entropy

MDPI

Article

A New Regression Model for the Analysis of Overdispersed and Zero-Modified Count Data

Wesley Bertoli [1,*], Katiane S. Conceição [2], Marinho G. Andrade [2] and Francisco Louzada [2]

[1] Department of Statistics, Federal University of Technology, Paraná, Av. Sete de Setembro, 3165 Rebouças, Curitiba 80230-901, PR, Brazil
[2] Department of Applied Mathematics and Statistics, Institute of Mathematical and Computer Sciences, University of São Paulo, Av. Trab. São Carlense, 400 Parque Arnold Schimidt, São Carlos 13566-590, SP, Brazil; katiane@icmc.usp.br (K.S.C.); marinho@icmc.usp.br (M.G.A.); louzada@icmc.usp.br (F.L.)
* Correspondence: wbsilva@utfpr.edu.br; Tel.: +55-41-3310-4888

Abstract: Count datasets are traditionally analyzed using the ordinary Poisson distribution. However, said model has its applicability limited, as it can be somewhat restrictive to handling specific data structures. In this case, the need arises for obtaining alternative models that accommodate, for example, overdispersion and zero modification (inflation/deflation at the frequency of zeros). In practical terms, these are the most prevalent structures ruling the nature of discrete phenomena nowadays. Hence, this paper's primary goal was to jointly address these issues by deriving a fixed-effects regression model based on the hurdle version of the Poisson–Sujatha distribution. In this framework, the zero modification is incorporated by considering that a binary probability model determines which outcomes are zero-valued, and a zero-truncated process is responsible for generating positive observations. Posterior inferences for the model parameters were obtained from a fully Bayesian approach based on the g-prior method. Intensive Monte Carlo simulation studies were performed to assess the Bayesian estimators' empirical properties, and the obtained results have been discussed. The proposed model was considered for analyzing a real dataset, and its competitiveness regarding some well-established fixed-effects models for count data was evaluated. A sensitivity analysis to detect observations that may impact parameter estimates was performed based on standard divergence measures. The Bayesian *p*-value and the randomized quantile residuals were considered for the task of model validation.

Keywords: Bayesian inference; hurdle model; Monte Carlo simulation; overdispersion; Poisson–Sujatha distribution; zero-modified data

PACS: 02.50.-r

MSC: 62E15; 62J20; 62F15

Citation: Bertoli, W.; Conceição, K.S.; Andrade, M.G.; Louzada, F. A New Regression Model for the Analysis of Overdispersed and Zero-Modified Count Data. *Entropy* **2021**, *23*, 646. https://doi.org/10.3390/e23060646

Academic Editor: Jacinto Martín

Received: 17 December 2020
Accepted: 24 January 2021
Published: 21 May 2021

1. Introduction

The ordinary Poisson ($\mathcal{P}$) distribution is often adopted for the analysis of count data, mainly due to its simplicity and having computational implementations available for most of the standard statistical packages. However, it is well-known that such a model is not suitable to describe over/underdispersed counts. Apart from data transformation, the most popular approach to circumvent such an issue is based on using hierarchical models that can accommodate different overdispersion levels [1].

The negative binomial ($\mathcal{NB}$) distribution (that may arise as a $\mathcal{P}$ mixture model by using a gamma distribution for the continuous part) is undoubtedly the most popular alternative to model extra-$\mathcal{P}$ variability. There is extensive literature regarding other discrete mixed distributions that can accommodate different levels of overdispersion, for example, the Poisson–Lindley [2], the Poisson–lognormal [3], the Poisson–inverse Gaussian [4],

the negative binomial–Lindley [5], the Poisson–Janardan [6], the two-parameter Poisson–Lindley [7], the Poisson–Amarendra [8], the Poisson–Shanker [9], the Poisson–Sujatha ($\mathcal{PS}$) [10], the quasi-Poisson–Lindley [11], the weighted negative binomial–Lindley [12] the Poisson-weighted Lindley [13], the binomial-discrete Lindley [14], and the two-parameter Poisson–Sujatha [15], among many others.

Unfortunately, there is a significant drawback regarding such mixture models: they do not fit well when data present a modification in the frequency of zeros (typically underestimates the data dispersion and the frequency of zero-valued outcomes). The most common case in practice is the presence of an excessive number of zero-valued observations and a skewed distribution of positive values. In this way, developing $\mathcal{P}$-based two-part models (zero-inflated/hurdle models) became necessary. Prominent works addressing this task are [16–22].

Several authors have considered these approaches to analyze real data, and here we point out a few. Ref. [23] have sought to deal with the excess of zeros on data from recreational trips. Ref. [24] have shown that the modeling of migration frequency data can be improved using zero-inflated Poisson models. Ref. [25] have exploited the apple shoot propagation data, and they have addressed the modeling task by using several zero-inflated regression models. In the social sciences, ref. [26] have considered the hurdle version of the $\mathcal{P}$ model for the number of homicides in Chicago (State of Illinois, US). Ref. [27] provided an application to private health insurance count data using ordinary and zero-inflated Poisson regression models. Further applications of these models were considered in quantitative studies about HIV-risk reduction [28,29], for the modeling of some occupational allergic diseases in France [30], for the analysis of DNA sequencing data [31], and for the modeling of several datasets on chromosomal aberrations induced by radiation [32]. A Bayesian approach for the zero-inflated Poisson ($\mathcal{ZIP}$) distribution was considered by [33], and by [34] in a regression framework with fixed-effects.

Noticeably, most developed works are focused on the modeling of zero inflation, but zero-deflated data are also frequently observed in practice. However, there are still very few studies addressing this case [35], but this situation is often referred to in works handling zero inflation. In this context, a more comprehensive approach is provided by zero-modified models, which are flexible tools to handle count data with inflation/deflation at zero when there is no information about the nature of such a phenomenon.

Some of the most relevant works about zero-modified and hurdle models are cited in the following. Ref. [35] have introduced the zero-modified Poisson ($\mathcal{ZMP}$) regression model, and ref. [36] have considered such a model as an alternative for the analysis of Brazilian leptospirosis notification data. The possible loss due to the specification of a $\mathcal{ZMP}$ model for analyzing samples without zero modification was studied by [37] using the Kullback–Leibler divergence. The hurdle version of the power series distribution was presented and well discussed by [38], and ref. [39] have adopted a Bayesian approach for the zero-modified Poisson model to predict match outcomes of the Spanish La Liga (2012–2013). Besides, ref. [40] have proposed the zero-modified Poisson–Shanker regression model, whose usefulness was illustrated through its application to fetal death notification data, and ref. [41] have introduced the zero-modified Poisson–Lindley regression model with fixed-effects under a fully Bayesian approach.

Accordingly, this paper aims to extend the works of [42,43] in the sense of developing a new fixed-effects regression model for count data based on the zero-modified Poisson–Sujatha distribution ($\mathcal{ZMPS}$). Ref. [42] have introduced and exploited the theoretical $\mathcal{ZMPS}$ distribution's main statistical properties. On the other hand, ref. [43] have proposed a new class of zero-modified models, whose baseline distributions are Poisson mixtures, including the $\mathcal{PS}$. The present paper also extends the works of [40,41] since the $\mathcal{ZMPS}$ model differentiates from the zero-modified Poisson–Lindley and Poisson–Shanker by the ability, for example, to describe better (by adjusting its shape parameter) those discrete phenomena in which the probabilities of observing 0 s and 1 s are low (see [43], Figure 2).

Formally, a discrete random variable Y defined into $\mathbb{N}_0 = \{0, 1, \ldots\}$ is said to follow a $\mathcal{ZMPS}$ distribution if its probability mass function (pmf) can be written as

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

where p is the zero-modification parameter and δ_y is an indicator function, so that $\delta_y = 1$ if $y = 0$ and $\delta_y = 0$ otherwise. Additionally, $\mu \in \mathbb{R}_+$ is the expected value of the ordinary $\mathcal{PS}$ distribution, whose reparameterized pmf is given by

$$\begin{aligned} \mathrm{P}(Y = y; \mu) &= \frac{h^3(\mu)}{h^2(\mu) + h(\mu) + 2} \\ &\quad \left[\frac{y^2 + y[h(\mu) + 4] + [h^2(\mu) + 3h(\mu) + 4]}{[h(\mu) + 1]^{y+3}}\right], \quad y \in \mathbb{N}_0, \end{aligned}$$

where

$$h(\mu) = \frac{1}{3\mu}\left[(s(\mu) - \mu + 1) - \frac{(\mu - 1)(5\mu + 1)}{s(\mu)}\right], \tag{2}$$

with

$$s(\mu) = \left[3\mu\sqrt{21\mu^4 + 84\mu^3 + 513\mu^2 + 96\mu + 15} + 2\mu\left(4\mu^2 + 33\mu + 3\right) + 1\right]^{1/3},$$

and $\mu = (\theta^2 + 2\theta + 6)[\theta(\theta^2 + \theta + 2)]^{-1}$ for $\theta \in \mathbb{R}_+$ (shape parameter). This parameterization is particularly useful since our primary goal is to derive a regression model, in which the influence of fixed-effects can be evaluated directly over the mean of a zero-modified response variable. Unlike in zero-inflated models, here parameter p is defined on the interval $[0, \mathrm{P}(Y > 0; \mu)^{-1}]$, and so the $\mathcal{ZMPS}$ model is not a mixture distribution since p may assume values greater than 1. The expected value and variance of Y are given, respectively, by $\mathbb{E}(Y) = \lambda = \mu p$ and $\mathbb{V}(Y) = \varsigma^2 = p[\sigma^2 + (1 - p)\mu^2]$, where $\sigma^2 \in \mathbb{R}_+$ is the variance of the $\mathcal{PS}$ distribution (see [43], Table 4).

The hurdle version of the $\mathcal{PS}$ distribution can be obtained by taking $\omega = p\,\mathrm{P}(Y > 0; \mu)$, and so rewriting Equation (1) as

$$\mathrm{P}_*(Y = y; \mu, \omega) = (1 - \omega)\delta_y + \omega\,\mathrm{P}^*(Y = y; \mu), \quad y \in \mathbb{N}_0, \tag{3}$$

for $\omega \in [0, 1]$ and where $\mathrm{P}^*(Y = y; \mu)$ is the pmf of the zero-truncated Poisson–Sujatha ($\mathcal{ZTPS}$) distribution [44]. Noticeably, Equation (3) is only a reparameterization of the standard $\mathcal{ZMPS}$, and so one can conclude that these models are interchangeable. For ease of notation and understanding, the acronym $\mathcal{ZMPS}$ will be used when we refer to the hurdle version of the $\mathcal{PS}$ distribution.

The corresponding cumulative distribution function (cdf) of Y is given by

$$\begin{aligned} \mathrm{F}^*(y; \mu, \omega) &= 1 - \frac{\omega}{\mathrm{P}(Y > 0; \mu)}\left\{\frac{yh(\mu)\left[h^2(\mu) + (y + 6)h(\mu) + 2\right]}{[h^2(\mu) + h(\mu) + 2][h(\mu) + 1]^{y+3}} + \right. \\ &\quad \left. \frac{h^4(\mu) + 4h^3(\mu) + 10h^2(\mu) + 7h(\mu) + 2}{[h^2(\mu) + h(\mu) + 2][h(\mu) + 1]^{y+3}}\right\}, \quad y \in \mathbb{N}_0. \end{aligned} \tag{4}$$

Comparatively, the proposed model can be considered more flexible than zero-inflated models as it allows for zero-deflation, which is a structure often encountered when handling count data (see, for example, [45,46]). Besides, it can incorporate overdispersion that does not come only from inflation/deflation of zeros, as one of its parts is dedicated to describing the positive values' behavior. In the regression framework that we have developed, discrepant points (outliers) can be identified, and through a careful sensitivity analysis, it is possible to quantify the influences of such observations. However, since the

$\mathcal{PS}$ distribution accounts for different levels of overdispersion, its zero-modified version is naturally a robust alternative, as it may accommodate discrepant points that would significantly impact the parameter estimates of the $\mathcal{ZMP}$ model.

In this paper, the inferential procedures are conducted under a fully Bayesian perspective—an adaptation of the g-prior method [47] for the fixed-effects parameters is considered. The random-walk metropolis algorithm was used to draw pseudo-random samples from the posterior distribution of the model parameters. Local influence measures based on some well-known divergences were considered for the task of detecting influential points. Model validation metrics such as the Bayesian p-value and the randomized quantile residuals are presented. Intensive Monte Carlo simulation studies were performed to assess Bayesian estimators' empirical properties; the obtained results are discussed, and the overall performance of the adopted methodology was evaluated. Additionally, an application using a real dataset is presented to assess the proposed model's usefulness and competitivity.

This paper is organized as follows. In Section 2, we present the fixed-effects regression model based on the hurdle version of the $\mathcal{PS}$ distribution. In Section 3, we describe all the Bayesian methodologies and associated numerical procedures considered for inferential purposes. In Section 4, we discuss the results of an intensive simulation study, and in Section 5, a real data application using the proposed model is exhibited. General comments and concluding remarks are addressed in Section 6.

2. The ZMPS Regression Model

Suppose that a random experiment (designed or observational) is conducted with n subjects. The primary response for such an experiment is described by a discrete random variable Y_i denoting the outcome for the i-th subject. The full response vector is given by $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, and we assume that the observed vector $\boldsymbol{y}$ is obtained conditionally to fixed-effects, here denoted by $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Assuming that $Y_i|\boldsymbol{\beta} \sim \mathcal{ZMPS}(\mu_i, \omega_i)$ holds for all i, a general fixed-effects regression model for count data based on the $\mathcal{ZMPS}$ distribution can be derived by rewriting Equation (3) as

$$\mathrm{P}_*(Y_i = y_i; \boldsymbol{\beta}) = (1 - \omega_i)\delta_{y_i} + \omega_i \mathrm{P}^*(Y_i = y_i; \mu_i), \quad y_i \in \mathbb{N}_0, \tag{5}$$

where $\mu_i \equiv \mu(\boldsymbol{x}_{1i}, \boldsymbol{\beta}_1)$ and $\omega_i \equiv \omega(\boldsymbol{x}_{2i}, \boldsymbol{\beta}_2)$ are parameterized nonlinear functions. In this framework, we have $\boldsymbol{\beta}_k^\intercal = (\beta_{k0}, \ldots, \beta_{kq_k})$ $(k = 1, 2)$ related to $\boldsymbol{x}_{ki}^\intercal = (1, x_{ki}^1, \ldots, x_{ki}^{q_k})$, where $\boldsymbol{x}_{ki}$ is a vector of covariates that may include, for example, dummy variables, cross-level interactions, and polynomials. The quantity q_1 (q_2) denotes the number of covariates considered in the systematic component of a linear predictor for parameter μ_i (ω_i). The full regression matrices of model (5) can be written as $\mathbf{X}_k = (\mathbf{1}_n, \mathbf{X}_{k,n\times q_k})$, where $\mathbf{1}_n$ is the intercept column and the submatrix $\mathbf{X}_{k,n\times q_k}$ is defined in such a way that its i-th row contains the vector $(x_{ki}^1, \ldots, x_{ki}^{q_k})$. The overall dimension of $\mathbf{X}_k$ is $n \times (q_k + 1)$.

Now, we have to specify two monotonic, invertible, and twice differentiable link functions, say g_1 and g_2, in which $\mu_i = g_1^{-1}(\boldsymbol{x}_{1i}^\intercal \boldsymbol{\beta}_1)$ and $\omega_i = g_2^{-1}(\boldsymbol{x}_{2i}^\intercal \boldsymbol{\beta}_2)$ are well defined on $\mathbb{R}_+$ and $(0, 1)$, respectively. For this purpose, one may choose any suitable mappings g_1 and g_2 such that $g_1^{-1}\colon \mathbb{R} \to \mathbb{R}_+$ and $g_2^{-1}\colon \mathbb{R} \to (0, 1)$. The logarithm link function, $\log(\mu_i) = \boldsymbol{x}_{1i}^\intercal \boldsymbol{\beta}_1$, is the natural choice for g_1. For g_2, the popular choice is the logit link function,

$$\mathrm{logit}(\omega_i) = \log\left(\frac{\omega_i}{1 - \omega_i}\right) = \boldsymbol{x}_{2i}^\intercal \boldsymbol{\beta}_2. \tag{6}$$

The probit link function,

$$\Phi^{-1}(\omega_i) = \boldsymbol{x}_{2i}^\intercal \boldsymbol{\beta}_2, \tag{7}$$

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

$$\log[-\log(1 - \omega_i)] = \boldsymbol{x}_{2i}^\intercal \boldsymbol{\beta}_2, \tag{8}$$

which corresponds to the complementary log–log link function. One can notice that these link functions exclude the limit cases $p_i = 0$ and $p_i = \mathrm{P}(Y > 0; \mu_i)^{-1}$. The link Function (8) is usually preferable when the occurrence probability of a specific outcome is considerably high/low as it accommodates asymmetric behaviors on the unit interval, which is not the case for link Functions (6) and (7). Besides, a more sophisticated approach considering power and reversal power link functions was proposed by [48], and can also be used to add even more flexibility when modeling parameter ω_i.

We may refer to the proposed model as a "semi-compatible" regression model. The term "compatible" alludes to "zero-altered," which defines the class proposed by [49], and extended by [50] in a setting including semiparametric zero-altered models that accommodate over/underdispersion. Zero-altered models are similar to zero-modified ones, but the compatibility arises from the linear predictors of μ_i and ω_i being the same. In our case, specifically, it is worthwhile to mention that identifiability problems may occur if one considers a fixed-effects regression model derived directly from (3), with parameters μ and p sharing covariates, even if $\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$. Therefore, the adopted structure allows for more flexibility and robustness as μ and ω may share covariates not necessarily with $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1$, and so the only requirement for ensuring model identifiability is the linear independence between covariates within linear predictors.

Given a set of covariates, the probability of a zero-valued count being observed for the i-th subject is given by $1 - g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)$. Under the logistic regression model (6), β_{2l} $(l = 1, \ldots, q_2)$ represents the direct change in the log-odds of Y_i, it being positive per 1-unit change in x_{2i}^{l}, while holding the other covariates at fixed values. On the other hand, the same not apply if one adopts the link Function (8) since $e^{\beta_{2l}}$ is not the odds ratio for the l-th covariate effect, and so β_{2l} does not have a straightforward interpretation in terms of contribution to log-odds. Likewise, it is not possible to interpret the coefficients of the probit model (7) directly, but one can evaluate the marginal effect of β_{2l} by analyzing how much the conditional probability of Y_i being positive is affected when the value of x_{2i}^{l} is changed. The exact interpretation of β_{1l} $(l = 1, \ldots, q_1)$ is not direct in terms of the mean of the hurdle model since the positive counts are modeled by a zero-truncated distribution $(\mathcal{ZTPS})$, and therefore, β_{1l} represents the overall effect of x_{1i}^{l} on the expected value μ_i when $y_i > 0$, while holding the other covariates at fixed values.

The proposed model has $d = \dim(\boldsymbol{\beta}) = q_1 + q_2 + 2$ unknown quantities to be estimated. A fully Bayesian approach will be considered for parameter estimation and associated inference. The next section is dedicated to present details of such an approach.

3. Inference

In this section, we address the problem of estimating and making inferences about the proposed model from a fully Bayesian perspective. Firstly, we derive the model likelihood function, and then, a suitable set of prior distributions is considered to obtain a computationally tractable posterior density for the vector $\boldsymbol{\beta}$. Beyond the primary distributional assumption that $Y_i|\boldsymbol{\beta} \sim \mathcal{ZMPS}(\mu_i, \omega_i)$ holds for all i, here we also assume that the outcomes for different subjects are unconditionally independent.

Let Y be a discrete random variable assuming values on $\mathbb{N}_0$. Suppose that a random experiment is carried out n times independently and, subject to $\boldsymbol{x}_{ki}$ for each i, a vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ of observed values from Y is obtained. Considering model Formulation (5), the likelihood function of $\boldsymbol{\beta}$ can be written as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}; \boldsymbol{y}) &= \prod_{i=1}^{n} \omega_i \left(\frac{1-\omega_i}{\omega_i}\right)^{\delta_{y_i}} \left[\frac{\mathrm{P}(Y_i = y_i; \mu_i)}{\mathrm{P}(Y_i > 0; \mu_i)}\right]^{1-\delta_{y_i}} \\
&= \prod_{i=1}^{n} g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2) \left[\frac{1 - g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)}{g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)}\right]^{\delta_{y_i}} \left\{\frac{\mathrm{P}\left[Y_i = y_i; g_1^{-1}(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1)\right]}{\mathrm{P}\left[Y_i > 0; g_1^{-1}(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1)\right]}\right\}^{1-\delta_{y_i}},
\end{aligned}$$

and so the corresponding log-likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\beta};\boldsymbol{y}) &= \sum_{i=1}^{n}(1-\delta_{y_i})\log\left\{\frac{\mathrm{P}\left[Y_i=y_i;g_1^{-1}(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1)\right]}{\mathrm{P}\left[Y_i>0;g_1^{-1}(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1)\right]}\right\}+\\ &\quad \sum_{i=1}^{n}\left\{\log\left[g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)\right]-\delta_{y_i}\log\left[\frac{g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)}{1-g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)}\right]\right\} \qquad (9)\\ &= \ell_1(\boldsymbol{\beta}_1;\boldsymbol{y})+\ell_2(\boldsymbol{\beta}_2;\boldsymbol{y}).\end{aligned}$$

In this work, we will consider a log-linear model for parameter μ_i, that is, $g_1(\mu_i) = \log(\mu_i) = \boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1$. The choice of g_2 is left open and the notation $\omega_i = g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)$ will be used when necessary. From Equation (9), one can easily notice that the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal and that ℓ_1 depends only on the positive values of $\boldsymbol{y}$. In this way, the log-likelihood function of $\boldsymbol{\beta}_1$ takes the form

$$\begin{aligned}\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y}) &= \sum_{j\in\mathcal{J}_1}\log\left\{y_j^2+y_j\left[h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+4\right]+\left[h^2\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+3h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+4\right]\right\}-\\ &\quad \sum_{j\in\mathcal{J}_1}\log\left[h^4\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+4h^3\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+10h^2\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+7h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+2\right]+ \qquad (10)\\ &\quad 3\sum_{j\in\mathcal{J}_1}\log\left[h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)\right]-\sum_{j\in\mathcal{J}_1}y_j\log\left[h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+1\right],\end{aligned}$$

where $\mathcal{J}_1 = \{j : y_j > 0, y_j \in \boldsymbol{y}\}$ is the finite set of indexes regarding the positive observations of $\boldsymbol{y}$. Adopting this setup is equivalent to assuming that each positive element of $\boldsymbol{y}$ comes from a $\mathcal{ZTPS}$ distribution. Here, we are extending the fact that estimating the $\mathcal{P}$ parameter θ using the zero-truncated Poisson ($\mathcal{ZTP}$) distribution results in a loss of efficiency in the inference if there is no zero modification [35,37]. Now, the log-likelihood function of $\boldsymbol{\beta}_2$ can be written as

$$\ell_2(\boldsymbol{\beta}_2;\boldsymbol{y}) = \sum_{i=1}^{n}\log\left[g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)\right]-\sum_{j\in\mathcal{J}_2}\log\left[\frac{g_2^{-1}\left(\boldsymbol{x}_{2j}^{\intercal}\boldsymbol{\beta}_2\right)}{1-g_2^{-1}\left(\boldsymbol{x}_{2j}^{\intercal}\boldsymbol{\beta}_2\right)}\right], \qquad (11)$$

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

3.1. Prior Distributions

The g-prior [47] is a popular choice among Bayesian users of the multiple linear regression model, mainly due to the fact of providing a closed-form posterior distribution for the regression coefficients. The g-prior is classified as an objective prior method which uses the inverse of the Fisher information matrix up to a scalar variance factor to obtain the prior correlation structure of the multivariate normal distribution. Such specification is quite attractive since the Fisher information plays a major role in determining large-sample covariance in both Bayesian and classical inference.

The problem of eliciting conjugate priors for a GLM was addressed by [51]. Their approach can be considered as a generalization of the original g-prior method. Still, its application is restricted for the class of GLMs since the proposed prior does not have closed-form for non-normal exponential families. Alternatively, ref. [52] have proposed the information matrix prior as a way to assess the prior correlation structure between the coefficients, not including the intercept since the regression matrix is centered as to ensure that β_0 is orthogonal to the other coefficients. This method uses the Fisher information similarly to a precision matrix whose elements are shrunken by a fixed variance factor. However, the authors have pointed out that such class of priors can only be considered Gaussian priors if the Fisher information matrix does not depend on the vector

$\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_q)$. In this way, ref. [53] had considered a similar approach when they proposed a class of hyper-g priors for GLMs, where the precision matrix is evaluated at the prior mode, hence obtaining an information matrix that is $\boldsymbol{\beta}'$ free.

The formal concept behind the information matrix prior is closely related to the unit information prior [54], whose main idea is that the amount of information provided by a prior distribution must be the same as the amount of information contained in a single observation. Such an idea can be applied in the previously mentioned approaches by simply considering the total sample size (n) as the variance factor. Ref. [52] have also considered fixed values for the scalar variance factor. On the other hand, some works, including [53,55,56] do consider prior elicitation and inference procedures for the variance scale factor. Here, we will adopt a methodology based on the unit information prior idea combined with the "noninformative g-prior" proposed by [57] for binary regression models. Based on such an approach, it is possible to obtain a quite simple prior distribution for the fixed-effects of the proposed model as $\boldsymbol{\beta}_k \sim \mathcal{N}_{\overline{q}_k}(\mathbf{0}, n(\mathbf{X}_k^\intercal \mathbf{X}_k)^{-1})$, where $\overline{q}_k = q_k + 1$.

It is worthwhile to mention that, in cases where $\mathbf{X}_k$ is rank deficient $(n < q_k + 1)$ or contains collinear covariates, it is highly advisable to compute the generalized inverse of $\mathbf{X}_k^\intercal \mathbf{X}_k$ otherwise the prior covariance matrix of $\boldsymbol{\beta}_k$ may not be defined.

Analogously to Marin and Robert's approach, we do not consider centered regression matrices in the prior specification. Hence, we are able to include β_{10} in the proposed g-prior but, in this case, the intercept is a priori correlated with the other coefficients $(\beta_{11}, \ldots, \beta_{1q_1})$. The same applies for β_{20} and the vector $(\beta_{21}, \ldots, \beta_{2q_2})$.

3.2. Posterior Distributions and Estimation

Considering the outlined structure for the $\mathcal{ZMPS}$ regression model, the unnormalized joint posterior distribution of the unknown vector $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}; \boldsymbol{y}) \propto \exp\{\ell(\boldsymbol{\beta}; \boldsymbol{y})\}\pi(\boldsymbol{\beta}). \tag{12}$$

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

$$\pi_1(\boldsymbol{\beta}_1; \boldsymbol{y}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1; \boldsymbol{y})\}\pi_1(\boldsymbol{\beta}_1) \quad \text{and} \quad \pi_2(\boldsymbol{\beta}_2; \boldsymbol{y}) \propto \exp\{\ell_2(\boldsymbol{\beta}_2; \boldsymbol{y})\}\pi_2(\boldsymbol{\beta}_2), \tag{13}$$

where ℓ_1 and ℓ_2 are given by (10) and (11), respectively. Naturally, in the discrete setting, the use of proper (Gaussian) priors prevents π_1 and π_2 from being improper.

From the Bayesian point of view, inferences for the elements of $\boldsymbol{\beta}_k$ can be obtained from their marginal posterior distributions. However, deriving analytical expressions for these densities is infeasible, mainly due to the associated log-likelihood function's complexity. In this case, to make inferences for $\boldsymbol{\beta}_k$, we must resort to a suitable iterative procedure to drawn pseudo-random samples from their posterior densities. Hence, aiming to generate N chains for $\boldsymbol{\beta}_k$, we will adopt the well-known random-walk metropolis (RwM) algorithm [58,59]. For the posterior densities in (13), we consider a multivariate normal distributions for the proposal (candidate-generating) densities in the algorithm. These distributions will be used as the main terms in the transition kernels when computing the acceptance probabilities. Hence, at any state $t > 0$, the MCMC simulation are performed by proposing a candidate $\boldsymbol{\psi}_k$ for $\boldsymbol{\beta}_k$ as

$$\boldsymbol{\psi}_k | \boldsymbol{\beta}_k^{(t-1)} \sim \mathcal{N}_{\overline{q}_k}\left[\nu \boldsymbol{\beta}_k^{(t-1)}, \nu \mathcal{S}_k^{(t-1)}\right],$$

where $\nu = n(n+1)^{-1}$. One can notice that transitions depend on the acceptance of pseudo-random vectors generated with mean given by the actual state of the chain, which is shrunken by the factor ν. Besides, at any state $t > 0$, the covariance matrix of the candidate vector $\boldsymbol{\psi}_k$ can be approximated numerically by evaluating $\mathcal{S}_k = \mathcal{H}_k^{-1}$ at $\boldsymbol{\beta}_k = \nu \boldsymbol{\beta}_k^{(t-1)}$, where

$$\mathcal{H}_k = -\frac{\partial^2 \log[\pi_k(\boldsymbol{\beta}_k; \boldsymbol{y})]}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_k^\intercal}.$$

The procedure to generate pseudo-random samples from the approximate posterior distribution of $\boldsymbol{\beta}$ is summarized in Algorithm A1 (see Appendix A). To run it, one has to specify the size of chains to be generated (N) and the initial state vectors $\boldsymbol{\beta}_1^{(0)}$ and $\boldsymbol{\beta}_2^{(0)}$ beforehand. For a specific asymptotic Gaussian environment, [59] have shown that the optimal acceptance rate should be around 45% for 1-dimensional problems and asymptotically approaches to 23.40% in higher-dimensional problems. We consider acceptance rates varying between 23.40% and 32% as quite reasonable since the proposed model will generally have at least four parameters to be estimated. Indeed, the higher the value of n, the lower the acceptance rate in the RwM algorithm, which results in lower variability of estimates.

The convergence of the simulated sequences can be monitored by using trace and autocorrelation plots, and the run-length control method with a half-width test [60], the Geweke z-score diagnostic [61], and the Brooks-Gelman-Rubin scale-reduction statistic [62]. After diagnosing convergence, some samples can be discarded as burn-in. The strategy to decrease the correlation between and within generated chains is based on getting thinned steps, and so the final sample is supposed to have size $M \ll N$ for each parameter. A full descriptive summary of the posterior distribution (12) can be obtained through Monte Carlo (MC) estimators using the sequence $\{\boldsymbol{\beta}^t\}_{t=1}^{M}$. We choose the posterior expected value as the Bayesian point estimator for $\boldsymbol{\theta}$, that is,

$$\hat{\boldsymbol{\beta}} = \frac{1}{M}\sum_{t=1}^{M}\boldsymbol{\beta}^{(t)}, \tag{14}$$

which is also known as the minimum mean square error estimator.

In the next section, we discuss the results of the Monte Carlo simulation studies performed to assess the proposed Bayesian methodology's performance. In Section 5, the proposed model's usefulness and competitivity are illustrated by using a real dataset. All computations were performed using the `R` environment [63]. The executable scripts were made available at the publisher's website.

3.3. Posterior Predictive Distribution

In a Bayesian context, the posterior predictive distribution (ppd) is defined as the distribution of possible future (unobserved) values conditioned on the observed ones. Under the $\mathcal{ZMPS}$ distribution, the pmf of any observation $w \in \mathbb{N}_0$ (subject to the vectors $\boldsymbol{x}_{1w}^{\intercal}$ and $\boldsymbol{x}_{2w}^{\intercal}$ of covariates) is given by

$$\begin{aligned}
\mathrm{P}_{\pi}(Y=w) &= \int_{\mathbb{R}^d} \mathrm{P}^{*}(Y=w;\mu_w,\omega_w)\pi(\boldsymbol{\beta};\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta} \\
&= \int_{\mathbb{R}^{\bar{q}_1}} \left[\frac{\mathrm{P}\left(Y=w;e^{\boldsymbol{x}_{1w}^{\intercal}\boldsymbol{\beta}_1}\right)}{\mathrm{P}\left(Y>0;e^{\boldsymbol{x}_{1w}^{\intercal}\boldsymbol{\beta}_1}\right)}\right]^{1-\delta_w} \pi_1(\boldsymbol{\beta}_1;\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta}_1 \times \\
&\quad \int_{\mathbb{R}^{\bar{q}_2}} g_2^{-1}(\boldsymbol{x}_{2w}^{\intercal}\boldsymbol{\beta}_2)\left[\frac{1-g_2^{-1}(\boldsymbol{x}_{2w}^{\intercal}\boldsymbol{\beta}_2)}{g_2^{-1}(\boldsymbol{x}_{2w}^{\intercal}\boldsymbol{\beta}_2)}\right]^{\delta_w} \pi_2(\boldsymbol{\beta}_2;\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta}_2,
\end{aligned}$$

where $\delta_w = 1$ if $w = 0$ and $\delta_w = 0$ otherwise. Noticeably, the ppd has no closed-form available, and therefore, an MC estimator for this quantity is given by

$$\hat{\mathrm{P}}_{\pi}(Y=w) = \frac{1}{M^2}\sum_{t=1}^{M} g_2^{-1}\left(\boldsymbol{x}_{2w}^{\intercal}\boldsymbol{\beta}_2^{(t)}\right)\left[\frac{1-g_2^{-1}\left(\boldsymbol{x}_{2w}^{\intercal}\boldsymbol{\beta}_2^{(t)}\right)}{g_2^{-1}\left(\boldsymbol{x}_{2w}^{\intercal}\boldsymbol{\beta}_2^{(t)}\right)}\right]^{\delta_w}\sum_{t=1}^{M} b_t(w), \tag{15}$$

where

$$b_t(w) = \left\{ \frac{h^3\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right)}{h^4\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 4h^3\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 10h^2\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 7h\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 2} \times \frac{w^2 + w\left[h\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 4\right] + \left[h^2\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 3h\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 4\right]}{\left[h\left(e^{x_{1w}^\intercal \beta_1^{(t)}}\right) + 1\right]^w} \right\}^{1-\delta_w}.$$

From Equation (15), one can easily estimate, for example, the posterior probability of $Y = 0$ (subject to $\boldsymbol{x}_{10}^\intercal$ and $\boldsymbol{x}_{20}^\intercal$) as

$$\hat{\mathrm{P}}_\pi(Y = 0; \boldsymbol{x}_{10}^\intercal, \boldsymbol{x}_{20}^\intercal) = \frac{1}{M} \sum_{t=1}^{M} g_2^{-1}\left(\boldsymbol{x}_{20}^\intercal \boldsymbol{\beta}_2^{(t)}\right) \left[\frac{1 - g_2^{-1}\left(\boldsymbol{x}_{20}^\intercal \boldsymbol{\beta}_2^{(t)}\right)}{g_2^{-1}\left(\boldsymbol{x}_{20}^\intercal \boldsymbol{\beta}_2^{(t)}\right)}\right].$$

4. Simulation Study

The empirical properties of an estimator can be accessed through Monte Carlo simulations. In this way, we have performed an intensive simulation study aiming to validade the Bayesian approach in some specific situations. The simulation process was carried out by generating 500 pseudo-random samples of sizes $n = 50, 100, 200,$ and 500 of a variable Y following a $\mathcal{ZMPS}$ distribution under the regression framework presented in Section 2. For the whole process, it was considered a $n \times 2$ regression matrix $\mathbf{X}_1 = (\mathbf{1}_n, \mathbf{X}_{1,n\times 1})$ in which $\mathbf{X}_{n\times 1}$ is a vector containing n generated values from a Uniform distribution on the unit interval. Here, we have fixed $\mathbf{X}_2 = \mathbf{X}_1$. Moreover, we have assigned different values for the vectors $\boldsymbol{\beta}_1^\intercal = (\beta_{10}, \beta_{11})$ and $\boldsymbol{\beta}_2^\intercal = (\beta_{20}, \beta_{21})$ in order to generate both zero-inflated and zero-deflated artificial samples. The logarithm link function was considered for g_1. For g_2, we have considered the link Functions (6)–(8) as a way to evaluate how these different specifications affect the estimation of $\boldsymbol{\beta}$.

Algorithm A2 (see Appendix A) can be used to generate a single pseudo-random realization from the $\mathcal{ZMPS}$ distribution in the regression framework with covariate $\mathcal{U}(0,1)$ for μ and ω. The extension for the use of more covariates is straightforward. The process to generate a pseudo-random sample of size n consists of running the algorithm as often as necessary, say n^* times $(n^* \geqslant n)$. The sequential search is a black-box algorithm and works with any computable probability vector. The main advantage of such a procedure is its simplicity. On the other hand, sequential search algorithms may be slow as the while-loop may have to be repeated very often. More details about this algorithm can be found at [64].

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

$$\mathbb{E}(\mathrm{NI}) = \lambda + 1 = \frac{\omega\mu\left[h^2(\mu) + h(\mu) + 2\right]\left[h(\mu) + 1\right]^3}{h^4(\mu) + 4h^3(\mu) + 10h^2(\mu) + 7h(\mu) + 2} + 1,$$

where $h(\mu)$ is given by Equation (2).

We have considered four scenarios for each kind of zero-modification. Table 1 presents the true parameter values that were considered in our study. For the zero-inflated (zero-deflated) case, the samples were generated from the $\mathcal{ZMPS}$ distribution by considering that $p_i \in (0,1)$ $(p_i \in [1, \mathrm{P}(Y > 0; \mu_i)^{-1}])$ for all i. Here, the regression coefficients were chosen by taking into account that zero-inflated (zero-deflated) samples have, naturally, proportion of zeros greater (lower) than expected under an ordinary count distribution and therefore, the variable Y_i $(i = 1, \ldots, n)$ was generated with mean far from zero (close to zero). Table 1 also presents the range of parameters μ_i and p_i in each scenario. The bounds

were obtained by evaluating the linear predictors $\beta_{10} + \beta_{11}x$ and $\beta_{20} + \beta_{21}x$ at $x = 0$ and $x = 1$ (limit values of the adopted covariate). Scenarios 1 and 2 of the zero-inflated case were considered to illustrate the Bayesian estimators' behavior when the proposed model is used to fit (right) long-tailed count data.

Table 1. Actual parameter values for simulation of zero-modified artificial datasets.

Case	Scenario	Link	β_{10}	β_{11}	β_{20}	β_{21}	Range μ_i	Range p_i
I	1	Logit	1.50	3.00	−1.00	−1.00	(4.48; 90.02)	(0.12; 0.30)
		Probit						(0.02; 0.18)
		CLL						(0.13; 0.34)
	2	Logit	1.50	3.00	−1.00	0.50	(4.48; 90.02)	(0.30; 0.38)
		Probit						(0.18; 0.31)
		CLL						(0.34; 0.45)
	3	Logit	1.50	−1.50	−1.00	−1.00	(1.00; 4.48)	(0.23; 0.30)
		Probit						(0.04; 0.18)
		CLL						(0.24; 0.34)
	4	Logit	1.50	−1.50	−1.00	0.50	(1.00; 4.48)	(0.30; 0.73)
		Probit						(0.18; 0.59)
		CLL						(0.34; 0.88)
D	1	Logit	−1.00	1.00	0.50	0.50	(0.37; 1.00)	(1.41; 2.30)
		Probit						(1.62; 2.56)
		CLL						(1.80; 2.99)
	2	Logit	−1.00	1.00	1.50	−1.00	(0.37; 1.00)	(1.20; 3.02)
		Probit						(1.33; 3.45)
		CLL						(1.56; 3.66)
	3	Logit	−1.00	−1.50	0.50	0.50	(0.08; 0.37)	(2.30; 9.64)
		Probit						(2.56; 11.09)
		CLL						(2.99; 12.31)
	4	Logit	−1.00	−1.50	1.50	−1.00	(0.08; 0.37)	(3.02; 8.21)
		Probit						(3.45; 9.12)
		CLL						(3.66; 10.65)

I: inflation; D: deflation; and CLL: complementary log–log.

To apply the proposed Bayesian approach to each scenario, we have considered the RwM algorithm for MCMC sampling. For each generated sample, a chain with $N = 50{,}000$ values was generated for each parameter, considering a burn-in period of 20% of the chain size. To obtain pseudo-independent samples from the posterior distributions given in (13), one out every 10 generated values were kept, resulting in chains of size $M = 4000$ for each parameter. Using trace plots and Geweke's z-score diagnostic, the remaining chains' stationarity was revealed. When running the simulations, the acceptance rates were ranging between 23.40% and 32%. The posterior mean (14) was considered as the Bayesian point estimator, and its performance was studied by assessing its bias (B), its mean squared error (MSE), and its mean absolute percentage error (MAPE). Besides, the coverage probability (CP) of the 95% highest posterior density intervals (HPDIs) was also estimated.

Using the generated samples and letting $\gamma = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the MC estimators for these measures are given by

$$\widehat{\mathrm{B}}_{\hat{\gamma}} = \frac{1}{500}\sum_{j=1}^{500}(\hat{\gamma}_j - \gamma),\ \widehat{\mathrm{MSE}}_{\hat{\gamma}} = \frac{1}{500}\sum_{j=1}^{500}(\hat{\gamma}_j - \gamma)^2,\ \text{and}\ \widehat{\mathrm{MAPE}}_{\hat{\gamma}} = \frac{1}{500}\sum_{j=1}^{500}\left|\frac{\hat{\gamma}_j - \gamma}{\gamma}\right|.$$

The variance of $\hat{\gamma}$ was estimated as the difference between the MSE and the square of the bias. Moreover, the CP of the HPDIs was estimated by

$$\widehat{\mathrm{CP}}_{\gamma} = \frac{1}{500}\sum_{j=1}^{500}\delta_j(\gamma),$$

where $\delta_j(\gamma)$ assumes 1 if the j-th HPDI contains the true value γ and 0 otherwise. We have also estimated the below noncoverage probability (BNCP) and the above noncoverage prob-

ability (ANCP) of the HPDIs. These measures are computed analogously to CP. The BNCP and ANCP may be useful measures to determine asymmetrical behaviors as they provide the probabilities of finding the actual value of γ on the tails of its posterior distribution.

Due to the massive amount of results, the obtained results were made available on the publisher's website as supplementary material. In our study, we have noticed that, as expected, the parameter estimates became more accurate with increasing sample sizes since the estimated biases and mean squared errors have decreased considerably as n increased. The squared ratio between the mean squared error and the estimated variance approaches 1 as n increases. Although high MAPE values were obtained for some parameters (when using small sample sizes), this does not compromise the overall estimation accuracy. For example, when $n = 100$, we have obtained a estimated MAPE value of approximately 56% for β_{11} (see Table S25, Scenario S1, Supplementary Material). Taking into account the true value of such parameter (1.00), we have that the estimates for β_{11} were ranging mostly between 0.44 and 1.56, which do not represent a significant impact on the estimated mean (μ). When (right) long-tailed count data are available, the CP of the HPDI for β_{11} is considerably lower than the adopted nominal level (for small sample sizes) as its posterior distribution tends to be more asymmetric towards higher values on the parameter space. However, we have observed that the estimated CP of the HPDIs is converging to 95% in both zero-modified cases, and the posterior distributions became more symmetric with increasing sample sizes.

Considering the predefined scenarios, we conclude that our simulation study provides favorable indications about the adopted Bayesian methodology's suitability to estimate the parameters of the proposed model. We believe that in a similar procedure with a different set of actual values, the estimators' overall behavior should resemble the results that we have described here. Besides, the adopted methodology would also be reliable if one or more than one covariates (possibly of other nature) were included in the linear predictors of μ_i and ω_i.

5. Chromosomal Aberration Data Analysis

In this section, the $\mathcal{ZMPS}$ regression model is considered for analyzing a real dataset obtained from a cytogenetic dosimetry experiment that was first presented by [65]. In this study, the response variable is the number of cytogenetic chromosomal aberrations after the DNA molecule is treated with induced radiation. The dataset was obtained by irradiating five blood samples from a healthy donor with different doses x_i $(i = 1, \ldots, 5)$ ranging between 0.1 and 1.0 Gy with 2.1 MeV neutrons in three different culture times (48 h, 56 h, and 72 h), considering partial-body exposure-densely ionizing radiation. In the following, n_i cells were examined in each irradiated sample and the number of dicentrics and centric ring aberrations y_{ij} $(j = 1, \ldots, n_i)$ was recorded.

While [65] have used a t-test to analyze whether the averages of the relative number of dicentrics plus centric ring aberration frequencies differed significantly between the three different culture times, we are primarily interested in evaluating if the averages of the number of dicentrics plus centric ring aberration differ significantly between doses of ionizing radiation, considering data from culture times of 72 h.

The frequency distribution of the collected data is available in Table 2, along with some descriptive statistics. From the observed dataset, there exist evidences that the response variable is slightly overdispersed since $\overline{y}_{\cdot} = 0.131 < s^2_{\cdot} = 0.210$ and $s^2_{\cdot}/\overline{y}_{\cdot} = 1.607$. Additionally, the number of aberrations appears to be heavily zero-inflated, as shown in the left-panel of Figure 1. On the other hand, one can notice that, as the dose of ionizing radiation increases, the number of observed zeros decreases. Still, the distribution becomes more overdispersed since it naturally increases the number of aberrations.

Table 2. Descriptive summary of the numbers of dicentrics and centric ring aberrations.

x_i	y_{ij}						n_i	$\overline{y}_i$	s_i	$s_i^2/\overline{y}_i$
	0	**1**	**2**	**3**	**4**	**5**				
0.1	2130	59	9	2	0	0	2200	0.038	0.224	1.316
0.3	1088	84	19	6	3	0	1200	0.127	0.449	1.591
0.5	875	88	30	7	0	0	1000	0.169	0.493	1.438
0.7	679	88	23	8	1	1	800	0.209	0.568	1.545
1.0	480	75	27	13	5	0	600	0.313	0.732	1.712

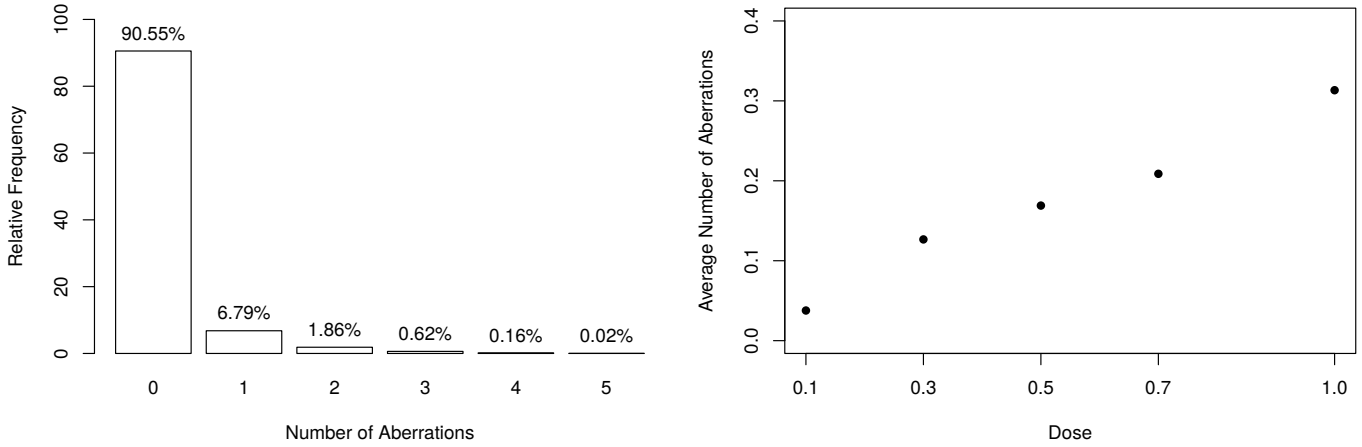


Figure 1. Summary of the numbers of dicentrics and centric ring aberrations.

According to [32], when considering higher linear energy transfer radiations, the incidence of chromosomal aberrations becomes a linear function of the dose because the more densely ionizing nature of the radiation leads to an "one track" distribution of damage. Such an aspect can be seen in the right-panel of Figure 1, which highlights the linear behavior between the average number of aberrations and the doses. In this way, our assumption is that $Y_{ij}|x_i \sim \mathcal{ZMPS}(\mu_{ij}, \omega_{ij})$, where parameters μ_{ij} and ω_{ij} are specified as linear dose models, that is,

$$\log(\mu_{ij}) = \beta_{10} + \beta_{11}x_i \quad \text{and} \quad g_2(\omega_{ij}) = \beta_{20} + \beta_{21}x_i.$$

To fit the $\mathcal{ZMPS}$ regression model with dose as the only covariate, we have adopted the same procedure used in the previous section. The link Function (7) was chosen to relate ω_{ij} with the linear predictor $\beta_{20} + \beta_{21}x_i$ and so we have the probit hurdle regression model. In this framework, the coefficient β_{11} represents the effect of the dose of ionizing radiation on the expected count μ_i when $Y_{ij} > 0$, and β_{21} indicates the effect of the dose on the probability of aberrations to occur. We have considered the RwM for MCMC sampling, generating a chain of size $N = 50{,}000$ for each parameter whereby the first 10,000 values were discarded as burn-in. The stationarity of the chains was revealed using the Geweke z-score diagnostic of convergence. To obtain the pseudo-independent samples from the posterior distributions given in (13), we have considered one value out of every 10 generated ones, resulting in chains of size $M = 4000$ for each parameter.

Table 3 presents the posterior parameter estimates and 95% HPDIs from $\mathcal{ZMPS}$ fitted model. When obtaining the MCMC samples, the acceptance rate in the RwM algorithm was approximately 32%. Besides, we have computed the number of effectively pseudo-independent draws, that is, the Effective Sample Size (ESS) for each parameter. Figures 2 and 3 depict the chains' history (trace plots) and the marginal posterior distributions of the regression coefficients. The normality assumption of the generated chains is quite reasonable, even with slight tails on the estimated densities. Additionally, there exists

evidence of symmetry since the posterior means and medians are very close to each other. For each parameter, the ESS was estimated at approximately half of M, indicating a good mixing of the generated chains without computational waste.

Table 3. Posterior parameter estimates and 95% highest posterior density intervals (HPDIs) from $\mathcal{ZMPS}$ fitted model.

Parameter	Mean	Median	Std. Dev.	ESS	95% HPDI	
					Lower	Upper
β_{10}	−1.481	−1.479	0.192	1874.876	−1.868	−1.119
β_{11}	0.935	0.937	0.279	1912.372	0.411	1.497
β_{20}	−1.790	−1.789	0.044	1834.592	−1.873	−1.706
β_{21}	1.062	1.063	0.074	1910.648	0.924	1.211

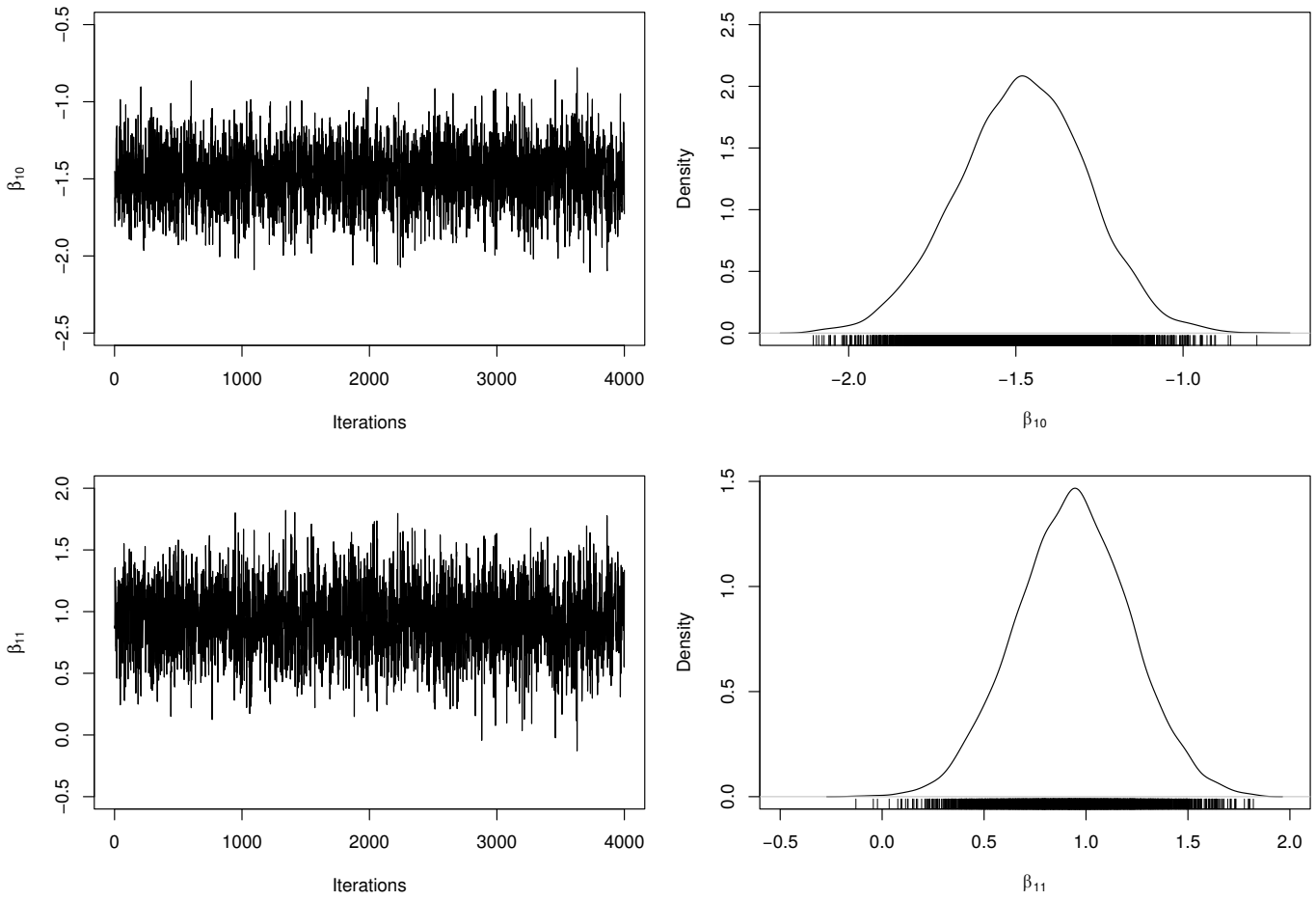


Figure 2. Trace plots and marginal posterior distributions of parameters β_{10} and β_{11} from the $\mathcal{ZMPS}$ regression model.

A sensitivity analysis to verify the existence of influential points is presented in Figure 4. We have estimated all divergence measures presented in Table A1 but, since the obtained results led to the same conclusions, we are only reporting the KL and H divergences and their calibration for each observation. Even being very conservative by considering an observation whose distance has a calibration exceeding 0.65 as an influential point, we do not have found evidence that any observation has influenced the estimation of any coefficient of the $\mathcal{ZMPS}$ regression model significantly.

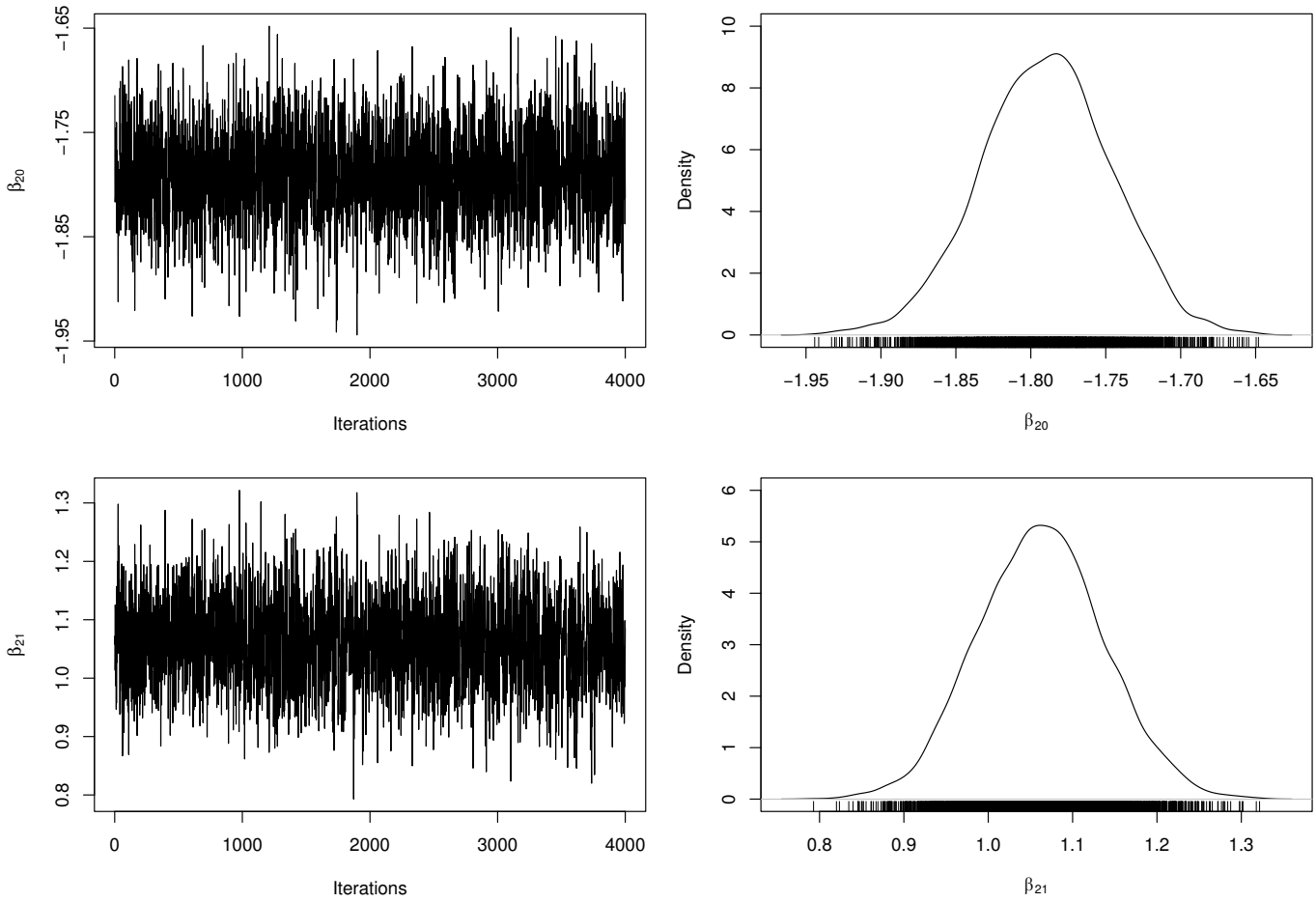
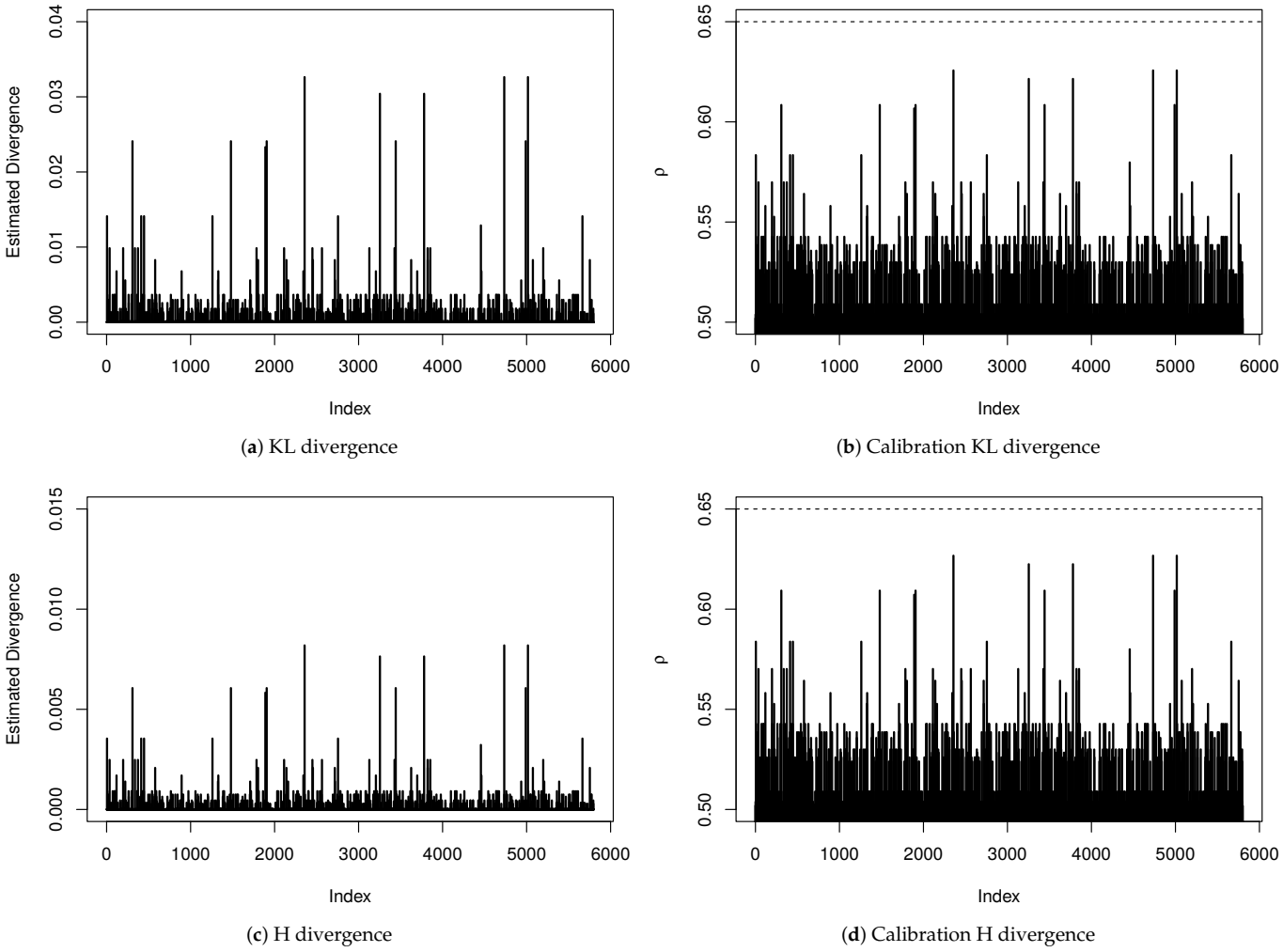


Figure 3. Trace plots and marginal posterior distributions of parameters β_{20} and β_{21} from the $\mathcal{ZMPS}$ regression model.

For comparison purposes, identical Bayesian procedures were adopted to fit the $\mathcal{P}$, the $\mathcal{NB}$, the $\mathcal{PS}$, the $\mathcal{ZMP}$ and the $\mathcal{ZMNB}$ regression models. To estimate the fixed dispersion parameter (ϕ) of $\mathcal{NB}$ and $\mathcal{ZMNB}$ models, we have considered a noninformative inverse-gamma prior distribution with hyperparameters $a = b = 1.0$. For each fitted model, we have estimated the measures presented in Appendix C. The model comparison procedure is summarized in Table 4. One can notice that the zero-modified models have performed considerably better with $\mathcal{ZMPS}$ outperforming all. These results are highlighting that the proposed model is highly competitive with well-established models in the literature. This feature can be considered one of the most relevant achievements of the $\mathcal{ZMPS}$ model since it has to deal with the positive observations using fewer parameters than, for example, the $\mathcal{ZMNB}$ model.

In Table 4, we have also reported the Bayesian p-values as a way to evaluate the adequacy of the fitted models. As expected, the $\mathcal{P}$ model is unsuitable to describe the considered dataset, and the fit provided by the $\mathcal{NB}$ regression model is also highly questionable. For the zero-modified models, there is no indication of overall lack-of-fit, since the posterior values of p_{B} were estimated close to 0.50. Figure 5 depicts additional evidence based on the RQRs for validating the fitted $\mathcal{ZMPS}$ regression. This residual metric was computed as discussed in Appendix D, using Equation (4). One can notice that the normality assumption of the residuals is easily verified by the behavior of its frequency distribution (left-panel). Additionally, the half-normal probability plot indicates that the fit of the $\mathcal{ZMPS}$ model was very satisfactory since all estimated residuals are lying within the simulated envelope (right-panel).



(**a**) KL divergence

(**b**) Calibration KL divergence

(**c**) H divergence

(**d**) Calibration H divergence

Figure 4. Sensitivity analysis for diagnosis of influential points.

Table 4. Comparison criteria and adequacy measures for the fitted models.

Model	DIC	EAIC	EBIC	NLMPL	p_B
$\mathcal{P}$	4650.631	4652.624	4665.955	2325.750	1.000
$\mathcal{NB}$	4340.938	4343.915	4363.912	2170.321	0.936
$\mathcal{PS}$	4436.313	4438.312	4451.643	2218.355	0.578
$\mathcal{ZMP}$	4323.300	4327.164	4353.826	2161.734	0.516
$\mathcal{ZMNB}$	4321.668	4326.960	4360.288	2160.530	0.598
$\mathcal{ZMPS}$	4320.539	4324.549	4351.212	2160.138	0.542

From the results displayed in Table 3, one can make some conclusions. Firstly, we have observed that the HPDIs of parameters β_{11} and β_{21} do not contain the value zero, which constitutes the dose of ionizing radiation as a relevant covariate to describe the average number of chromosomal aberrations as well the probability of not observing at least one aberration (p_0). For example, the expected number of dicentrics and centric rings in a cell that was exposed to 1.0 Gy is 0.363, and the probability of such aberrations not to occur is $\hat{p}_0 = \Phi(1.790 - 0.319) = 0.929$. Therefore, based on the posterior estimates, the components of the fitted $\mathcal{ZMPS}$ model can be expressed by

$$\hat{\mu}_{ij} = \exp\{-1.481 + 0.935x_i\} \quad \text{and} \quad \hat{\omega}_{ij} = \Phi(-1.790 + 1.062x_i),$$

where x_i is the dose of ionizing radiation.

Figure 6 present the Bayesian estimates, by dose, for the probability of not observing at least one aberration (left-panel) and for parameter p (right-panel). Noticeably, inferences about parameter p confirm the initial assumption that the analyzed sample has an excessive amount of zeros.

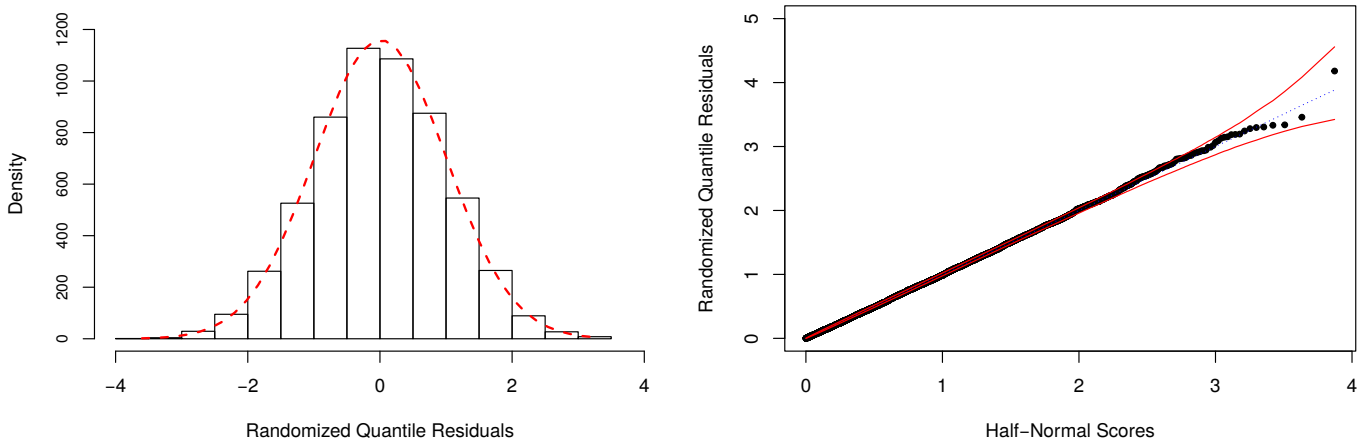


Figure 5. Frequency distribution and half-normal plot with simulated envelope for the randomized quantile residuals (RQRs).

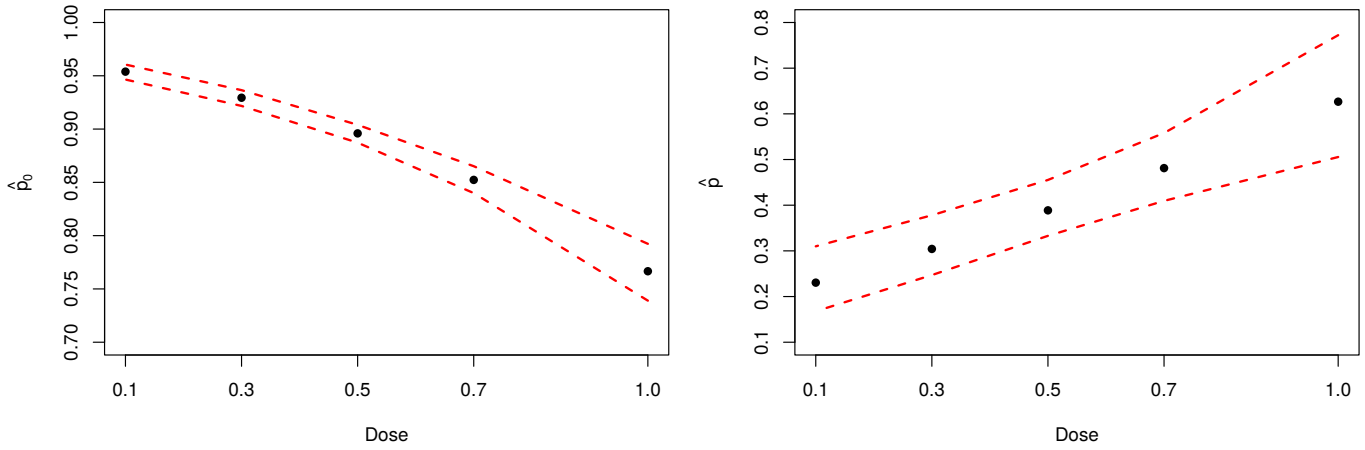


Figure 6. Posterior estimates of parameters p_0 and p. The dashed red lines represent the 95% HPDIs.

Table 5 presents a general posterior summary of the models that were fitted to the chromosomal aberration data. Here, parameter λ as estimated as $n^{-1}\sum_{i=1}^{5}\sum_{j=1}^{n_i}\hat{\lambda}_{ij}$ and ς^2 was estimated analogously. One can notice that the expected number of zeros $(\hat{n}_0)$ obtained by the $\mathcal{P}$, the $\mathcal{NB}$ and the $\mathcal{PS}$ models are slightly lower than the observed n_0, while those provided by the zero-modified models are very close (or exactly equal) to 5252. Through these measures, one can better understand how the fitted models are adhering to the data since the nature of the observed counts should be well described regarding its overdispersion level and the frequency and the average number of nonzero observations.

The goodness-of-fit of the fitted models can be evaluated by the χ^2 statistic obtained from the observed and expected frequencies. To compute such measure, we have grouped cells with frequencies lower or equal than 5, resulting in 4 degrees of freedom. The obtained statistics are also presented in Table 5. Figure 7 depict the positive expected frequencies (left-panel) and the dose-response curves (right-panel) that were estimated by the zero-modified models. Noticeably, the zero-modified models describe much better the data's behavior, especially the $\mathcal{ZMNB}$ and the $\mathcal{ZMPS}$ distributions.

Table 5. Posterior parameter estimates and goodness-of-fit evaluation.

Model	Parameter	$\hat{\lambda}$	$\hat{\varsigma}^2$	$\hat{n}_0$	χ^2	p-Value
$\mathcal{P}$	$\hat{\beta}_{10} = -2.97$ $\hat{\beta}_{11} = 1.95$	0.131	0.131	5086	2343.773	<0.001
$\mathcal{NB}$	$\hat{\beta}_{10} = -3.02$ $\hat{\beta}_{11} = 2.07$ $\hat{\phi} = 0.28$	0.133	0.232	5202	20.050	<0.001
$\mathcal{PS}$	$\hat{\beta}_{10} = -2.99$ $\hat{\beta}_{11} = 1.98$	0.132	0.157	5126	266.458	<0.001
$\mathcal{ZMP}$	$\hat{\beta}_{10} = -0.86$ $\hat{\beta}_{11} = 0.82$ $\hat{\beta}_{20} = -1.79$ $\hat{\beta}_{21} = 1.06$	0.132	0.199	5251	16.456	0.002
$\mathcal{ZMNB}$	$\hat{\beta}_{10} = -1.33$ $\hat{\beta}_{11} = 0.88$ $\hat{\beta}_{20} = -1.79$ $\hat{\beta}_{21} = 1.07$ $\hat{\phi} = 1.51$	0.131	0.206	5251	7.255	0.123
$\mathcal{ZMPS}$	Table 3	0.132	0.210	5252	5.298	0.258

From the obtained results, one can conclude that despite the suitable fit provided by the $\mathcal{ZMNB}$ regression model, the proposed model have adhered better to the chromosomal aberration data. This achievement can be regarded as extremely relevant since the $\mathcal{ZMNB}$ model has an additional (dispersion) parameter to handle the non-zero observations. In contrast, the proposed model was proved highly competitive by its ability to accommodate the data overdispersion and zero modification using fewer parameters.

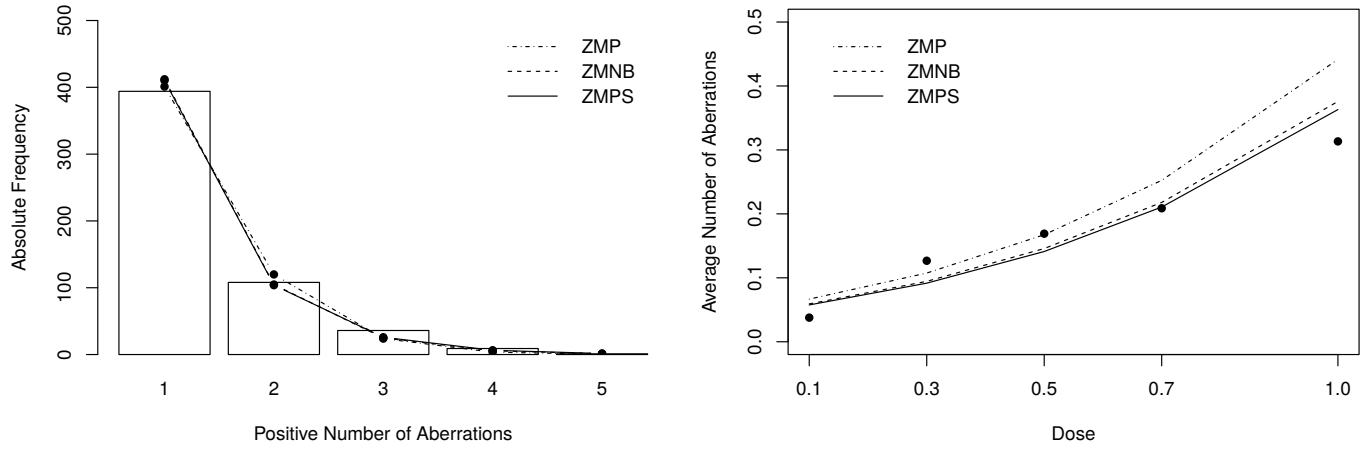


Figure 7. Posterior expected frequencies and dose-response curve fitted by the zero-modified models.

6. Concluding Remarks

This work aimed to introduce the $\mathcal{ZMPS}$ regression model as an alternative for the analysis of overdispersed datasets exhibiting zero-modification in the presence of covariates. Intensive Monte Carlo simulation studies were performed, and the obtained results have allowed us to assess the empirical properties of the Bayesian estimators and then conclude about the suitability of the adopted methodology to the predefined scenarios. The proposed model was considered for analyzing a real dataset on the number of cytogenetic

chromosomal aberrations, considering the dose of ionizing radiation as the covariate for both model components. The response variable was identified as overdispersed and heavily zero-inflated, which justified using the $\mathcal{ZMPS}$ regression model. The main conclusion one can make from the fitted models is that the dose is statistically relevant to describe either the probability of occurrence and the average incidence of aberrations. Besides, when looking at the χ^2 statistic and the posterior-based comparison criteria, we have noticed that the proposed model has presented a better fit when compared to its competitors and therefore, it can be considered an excellent addition to the set of models that can be used for the analysis of overdispersed and zero-modified count data.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/e23060646/s1.

Author Contributions: All authors equally contributed to developing this work. All authors have read and agreed to the published version of the manuscript.

Funding: The research of Wesley Bertoli is supported by the Federal University of Technology—Paraná. The research of Francisco Louzada is supported by the National Council for Scientific and Technological Development (CNPq— Grant: 301976/2017-1) and by the São Paulo Research Foundation (FAPESP— Grant: 2013/07375-0). The research of Katiane S. Conceição and Marinho G. Andrade is supported by FAPESP (Grants: 2019/22412-5 and 2019/21766-8). This work was supported by FAPESP (Grant: 2021/00407-0).

Data Availability Statement: The dataset analyzed in this work was made available in the Supplementary Material.

Acknowledgments: We would like to thank the associate editor and the four anonymous referees for their careful reading and thoughtful suggestions, which certainly improved this work's content.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANCP	Above noncoverage probability
B	Bias
BNCP	Below noncoverage probability
CDF	Cumulative distribution function
CP	Coverage probability
CPO	Conditional predictive ordinate
CS	Chi-square
DIC	Deviance information criterion
EAIC	Expected Akaike information criterion
EBIC	Expected Bayesian information criterion
ESS	Effective sample size
GLM	Generalized linear model
H	Hellinger
HPDI	Highest posterior density interval
J	Jeffrey
KL	Kullback–Leibler
L^1	Variational divergence
LMPL	Log-marginal pseudo-likelihood
MAPE	Mean absolute percentage error
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MSE	Mean squared error
$\mathcal{NB}$	Negative binomial
NI	Number of iterations

$\mathcal{P}$	Poisson
PMF	Probability mass function
$\mathcal{PS}$	Poisson–Sujatha
RQR	Randomized quantile residuals
RwM	Random-walk metropolis
$\mathcal{ZIP}$	Zero-inflated Poisson
$\mathcal{ZMP}$	Zero-modified Poisson
$\mathcal{ZMPS}$	Zero-modified Poisson–Sujatha
$\mathcal{ZTP}$	Zero-truncated Poisson
$\mathcal{ZTPS}$	Zero-truncated Poisson–Sujatha

Appendix A. Algorithms

Appendix A.1. Random-Walk Metropolis

Algorithm A1 Random-walk metropolis.

procedure RWM($N, \boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}$)
  Set $t \leftarrow 1$ and $\nu \leftarrow n(n+1)^{-1}$
  while $t \leqslant N$ **do**
    Generate $\boldsymbol{\psi}_1 \sim \mathcal{N}_{\overline{q}_1}[\nu\boldsymbol{\beta}_1^{(t-1)}, \nu\mathcal{S}_1^{(t-1)}]$ and $\boldsymbol{\psi}_2 \sim \mathcal{N}_{\overline{q}_2}[\nu\boldsymbol{\beta}_2^{(t-1)}, \nu\mathcal{S}_2^{(t-1)}]$
    Set $\alpha_1 \leftarrow \exp\{\pi_1(\boldsymbol{\psi}_1; \boldsymbol{y}) - \pi_1(\boldsymbol{\beta}_1^{(t-1)}; \boldsymbol{y})\}$
    Set $\alpha_2 \leftarrow \exp\{\pi_2(\boldsymbol{\psi}_2; \boldsymbol{y}) - \pi_2(\boldsymbol{\beta}_2^{(t-1)}; \boldsymbol{y})\}$
    Set $\boldsymbol{\beta}_1^{(t)} \leftarrow \boldsymbol{\beta}_1^{(t-1)}$ and $\boldsymbol{\beta}_2^{(t)} \leftarrow \boldsymbol{\beta}_2^{(t-1)}$
    Generate $u_1, u_2 \sim \mathcal{U}(0,1)$
    if $u_1 \leqslant \min\{1, \alpha_1\}$ and $u_2 \leqslant \min\{1, \alpha_2\}$ **then**
      Set $\boldsymbol{\beta}_1^{(t)} \leftarrow \boldsymbol{\psi}_1$ and $\boldsymbol{\beta}_2^{(t)} \leftarrow \boldsymbol{\psi}_2$
    end if
    Set $t \leftarrow t+1$
  end while
  return $\{\boldsymbol{\beta}^t\}_{t=1}^N = \{\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t\}_{t=1}^N$
end procedure

Appendix A.2. Sequential Search

Algorithm A2 Sequential search.

procedure SEQSEA($\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}$)
  Generate $x, u \sim \mathcal{U}(0,1)$
  Set $\mu \leftarrow \exp\{\beta_{10} + \beta_{11}x\}$ and $\omega \leftarrow g_2^{-1}(\beta_{20} + \beta_{21}x)$
  Set $k \leftarrow (1-\omega)$ and $y \leftarrow 0$
  while $u > k$ **do**
    Set $y \leftarrow y+1$ and $k \leftarrow k + \omega\, \mathrm{P}_*(Y = y; \mu)$
  end while
  return y
end procedure

Appendix B. Influential Points

Identifying influential observations is a crucial step in any statistical analysis. Usually, the presence of influential points impacts the inferential procedures and the subsequent conclusions considerably. In this way, this subsection is dedicated to present some case deletion Bayesian diagnostic measures that can be used to quantify the influence of observations from each subject in a given dataset.

The computation of divergence measures between posterior distributions is a useful way to quantify influence. According [66], the φ-divergence measure between two densities f and g for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is defined by

$$d_\varphi = \int_{\boldsymbol{\Theta}} g(\boldsymbol{\theta}) \varphi\left[\frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\right] \mathrm{d}\boldsymbol{\theta},$$

where φ is a smooth convex, lower semicontinuous function such that $\varphi(1) = 0$. Some popular divergence measures can be obtained by choosing specific functions for φ. The well-known Kullback–Leibler (KL) divergence is obtained by considering $\varphi(z) = -\log(z)$. A symmetric version of the KL divergence, the Jeffrey (J) divergence, can be obtained by specifying $\varphi(z) = (z-1)\log(z)$ and the variational divergence (L^1 norm) is obtained when $\varphi(z) = 0.50|z-1|$. In addition, the Chi-square (CS) divergence is obtained by considering $\varphi(z) = (z-1)^2$ and the Hellinger (H) distance arises when $\varphi(z) = 0.50(\sqrt{z}-1)^2$. We refer to [67] for a detailed study on several types of φ-divergence.

Let $g(\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}; \boldsymbol{y}_i)$ be the joint posterior distribution of $\boldsymbol{\beta}$ based only on the i-th observation and let $f(\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}; \boldsymbol{y}_i^j)$, where $\boldsymbol{y}_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ is the response vector without the i-th observation. After some algebra (see [68] for the KL divergence case), one can verify that the φ-divergence corresponds to

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

where $\mathbb{E}_{\boldsymbol{\beta}}[\mathrm{P}_*(Y_i = y_i; \boldsymbol{\beta})^{-1}; \boldsymbol{y}]^{-1}$ is the conditional predictive ordinate (CPO) statistic [69] for the i-th observation. Here, we are also not able to compute the inner expectation over $\boldsymbol{\beta}$ analytically and so, an MC estimator for the CPO_i is given by

$$\widehat{\mathrm{CPO}}_i = \left[\frac{1}{M} \sum_{t=1}^{M} \mathrm{P}_*\left(Y_i = y_i; \boldsymbol{\beta}^{(t)}\right)^{-1} \right]^{-1}. \tag{A1}$$

According to [70], the harmonic mean estimator (A1) is stable when most of the individual log-likelihood values exceed -10. Using the estimated CPO, one can approximate the local influence of a particular y_i on the joint posterior distribution (12) as

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

One can notice that, if $\pi(\boldsymbol{\beta}; \boldsymbol{y}_{-i}) = \pi(\boldsymbol{\beta}; \boldsymbol{y})$, then there is no divergence caused by observation y_i. In practice, however, it may not be elementary to define a threshold value for the divergence to decide about the magnitude of the influence [71]. A measure of calibration for the KL divergence was proposed by [72]. The idea is based on the typical toy binary example of tossing a coin once and observing its upper face. This experiment can be described by $\mathrm{P}(Y = y; \rho) = \rho^y (1-\rho)^{1-y}$, $y \in \{0, 1\}$, where $\rho \in [0, 1]$ is the probability of success. Regardless of what success means, if the coin is unbiased, then $\mathrm{P}(Y = y; \rho) = 0.50$. Thus, the φ-divergence between a (possibly) biased and an unbiased coin is given by

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

from which one can conclude that the divergence between two posteriors distributions can be associated with the biasedness of a coin [67]. By analogy, this implies that predict unobserved responses using $\pi(\boldsymbol{\beta}; \boldsymbol{y}_{-i})$ instead of $\pi(\boldsymbol{\beta}; \boldsymbol{y})$ is equivalent to describe an unobserved event as having probability ρ_i, when the correct probability is 0.50. Considering some specific choices for φ, in Table A1 we present MC estimators that can be used to

compute the local influence of each y_i. Besides, we also present the expression of $d_\varphi(\rho)$ for each φ. For ease of notation, we assume $f_i^t = \mathrm{P}(Y_i = y_i; \boldsymbol{\beta}^{(t)})$.

Table A1. MC estimators for some φ-divergence measures and their calibration.

$\boldsymbol{\varphi}$	$\hat{\boldsymbol{d}}_{\boldsymbol{\varphi}}$	$\boldsymbol{d_\varphi(\rho)}$	$\hat{\boldsymbol{\rho}}_{\boldsymbol{\varphi}}$
KL_i	$\frac{1}{M}\sum_{t=1}^{M}\log(f_i^t) - \log\left(\widehat{\mathrm{CPO}}_i\right)$	$-\frac{1}{2}\log[4\rho_i(1-\rho_i)]$	$\frac{1}{2}\left[1+\sqrt{1-e^{-2\hat{d}_i}}\right]$
J_i	$\frac{1}{M}\sum_{t=1}^{M}\left(\frac{\widehat{\mathrm{CPO}}_i}{f_i^t}-1\right)\log\left(\frac{\widehat{\mathrm{CPO}}_i}{f_i^t}\right)$	$-\frac{(1-2\rho_i)}{2}\log\left(\frac{\rho_i}{1-\rho_i}\right)$	no closed-form
L_i^1	$\frac{1}{2M}\sum_{t=1}^{M}\frac{\left\lvert\widehat{\mathrm{CPO}}_i - f_i^t\right\rvert}{f_i^t}$	$\frac{1}{2}\lvert 1-2\rho_i\rvert$	$\frac{1}{2}+\hat{d}_i$
CS_i	$\frac{1}{M}\sum_{t=1}^{M}\frac{\left(\widehat{\mathrm{CPO}}_i - f_i^t\right)^2}{(f_i^t)^2}$	$(1-2\rho_i)^2$	$\frac{1}{2}\left[1+\sqrt{\hat{d}_i}\right]$
H_i	$\frac{1}{2M}\sum_{t=1}^{M}\left(\sqrt{\frac{\widehat{\mathrm{CPO}}_i}{f_i^t}}-1\right)^2$	$1-\frac{1}{\sqrt{2}}\left(\sqrt{\rho_i}+\sqrt{1-\rho_i}\right)$	$\frac{1}{2}+\left[\sqrt{\hat{d}_i}-\sqrt{\hat{d}_i^3}\right]\sqrt{2-\hat{d}_i}$

KL: K; J: Jeffrey; L^1: Variational; CS: Chi-Square; and H: Hellinger.

The function $d_\varphi(\rho)$ is symmetric about 0.50 and increases as ρ moves away from 0.50. In addition, $\inf_{\rho\in(0,1)} d_\varphi(\rho) = 0$, which is attained at $\rho = 0.50$ since $d_\varphi(0.50) = \varphi(1) = 0$. Therefore, a general measure of calibration based on the φ-divergence can be obtained by solving

$$2d_\varphi(\rho) - \varphi(2\rho) - \varphi[2(1-\rho)] = 0.$$

An estimator for the calibration measure (ρ_φ) associated with each φ-divergence type is also presented in Table A1. Clearly, depending on the form of φ, such an equation may not have a closed-form, which is the case of the J divergence. Besides, one can notice that $\rho_i \in [0.50, 1]$ and so, for $\rho_i \gg 0.50$, the i-th observation may be considered an influential point. For example, if $\rho_i > 0.80$ is considered a significative bias, then y_i will be classified as influential if $\hat{d}_i > 0.223$ $(d_\varphi(0.80) \approx 0.223)$ under the KL divergence or yet if $\hat{d}_i > 0.051$ $(d_\varphi(0.80) \approx 0.051)$ under the H divergence.

Appendix C. Model Comparison and Adequacy

There are several techniques for Bayesian model selection that are useful to compare competing models. The most popular method is the deviance information criterion (DIC), which was proposed to work simultaneously to measure fit and complexity of the model. The DIC criterion is defined as

$$\mathrm{DIC} = \mathbb{E}_{\boldsymbol{\beta}}[\mathrm{D}(\boldsymbol{\beta})] + \varrho_{\mathrm{D}} = \underline{\mathrm{D}}(\boldsymbol{\beta}) + \varrho_{\mathrm{D}},$$

where $\mathrm{D}(\boldsymbol{\beta}) = -2\ell(\boldsymbol{\beta}; \boldsymbol{y})$ is the deviance function and $\varrho_{\mathrm{D}} = \underline{\mathrm{D}}(\boldsymbol{\beta}) - \mathrm{D}(\hat{\boldsymbol{\beta}})$ is the effective number of model parameters, with $\hat{\boldsymbol{\beta}}$ given by (14). A negative value for ϱ_{D} may suggest that the log-likelihood function is non-concave, the prior distribution is misspecified, or the posterior expected value is not a good estimator for $\boldsymbol{\beta}$. On the other hand, when $\varrho_{\mathrm{D}} \gg d$, then there is an indication of overfitting with estimate $\hat{\boldsymbol{\beta}}$.

Noticeably, we are not able to compute the expectation of $\mathrm{D}(\boldsymbol{\beta})$ over $\boldsymbol{\beta}$ analytically. In this case, an MC estimator for such a measure is given by

$$\hat{\underline{\mathrm{D}}}(\boldsymbol{\beta}) = -\frac{2}{M}\sum_{t=1}^{M}\ell\left(\boldsymbol{\beta}^{(t)}; \boldsymbol{y}\right),$$

and so the DIC can be estimated by $\widehat{\mathrm{DIC}} = 2\hat{\underline{\mathrm{D}}}(\boldsymbol{\beta}) - \mathrm{D}(\hat{\boldsymbol{\beta}})$.

The expected Akaike (EAIC) and the expected Bayesian (EBIC) information criteria can also be used when comparing Bayesian models [73,74]. Using the approximation for the expected value of $\mathrm{D}(\boldsymbol{\beta})$, these measures can be estimated by

$$\widehat{\mathrm{EAIC}} = \hat{\underline{\mathrm{D}}}(\boldsymbol{\beta}) + 2d \quad \text{and} \quad \widehat{\mathrm{EBIC}} = \hat{\underline{\mathrm{D}}}(\boldsymbol{\beta}) + d\log(n).$$

Another widely used criterion is derived from the CPO statistic, which is based on the cross-validation criterion to compare models. For the i-th observation, the CPO can be estimated through Equation (A1). A summary statistic of the estimated CPO's is the log-marginal pseudo-likelihood (LMPL) given by the sum of the logarithms of $\widehat{\mathrm{CPO}}_i$'s. Regarding model comparison, we have that the lower the values of DIC, EAIC, EBIC, and NLMPL (negative LMPL), the better the fit.

In addition to comparing, researchers are often interested in verifying the adequacy of the fitted models. An effective way to evaluate model suitability is based on the use of measures derived from the ppd. For instance, if any observation is extremely unlikely relative to the ppd, the obtained fit's adequacy might be questionable. Ref. [75] proposed a widespread discrepancy measure between model and data. In our case, we need a slightly adapted version of such a measure, which is given by

$$\mathrm{T}(\boldsymbol{y}, \boldsymbol{\beta}) = -2\sum_{i=1}^{n} \log[\mathrm{P}_*(Y_i = y_i; \boldsymbol{\beta})].$$

The Bayesian p-value (posterior predictive p-value), proposed by [76], is defined as

$$p_{\mathrm{B}} = \mathrm{P}[\mathrm{T}(\boldsymbol{y}_r, \boldsymbol{\beta}) \geqslant \mathrm{T}(\boldsymbol{y}, \boldsymbol{\beta}); \boldsymbol{y}],$$

where $\boldsymbol{y}_r$ denotes the response vector that might have been observed if the conditions generating $\boldsymbol{y}$ were reproduced. This predictive measure can be empirically estimated as the relative number of times that $\mathrm{T}(\boldsymbol{y}_r, \hat{\boldsymbol{\beta}})$ exceeds $\mathrm{T}(\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ out of B simulations. In general, the model fit becomes suspect if the discrepancy is of practical relevance, and the associated Bayesian p-value is close either to 0 or 1 [75]. A large (small) value of p_{B}, say greater than 0.95 (lower than 0.05), indicates model misspecification (lack-of-fit), that is, the observed behavior would be unlikely to be seen if we replicate the response vector using the fitted model.

Appendix D. Residual Analysis

The residual analysis plays an essential role in the task of validating the results obtained from a regression model. In general, residual metrics are responsible for indicating departures from the underlying model assumptions by quantifying the portion of data variability that the fitted model is not explaining. Assessing a regression model's adequacy using residual metrics is a common practice nowadays due to the availability of statistical packages providing diagnostic tools for well-established models. However, deriving appropriate residuals is not always an easy task for non-normal models that accommodate overdispersion. In this way, we will consider a popular residual metric proposed by [77], the randomized quantile residuals (RQRs), which can be straightforwardly used in our context to assess the appropriateness of the proposed model when fitted to real data.

For obvious reasons, we focus on the definition of RQRs for discrete random variables. In this case, the RQR associated with the i-th observation is defined as $r_i = \Phi^{-1}(u_i)$, where Φ denotes the cdf of the standard normal distribution and u_i is a Uniform random variable defined on $(a_i, b_i]$, with $a_i = \lim_{y \uparrow y_i} \mathrm{F}(y_i)$ and $b_i = \mathrm{F}(y_i)$, where $\mathrm{F}(y_i)$ is the cdf of the current model. In our case, we may obtain an MC estimator for the RQR as $\hat{r}_i = \Phi^{-1}(u_i)$, with $u_i \sim \mathcal{U}(\lim_{y \uparrow y_i} \hat{\mathrm{F}}_*(y_i), \hat{\mathrm{F}}_*(y_i)]$. Here, $\hat{\mathrm{F}}_*(y_i) \equiv \mathrm{F}_*(y_i; \hat{\mu}_i, \hat{\omega}_i)$ is an estimate for the probability of $Y_i \leqslant y_i$ using cdf (4), where $\hat{\mu}_i$ and $\hat{\omega}_i$ depend on the fitted model as $\hat{\mu}_i = \log(\boldsymbol{x}_{1i}^{\intercal}\hat{\boldsymbol{\beta}}_1)$ and $\hat{\omega}_i = g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\hat{\boldsymbol{\beta}}_2)$.

The primary assumption for this metric is that $\hat{r}_i \sim \mathcal{N}(0,1)$ must hold, whichever the variability degree of $\hat{\mu}_i$ and $\hat{\omega}_i$. In this case, after model fitting, one has to evaluate if these residuals are normally distributed around zero, which can be made through adherence tests and by using graphical techniques as histograms and half-normal probability plots. An excellent alternative for checking whether RQRs are consistent with the fitted model is the inclusion of simulated envelopes in their half-normal plot. Thus, if a significant subset of estimated residuals falls outside the envelope bands, then the fitted model's adequacy must be questioned, and further investigation on the corresponding observations is necessary. Ref. [78] provides an algorithm for obtaining simulated envelopes for a half-normal plot.

References

1. Karlis, D.; Xekalaki, E. Mixed Poisson distributions. *Int. Stat. Rev.* **2005**, *73*, 35–58. [CrossRef]
2. Sankaran, M. The discrete Poisson-Lindley distribution. *Biometrics* **1970**, *26*, 145–149. [CrossRef]
3. Bulmer, M.G. On fitting the Poisson-Lognormal distribution to species-abundance data. *Biometrics* **1974**, *30*, 101–110. [CrossRef]
4. Shaban, S.A. On the discrete Poisson-Inverse Gaussian distribution. *Biom. J.* **1981**, *23*, 297–303. [CrossRef]
5. Zamani, H.; Ismail, N. Negative Binomial-Lindley distribution and its application. *J. Math. Stat.* **2010**, *6*, 4–9. [CrossRef]
6. Shanker, R.; Sharma, S.; Shanker, U.; Shanker, R.; Leonida, T.A. The discrete Poisson-Janardan distribution with applications. *Int. J. Soft Comput. Eng.* **2014**, *4*, 31–33.
7. Shanker, R.; Mishra, A. A two parameter Poisson-Lindley distribution. *Int. J. Stat. Syst.* **2014**, *9*, 79–85.
8. Shanker, R. The discrete Poisson-Amarendra distribution. *Int. J. Stat. Distrib. Appl.* **2016**, *2*, 14–21. [CrossRef]
9. Shanker, R. The discrete Poisson-Shanker distribution. *Jacobs J. Biostat.* **2016**, *1*, 1–7.
10. Shanker, R. The discrete Poisson-Sujatha distribution. *Int. J. Probab. Stat.* **2016**, *5*, 1–9.
11. Shanker, R.; Mishra, A. A quasi Poisson-Lindley distribution. *J. Indian Stat. Assoc.* **2016**, *54*, 113–125.
12. Bakouch, H.S. A Weighted Negative Binomial-Lindley distribution with applications to dispersed data. *An. Acad. Bras. Ciências* **2018**, *90*, 2617–2642. [CrossRef]
13. Shanker, R.; Shukla, K.K. On Poisson-weighted Lindley distribution and its applications. *J. Sci. Res.* **2018**, *11*, 1–13. [CrossRef]
14. Kuş, C.; Akdoğan, Y.; Asgharzadeh, A.; Kınacı, İ.; Karakaya, K. Binomial-discrete Lindley distribution. *Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.* **2019**, *10*, 401–411. [CrossRef]
15. Shanker, R.; Shukla, K.K.; Leonida, T.A. A two-parameter Poisson-Sujatha distribution. *Am. J. Math. Stat.* **2020**, *68*, 70–78.
16. Mullahy, J. Specification and testing of some modified count data models. *J. Econom.* **1986**, *91*, 841–853. [CrossRef]
17. Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **1992**, *34*, 1–14. [CrossRef]
18. Zorn, C.J.W. Evaluating zero-inflated and hurdle Poisson specifications. *Midwest Political Sci. Assoc.* **1996**, *18*, 1–16.
19. Deb, P.; Trivedi, P.K. The structure of demand for health care: Latent class versus two-part models. *J. Health Econ.* **2002**, *21*, 601–625. [CrossRef]
20. Angers, J.F.; Biswas, A. A Bayesian analysis of zero-inflated generalized Poisson model. *Comput. Stat. Data Anal.* **2003**, *42*, 37–46. [CrossRef]
21. McDowell, A. From the help desk: Hurdle models. *Stata J.* **2003**, *3*, 178–184. [CrossRef]
22. Wagh, Y.S.; Kamalja, K.K. Zero-inflated models and estimation in zero-inflated Poisson distribution. *Commun. Stat. Simul. Comput.* **2018**, *47*, 1–18. [CrossRef]
23. Gurmu, S.; Trivedi, P.K. Excess zeros in count models for recreational trips. *J. Bus. Econ. Stat.* **1996**, *14*, 469–477.
24. Bohara, A.K.; Krieg, R.G. A zero-inflated Poisson model of migration frequency. *Int. Reg. Sci. Rev.* **1996**, *19*, 211–222. [CrossRef]
25. Ridout, M.; Demétrio, C.G.B.; Hinde, J. Models for count data with many zeros. In Proceedings of the XIXth International Biometric Conference, Cape Town, South Africa, 13–18 December 1998; Volume 19, pp. 179–192.
26. Bahn, G.D.; Massenburg, R. Deal with excess zeros in the discrete dependent variable, the number of homicide in Chicago census tract. In Proceedings of the Joint Statistical Meetings of the American Statistical Association, Denver, CO, USA, 3–7 August 2008; pp. 3905–3912.
27. Mouatassim, Y.; Ezzahid, E.H. Poisson regression and zero-inflated Poisson regression: Application to private health insurance data. *Eur. Actuar. J.* **2012**, *2*, 187–204. [CrossRef]
28. Heilbron, D.C.; Gibson, D.R. Shared needle use and health beliefs concerning AIDS: Regression modeling of zero-heavy count data. Poster session. In Proceedings of the Sixth International Conference on AIDS, San Francisco, CA, USA, 20–24 June 1990.
29. Hu, M.C.; Pavlicova, M.; Nunes, E.V. Zero-inflated and hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. *Am. J. Drug Alcohol Abus.* **2011**, *37*, 367–375. [CrossRef]
30. Ngatchou-Wandji, J.; Paris, C. On the zero-inflated count models with application to modelling annual trends in incidences of some occupational allergic diseases in France. *J. Data Sci.* **2011**, *9*, 639–659.
31. Beuf, K.D.; Schrijver, J.D.; Thas, O.; Criekinge, W.V.; Irizarry, R.A.; Clement, L. Improved base-calling and quality scores for 454 sequencings based on a hurdle Poisson model. *BMC Bioinform.* **2012**, *13*, 303. [CrossRef]

32. Oliveira, M.; Einbeck, J.; Higueras, M.; Ainsbury, E.; Puig, P.; Rothkamm, K. Zero-inflated regression models for radiation-induced chromosome aberration data: A comparative study. *Biom. J.* **2016**, *58*, 259–279. [CrossRef]
33. Rodrigues, J. Bayesian analysis of zero-inflated distributions. *Commun. Stat. Theory Methods* **2003**, *32*, 281–289. [CrossRef]
34. Ghosh, S.K.; Mukhopadhyay, P.; Lu, J.C. Bayesian analysis of zero-inflated regression models. *J. Stat. Plan. Inference* **2006**, *136*, 1360–1375. [CrossRef]
35. Dietz, E.; Böhning, D. On estimation of the Poisson parameter in zero-modified Poisson models. *Comput. Stat. Data Anal.* **2000**, *34*, 441–459. [CrossRef]
36. Conceição, K.S.; Andrade, M.G.; Louzada, F. Zero-modified Poisson model: Bayesian approach, influence diagnostics, and an application to a Brazilian leptospirosis notification data. *Biom. J.* **2013**, *55*, 661–678. [CrossRef] [PubMed]
37. Conceição, K.S.; Andrade, M.G.; Louzada, F. On the zero-modified Poisson model: Bayesian analysis and *posterior* divergence measure. *Comput. Stat.* **2014**, *29*, 959–980. [CrossRef]
38. Conceição, K.S.; Louzada, F.; Andrade, M.G.; Helou, E.S. Zero-modified Power Series distribution and its hurdle distribution version. *J. Stat. Comput. Simul.* **2017**, *87*, 1842–1862. [CrossRef]
39. Conceição, K.S.; Suzuki, A.K.; Andrade, M.G.; Louzada, F. A Bayesian approach for a zero modified Poisson model to predict match outcomes applied to the 2012-13 La Liga season. *Braz. J. Probab. Stat.* **2017**, *31*, 746–764. [CrossRef]
40. Bertoli, W.; Conceição, K.S.; Andrade, M.G.; Louzada, F. On the zero-modified Poisson-Shanker regression model and its application to fetal deaths notification data. *Comput. Stat.* **2018**, *33*, 807–836. [CrossRef]
41. Bertoli, W.; Conceição, K.S.; Andrade, M.G.; Louzada, F. Bayesian approach for the zero-modified Poisson-Lindley regression model. *Braz. J. Probab. Stat.* **2019**, *33*, 826–860. [CrossRef]
42. Bertoli, W.; Ribeiro, A.M.T.; Conceição, K.S.; Andrade, M.G.; Louzada, F. On zero-modified Poisson-Sujatha distribution to model overdispersed count data. *Austrian J. Stat.* **2018**, *47*, 1–19.
43. Bertoli, W.; Conceição, K.S; Andrade, M.G; Louzada, F. A Bayesian approach for some zero-modified Poisson mixture models. *Stat. Model.* **2020**, *20*, 467–501. [CrossRef]
44. Shanker, R.; Fesshaye, H. On zero-truncation of Poisson, Poisson-Lindley and Poisson-Sujatha distributions and their applications. *Biom. Biostat. Int. J.* **2016**, *3*, 1–13. [CrossRef]
45. Fernández-Fontelo, A.; Puig, P.; Ainsbury, E.A.; Higueras, M. An exact goodness-of-fit test based on the occupancy problems to study zero-inflation and zero-deflation in biological dosimetry data. *Radiat. Prot. Dosim.* **2018**, *179*, 317–326. [CrossRef] [PubMed]
46. Li, C.; Wang, D.; Sun, J. Control charts based on dependent count data with deflation or inflation of zeros. *J. Stat. Comput. Simul.* **2019**, *89*, 3273–3289. [CrossRef]
47. Zellner, A. On assessing *prior* distributions and Bayesian regression analysis with *g-prior* distributions. *Bayesian Inference Decis. Tech. Essays Honor Bruno De Finetti* **1986**, *6*, 233–243.
48. Bazán, J.L.; Torres-Avilés, F.; Suzuki, A.K.; Louzada, F. Power and reversal power links for binary regressions: An application for motor insurance policyholders. *Appl. Stoch. Model. Bus. Ind.* **2017**, *33*, 22–34. [CrossRef]
49. Heilbron, D.C. Zero-altered and other regression models for count data with added zeros. *Biom. J.* **1994**, *36*, 531–547. [CrossRef]
50. Ghosh, S.K.; Kim, H. Semiparametric inference based on a class of zero-altered distributions. *Stat. Methodol.* **2007**, *4*, 371–383. [CrossRef]
51. Chen, M.H.; Ibrahim, J.G. Conjugate *priors* for generalized linear models. *Stat. Sin.* **2003**, *30*, 461–476.
52. Gupta, M.; Ibrahim, J.G. An information matrix prior for Bayesian analysis in generalized linear models with high dimensional data. *Stat. Sin.* **2009**, *19*, 1641–1663.
53. Bové, D.S.; Held, L. Hyper-*g* priors for generalized linear models. *Bayesian Anal.* **2011**, *6*, 387–410.
54. Kass, R.E.; Wasserman, L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Am. Stat. Assoc.* **1995**, *90*, 928–934. [CrossRef]
55. Hansen, M.H.; Yu, B. Minimum description length model selection criteria for generalized linear models. *Lect. Notes Monogr. Ser.* **2003**, *40*, 145–163.
56. Wang, X.; George, E.I. Adaptive Bayesian criteria in variable selection for generalized linear models. *Stat. Sin.* **2007**, *17*, 667–690.
57. Marin, J.M.; Robert, C. *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*; Springer Texts in Statistics: New York, NY, USA, 2007.
58. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [CrossRef]
59. Roberts, G.O.; Gelman, A.; Gilks, W.R. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **1997**, *7*, 110–120. [CrossRef]
60. Heidelberger, P.; Welch, P.D. Simulation run length control in the presence of an initial transient. *Oper. Res.* **1983**, *31*, 1109–1144. [CrossRef]
61. Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of *posterior* moments. *Bayesian Stat.* **1992**, *4*, 641–649.
62. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434–455.
63. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

64. Hörmann, W.; Leydold, J.; Derflinger, G. *Automatic Nonuniform Random Variate Generation*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
65. Heimers, A.; Brede, H.J.; Giesen, U.; Hoffmann, W. Chromosome aberration analysis and the influence of mitotic delay after simulated partial-body exposure with high doses of sparsely and densely ionising radiation. *Radiat. Environ. Biophys.* **2006**, *45*, 45–54. [CrossRef]
66. Csiszár, I. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.* **1967**, *2*, 299–318.
67. Peng, F.; Dey, D.K. Bayesian analysis of outlier problems using divergence measures. *Can. J. Stat.* **1995**, *23*, 199–213. [CrossRef]
68. Cho, H.; Ibrahim, J.G.; Sinha, D.; Zhu, H. Bayesian case influence diagnostics for survival models. *Biometrics* **2009**, *65*, 116–124. [CrossRef] [PubMed]
69. Geisser, S.; Eddy, W.F. A predictive approach to model selection. *J. Am. Stat. Assoc.* **1979**, *74*, 153–160. [CrossRef]
70. Congdon, P. *Bayesian Models for Categorical Data*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
71. Weiss, R. An approach to Bayesian sensitivity analysis. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 739–750. [CrossRef]
72. McCulloch, R.E. Local model influence. *J. Am. Stat. Assoc.* **1989**, *84*, 473–478. [CrossRef]
73. Brooks, S.P. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2002**, *64*, 616–639.
74. Carlin, B.P.; Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2001.
75. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*; Chapman & Hall/CRC Texts in Statistical Science; CRC Press: Boca Raton, FL, USA, 2004.
76. Rubin, D.B. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Stat.* **1984**, *12*, 1151–1172. [CrossRef]
77. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.
78. Moral, R.A.; Hinde, J.; Demétrio, C.G.B. Half-Normal plots and overdispersed models in R: The `hnp` package. *J. Stat. Softw.* **2017**, *81*, 1–23. [CrossRef]