case(s). In the Table 1, dataset (a) denotes the original simulated dataset with no perturbation and datasets (b)-(d) denote datasets with perturbed cases.

Table 1: Mean and Standard Deviation (SD) of the parameters of ZICOM–Poisson model for each datasets.

Dataset names	Perturbed case	α_0	α_1	β_0	β_1	γ_1	γ_2
a	none	-0.260	0.684	1.978	1.046	-1.066	-0.868
		(0.271)	(0.382)	(0.184)	(0.225)	(0.360)	(0.503)
b	11	-0.431	0.7692	1.500	0.874	-1.342	-1.398
		(0.302)	(0.435)	(0.320)	(0.503)	(0.503)	(0.776)
c	59	-0.381	0.729	1.446	0.895	-1.216	-1.189
		(0.296)	(0.401)	(0.221)	(0.431)	(0.503)	(0.765)
d	11,59	-0.741	0.642	0.932	0.221	-2.050	-2.345
		(1.307)	(0.895)	(2.774)	(1.416)	(2.426)	(1.379)

Now we consider the sample from the posterior distributions of the parameters of the ZICOM–Poisson model to calculate the ψ-divergence measures in (9) described in Section 5. The results in Table 2 show, before perturbation (dataset (a)), that all the selected cases are not influential according to all ψ-divergence measures. However, after perturbation (dataset, (b)-(d)), the measures increase indicating that the perturbed cases are influential.

Table 2: ψ-divergence measures for the simulated data fitting the ZICOM–Poisson model.

Dataset names	Case number	$d_{\text{K-L}}$	d_{J}	d_{L_1}	d_{χ^2}
a	11	0.160	0.342	0.224	0.229
	59	0.053	0.110	0.131	0.127
b	11	1.646	3.847	0.690	41.334
c	59	0.861	1.951	0.516	8.451
d	11	1.630	3.545	0.585	53.049
	9	0.881	1.469	0.350	3.740

In the Figures 1 and 2, we have depicted the four ψ-divergence measures for the cases (a) and (b), respectively. Clearly we can see that all measures performed well to identifying influential case(s), providing larger ψ-divergence measures when compared to the other cases.

In Table 3 we report the Monte Carlo estimates of the DIC, $EAIC$, $EBIC$ and $LPML$ for each perturbed version of the original data set. We can see that, the ZICOM–Poisson Model before perturbation under dataset (a) stand out as the best ones.

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

Dataset names	LPML	DIC	EAIC	EBIC
a	-215.36	429.13	435.89	451.52
b	-236.83	470.66	477.28	492.91
c	-229.89	461.11	465.98	481.61
d	-233.77	441.69	447.09	462.72

6.2 Apple cultivar data

In this section we considerthe data set extracted from Ridout *et al.* (2001) and used by Garay *et al.* (2011) from a frequentist point of view. These data is referring to the number of roots produced by 270 micropropagated shoots of the columnar apple cultivar Trajan. The shoots had been produced under an 8- or 16-hour photoperiod in culture systems that utilized one of four different concentrations of the cytokinin BAP in the culture medium. There were 30 or 40 shoots of each of these eight treatment combinations. Of the 140 shoots produced under the 8-hour photoperiod, only 2 failed to produce roots, but 62 of the 130 shoots produced under the 16-hour photoperiods failed to root. The focus of the study is on the influence of BAP concentration on two photoperiod on rooting of apple cultivar. The sample size corresponds to $n = 270$ and the percentage of zeros observed was 23.7%. The following data were collected from each micropropagated: y_i : count of roots; x_{i1}: concentrations of the cytokinin BAP and x_{i2}: photoperiod (0=8-hour, 1=16-hour) , $i = 1, \ldots, 270$.

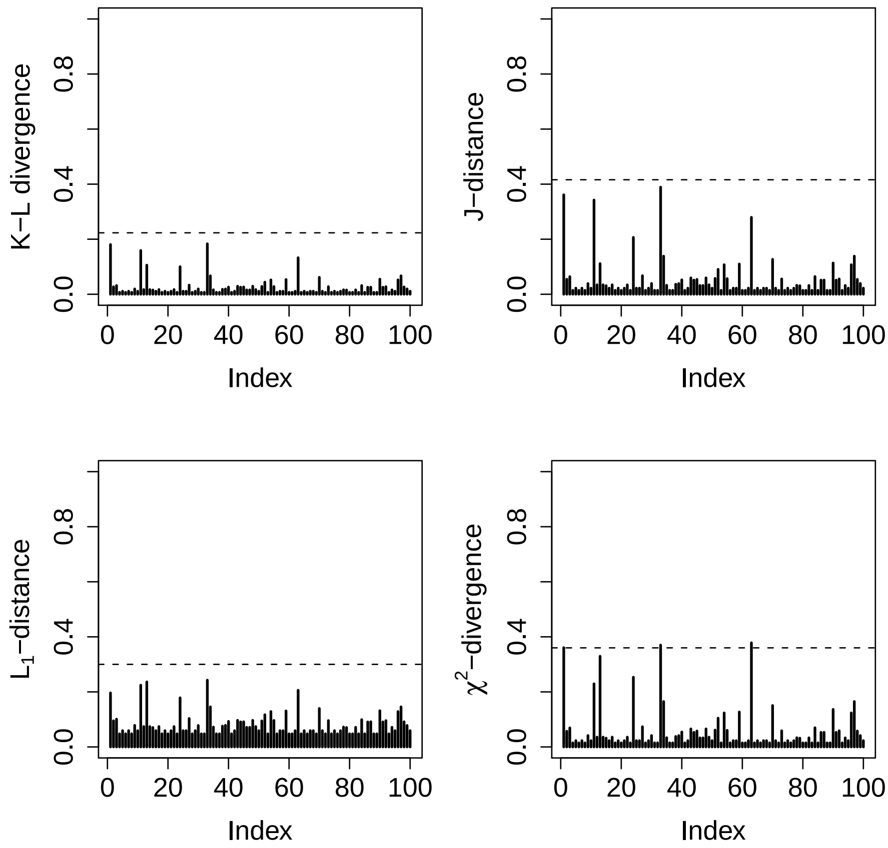

Figure 1: ψ-divergence measures from dataset (a) for the simulated data.

We fit a ZICOM–Poisson regression model described in Section 2 with

$$p_i = \frac{\exp\left(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{2i}\right)}{1 + \exp\left(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{2i}\right)}, \ \log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{2i}, \text{ and } \log(\phi_i) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{2i}.$$

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 $\gamma_l \sim N(0, 10^2)$, $l = 0, 1, 2$. The MCMC computations were done similar to those in the Section 6.1 and further to monitor the convergence of the Gibbs samples we used the methods recommended by Cowles & Carlin (1996). 40000 MCMC posterior samples were used in this analysis after burn-in. MCMC computations, we used every twenth sample from the 40,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results.

The posterior means, medians, standard deviations and 95% highest posterior density (HPD) intervals are in Table 4. The covariates have a significant effect on the dispersion parameter. But only the covariate x_2 (photoperiod) is significant in the mean of Y as well as the proportion of zeros.

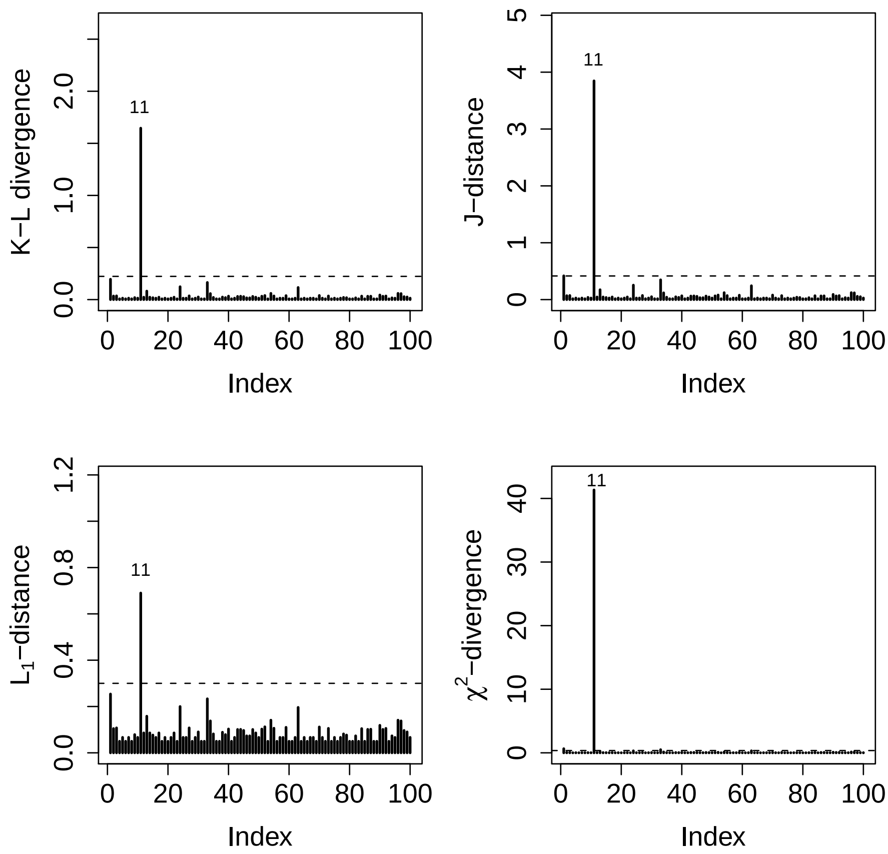

Figure 2: ψ-divergence measures from dataset (b) for the simulated data.

To compare the ZICOM–Poisson model with ZIP model, we obtained the values DIC, $EAIC$, $EBIC$ and $LPML$ criteria. These information criteria furnish the values given in Table 5. According to the all criteria, the ZICOM-Poisson model, stands out as the best one. Taking into account the criteria in Table 5,we select the ZICOM–Poisson regression model as our working model.

Considered the sample to the posterior distributions of the parameters of ZICOM–Poisson regression models, the ψ-divergence measures in (9) described in Section 5 were computed. The Figure 3 shows the index plot of the four ψ-divergence measures, where we observed that case 230 is possible influential observation in the posteriori distribution. In order to reveal the impact of this observation on the parameter estimates, we refitted the model under this situation.

The relative changes (in percentage) of 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 $j = 1, ..., 9$, after the observations $I = \{230\}$ has been removed.

Relative changes in posterior means and the corresponding 95% HPD intervals in paren-

Table 4: Posterior summaries of the parameters for the ZICOM–Poisson regression model.

				HPD Interval (95%)	
Parameter	Mean	Median	Standard deviation	LI	LS
α_0	-3.5679	-3.5474	0.5221	-4.6630	-2.6085
α_1	-0.0018	-0.0026	0.0292	-0.0591	0.0556
α_2	3.3037	3.2709	0.5039	2.3928	4.3683
β_0	1.8485	1.8512	0.07908	1.6848	1.9924
β_1	0.0108	0.0107	0.0060	-0.0009	0.0227
β_2	-0.6526	-0.6166	0.2083	-1.1568	-0.3302
γ_0	-0.7958	-0.7985	0.2082	-1.2102	-0.3940
γ_1	0.0592	0.0592	0.01736	0.0245	0.0929
γ_2	-0.8960	-0.8707	0.3317	-1.6011	-0.3321

Table 5: Bayesian criteria for the fitted models.

	Criterion			
Activation	LPLM	DIC	EAIC	EBIC
ZICOM–Poisson	-620.50	1241.59	1251.24	1283.63
ZIP	-639.01	1273.41	1280.73	1302.32

theses for the parameters of ZICOM–Poisson regression model are: $0.30[-4.68; -2.55]$ for α_0, $212.34[-0.06; 0.05]$ for α_1, $0.14[2.35; 4.38]$ for α_2, $0.20[1.69\ ; 2.00]$ for β_0, $10.76[-0.002; 0.0214]$ for β_1, $0.61[-1.24; -0.34]$ for β_2 $1.68[-1.25; -0.40]$ for γ_0, $8.49[0.03; 0.10]$, for γ_1 and $3.67[-1.67; -0.36]$ for γ_2. We notice that there are little relative changes in posterior mean with exception in α_1 after dropping. the observation 230. But the inferences in the coefficients there was no changes. Thus,the final selected model in our analysis is given by

$$p_i = \frac{\exp\left(\alpha_0 + \alpha_2 x_{2i}\right)}{1 + \exp\left(\alpha_0 + \alpha_2 x_{2i}\right)}, \ \log(\mu_i) = \beta_0 + \beta_2 x_{2i}, \ \text{and} \ \log(\phi_i) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{2i}, \ i = 1, \ldots, 270.$$

For final model the posterior means(standard deviations) and 95% HPD intervals for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are: -3.559 (0.469) and [-4.585, -2.732] for α_0,3.303 (0.504) and [2.383 , 4.332] for α_2, 1.996(0.040) and [1.882,2.041] for β_0, -0.615(0.229) and [-1.188, -0.310] for β_2, -0.686(0.194) and [-1.082, -0.320] for γ_0, 0.048(0.017) and [0.013 , 0.080] for γ_1 and -0.822 (0.337) and [-1.529 , -0.258] for γ_2. The values of (DIC, LPML) are (1240.2, -620.20), respectively. Comparing with the entries in Table 5, we realize that the final model provides a similar fit to the data. Table 6 presents the fitted frequency distribution obtained from the ZICOM–Poisson and ZIP models. Its is observed that the fit by ZICOM–Poisson is better than that of ZIP in terms of χ^2. This can be easily explained as the ZICOM–Poisson deals with more paramaters.

We turn our attention to the role of the covariates on the parameters (p, μ and ϕ) of ZICOM–Poisson regression model. Table 7 shows the posterior summaries of the p, μ and ϕ stratified by photoperiod levels with concentrations of the cytokinin BAP equal to 8.8 for ϕ, which correspond

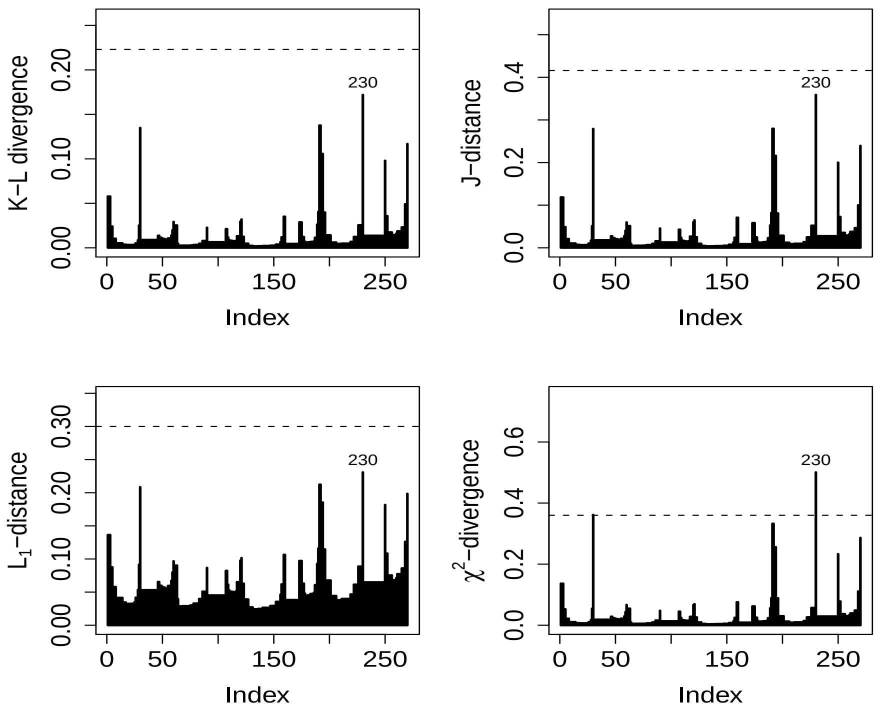

Figure 3: Index plots of ψ-divergence measures for the ZICOM–Poisson regression model.

to the median, under the ZICOM–Poisson model.

7 Final remarks

In this paper we proposed the ZICOM–Poisson model as a alternative model for modeling count data with a excess of zeros. We use the Markov Chain Monte Carlo methods to obtain a Bayesian inference approach for the proposed model. The model can be tested for the best fitting in a straightforwardly way. Moreover, we propose a case influence diagnostic Bayesian procedure based on the variational distance, J-distance, Kullback-Leibler divergente and χ^2-square divergence in order to study the sensitivity of the Bayesian estimates under perturbations in the model/data. Finally, we fitted our model to artificial and real datasets to show the potentiality of the methodology, where we realized that the ZICOM–Poisson model delivers the best fit.

References

Agarwal, D., Gelfand, A. & Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, **9**(4), 341–355.

Table 6: Fitted frequency distribution of the real datasets

Count	Observed	ZICOM–Poisson	ZIP
0	64	63.6	65.2
1	10	8.1	2.3
2	13	14.8	7.3
3	15	21.3	15.4
4	21	25.9	24.4
5	18	27.6	31.0
6	24	26.4	32.7
7	21	23.1	29.5
8	23	18.6	23.4
9	21	14.1	16.4
10	17	10.0	10.4
11	12	6.7	6.0
12	5	4.2	3.2
13	2	2.6	1.5
14	3	1.5	0.7
≥ 15	1	1.5	0.5
χ^2		18.4	33.9
p-Value		0.14	0.001

Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L. & Kirchner, U. (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)*, pages 195–209.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002). *Journal of the Royal Statistical Society 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. *Journal of Statistical Planning and 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. *Computational Statistics & Data Analysis*, **55**(1), 588–602.

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

Conway, R. & Maxwell, W. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, **12**(2).

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

Table 7: Posterior summaries of the proportion of zeros (p), mean of y and dispersion parameters stratified by photoperiod levels and median of concentrations of the cytokinin BAP under ZICOM–Poisson regression model.

Parameter	Photoperiod	Mean	Median	Standar Deviation	HPD Interval (95%) LI	HPD Interval (95%) LS
p	8-hour	0.0308	0.0288	0.0134	0.0103	0.062
	16-hour	0.4371	0.4380	0.0473	0.3421	0.528
μ	8-hour	7.150	7.150	0.283	6.56	7.69
	16-hour	3.960	4.050	0.748	2.18	5.14
ϕ	8-hour	0.774	0.772	0.104	0.583	0.984
	16-hour	0.354	0.349	0.105	0.166	0.574

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

Cowles, M. K. & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, **91**, 883–904.

Dey, D. & Birmiwal, L. (1994). Robust bayesian analysis using divergence measures. *Statistics & Probability Letters*, **20**(4), 287–294.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, second edition.

Garay, A. M., Hashimoto, E. M., Ortega, E. M. & Lachos, V. H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics and Data Analysis*, **55**, 1304–1318.

Geisser, S. & Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, pages 153–160.

Gelfand, A., Dey, D. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods (with discussion). bayesian statistics 4. eds: J. bernardo et al. *BernardoJM, BergerJO, Dawid A P, Sm ith A FM(editors). OxfordUniversityPress*, **1**(4), 7–167.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. bayesian statistics 4. eds: J. bernardo et al. *BernardoJ M, BergerJ O, Dawid A P, Sm ith A FM(editors). OxfordUniversityPress*, **1**(4), 169–188.

Guikema, S. & Goffelt, J. (2008). A flexible count data regression model for risk analysis. *Risk analysis*, **28**(1), 213–223.

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

Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York.

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

Lawless, J. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 209–225.

Lord, D., Guikema, S. & Geedipally, S. (2008). Application of the conway-maxwell-poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention*, **40**(3), 1123–1134.

Mwalili, S., Lesaffre, E. & Declerck, D. (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. & Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics*, **23**(2), 199–213.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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.

Shmueli, G., Minka, T., Kadane, J., Borle, S. & Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(1), 127–142.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, **64**, 583–639.

Weiss, R. (1996). An approach to bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 739–750.

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