Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: www.tandfonline.com/journals/cjas20

A doubly-inflated Poisson regression for correlated count data

Erfan Ghasemi, Alireza Akbarzadeh Baghban, Farid Zayeri, Asma Pourhoseingholi & Seyed Mohammadreza Safavi

To cite this article: Erfan Ghasemi, Alireza Akbarzadeh Baghban, Farid Zayeri, Asma Pourhoseingholi & Seyed Mohammadreza Safavi (2021) A doubly-inflated Poisson regression for correlated count data, Journal of Applied Statistics, 48:6, 1111-1127, DOI: 10.1080/02664763.2020.1757049

To link to this article: https://doi.org/10.1080/02664763.2020.1757049

Published online: 01 May 2020.

JOURNAL OF APPLIED STATISTICS
2021, VOL. 48, NO. 6, 1111–1127
https://doi.org/10.1080/02664763.2020.1757049

Taylor & Francis
Taylor & Francis Group

APPLICATION NOTE

A doubly-inflated Poisson regression for correlated count data

Erfan Ghasemi[a], Alireza Akbarzadeh Baghban[b], Farid Zayeri[c], Asma Pourhoseingholi[a] and Seyed Mohammadreza Safavi[d]

[a]Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [b]Proteomics Research Center, School of Rehabilitation, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [c]Proteomics Research Center and Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [d]Research Institute of Dental Sciences, Dental Research Center, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

Count data have emerged in many applied research areas. In recent years, there has been a considerable interest in models for count data. In modelling such data, it is common to face a large frequency of zeroes. The data are regarded as zero-inflated when the frequency of observed zeroes is larger than what is expected from a theoretical distribution such as Poisson distribution, as a standard model for analysing count data. Data analysis, using the simple Poisson model, may lead to over-dispersion. Several classes of different mixture models were proposed for handling zero-inflated data. But they do not apply to cases when inflated counts happen at some other points, in addition to zero. In these cases, a doubly-inflated Poisson model has been suggested which only be used for cross-sectional data and cannot consider correlations between observations. However, correlated count data have a large application, especially in the health and medical fields. The present study aims to introduce a Doubly-Inflated Poisson models with random effect for correlated doubly-inflated data. Then, the best performance of the proposed method is shown via different simulation scenarios. Finally, the proposed model is applied to a dental study.

ARTICLE HISTORY
Received 9 August 2018
Accepted 11 April 2020

KEYWORDS
Count data; doubly-inflated; Poisson regression; zero-inflated; correlated data

1. Introduction

Count data have attracted a lot of attention in many applied research areas, such as industry, economy, health, biology and dental epidemiology [2,15,18]. We may encounter a large frequency of zeroes in modelling such data. The data are called ‘zero-inflated’ when the real number of zeroes is larger than what is predicted from a Poisson model, which is considered as a standard model for analysing count data. The misleading results could happen in the inferences due to ‘over-dispersion’ if the zero inflation is not taken into consideration [20]. In other words, the variability in the data is larger than what is expected under an ordinary Poisson distribution.

CONTACT Alireza Akbarzadeh Baghban akbarzad@sbmu.ac.ir Proteomics Research Center, School of Rehabilitation, Shahid Beheshti University of Medical Sciences, Across Bu-Ali Hospital, Damavand Street, Tehran, Iran

Several classes of different mixture models, including ‘zero-inflated models’ and hurdle models, were proposed for handling zero-inflated data [6,17,25]. The hurdle models encompass a mixture of a point mass at zero and a truncated count distribution for the non-zero values [14]. On the other hand, zero-inflated models comprise a mixture of a point mass at zero and a count distribution. There are different kinds of zero-inflated models, such as Zero-Inflated Poisson (ZIP), Zero-Inflated Negative Binomial (ZINB) and Zero-Inflated Binomial (ZIB) [16,27,31].

Although the proposed models can overcome the accumulation problem at zero, they do not apply to cases where inflated counts happen at some other point ($k > 0$), in addition to zero. For example, in a study on ear infection of toddlers in a period of six months, the frequencies of 0 and 2 were high. The large frequency of zeroes showed the large number of toddlers who were never infected in their ears. On the other hand, a high probability of a relapse of infection at a later time was an interesting point of this study [5]. As another example, data from the Australian survey of 1977–1978 showed the count of the number of doctor visits. In these data, again the frequencies of 0 and 1 count were inflated. The reason was that many patients did not visit their doctors and many of them visited once [3]. In these cases, for obtaining more accurate results, a doubly-inflated model can be used instead of Poisson or zero-inflated Poisson regressions. This model can address not only the inflated frequencies of zero but also the inflated frequencies of k.

Therefore, designing a model based on both inflations is essential. Chandra proposed a Doubly-Inflated Poisson (DIP) model for cross-sectional data in order to handle the inflation in two points (zero and $k > 0$) [24]. Sengupta et al have proposed a bivariate doubly-inflated Poisson model in 2016 [23] and then in another study have improved it by using copula function in 2017 [22].

Unfortunately, these models can only be used for cross-sectional data and cannot consider correlations between observations. While, in most of studies, the data may be repeated in several occasions or they may use a clustering form, leading to a correlation structure in dataset [32]. The correlation between the data has the information to be considered and ignoring this correlation can introduce substantial bias in the estimates of regression coefficient variances that can affect inferences [4,7,8,9,19]. For example, in Iranian National Oral Health Survey done on 12-year-old children in 2012, the total number of decay, missed and filled teeth (DMFT) were recorded. In this study, oral area is divided into four areas, so that maxilla and mandible are divided to right and left sides. There was inflation at zero and one. According to the correlation between DMFT in different oral areas, a model for handling this correlation is interested.

Hall added a random effect in the zero-inflated model which has taken the correlation structure into consideration [12].

By considering all the above-mentioned problems, the aim of this study is to present a doubly-inflated Poisson model which incorporates over-dispersion and doubly-inflated and correlated count data. The proposed model was first studied on simulated data and then used to analyse the dental data. The organization of the paper is as follows. Section 2.1 introduces a zero-inflated Poisson model for the correlated data, along with estimation and inference. Section 2.2 is related to doubly-inflated Poisson model with random effect. The proposed method based on real data and the simulation study to compare the results of the proposed model with the current models is discussed in Section 3. Finally, conclusion and recommendation are explained in Section 4.

2. Methodology

2.1. ZIP model with a random effect

Suppose that y_{ij} refers to the response for subjects $i = 1, \ldots, n$ at the cluster $j = 1, \ldots, T_i$. The distribution of y_{ij} given the random effect b_i is as follows:

$$y_{ij} \sim \begin{cases} 0 & \text{with probability } \psi_{ij} \\ \text{poisson}(\mu_{ij}) & \text{with probability } 1 - \psi_{ij} \end{cases}$$

where ψ_{ij} represents the probability of the observation arising from the degenerate distribution at zero and μ_{ij} denotes the mean of Poisson distribution. ψ and μ are regarded as the parameters of the model which are modelled via logit and log link functions, respectively:

$$\text{logit}(\psi_{ij}) = Z'_{ij}\gamma$$

$$\log(\mu_{ij}) = X'_{ij}\beta + \sigma b_i$$

where Z_{ij} and X_{ij} are the vectors of covariates in zero and the Poisson part of the model, respectively. $b_1, \cdots, b_n \overset{iid}{\sim} N(0,1)$ is the vector of random effects and y_{ij} is conditionally independent given b_i. The log likelihood for each parameter is given below:

$$l(\gamma, \beta, \sigma; y) = \sum_{i=1}^{n} log \left\{ \int_{-\infty}^{+\infty} \left[\prod_{j=1}^{T_i} P(Y_{ij} = y_{ij}|b_i) \right] \phi(b_i) db_i \right\}$$

where

$$P(Y_{ij} = y_{ij}|b_i) = (1 + e^{Z'_{ij}\gamma})^{-1} \left\{ u_{ij}[\exp(Z'_{ij}\gamma) + exp(-exp(X'_{ij}\beta + \sigma b_i))] + (1 - u_{ij}) \times \frac{exp[y_{ij}(X'_{ij}\beta + \sigma b_i) - exp(X'_{ij}\beta + \sigma b_i)]}{y_{ij}!} \right\}$$

and $u_{ij} = 1$ if $y_{ij} = 0$ and $u_{ij} = 0$ otherwise and $\phi()$ is the probability density function of standard normal distribution. Finally, the model is fitted by using maximum likelihood via the EM algorithm.

2.2. DIP model

2.2.1. DIP distribution

Zero inflation models are the common models for analysing count data when the zero inflation takes place. Recently, these models have attracted a lot of attention. However, in some special cases, the inflation happens both in zero and other points (k > 0). An unobserved variable is used to model such data. The variable distribution is as follows:

$$\text{P}(W = w) = \begin{cases} p_1 & \text{for} \quad w = 2 \\ p_2 & \text{for} \quad w = 1 \\ p_3 & \text{for} \quad w = 0 \end{cases}$$

where $p_3 = 1 - p_1 - p_2$.

Using the above unobserved variable, the DIP distribution for y_{ij} is illustrated as

$$P(Y_{ij} = y_{ij}|W = w) = \begin{cases} 1 & \text{for } w = 2, \quad y_{ij} = 0 \\ 1 & \text{for } w = 1, \quad y_{ij} = k \\ \dfrac{exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!} & \text{for } w = 0, \quad y_{ij} = 0, 1, 2, \cdots \end{cases}$$

(p_1, p_2, μ_{ij}) represent the parameters of the DIP model and the marginal probability function of y_{ij} is as follows:

$$P(Y_{ij} = y_{ij}) = \begin{cases} p_1 + p_3 \exp(-\mu_{ij}) & \text{for } y_{ij} = 0 \\ p_2 + p_3 \dfrac{\exp(-\mu_{ij})\mu_{ij}^{k}}{k!} & \text{for } y_{ij} = k \\ p_3 \dfrac{\exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!} & \text{for } y_{ij} = 1, 2, \cdots \neq k \end{cases}$$

where $(p_3 = 1 - p_1 - p_2)$ and $E(Y) = p_2 k + p_3 \mu$, $Var(Y) = p_2 k^2 + p_3(\mu^2 + \mu) - (E(Y))^2$ demonstrate the expectation and variance, respectively[13]. When $p_2 \to 0$, the DIP distribution tends to be the ZIP model.

Deciding on k practically needs to explore the data in the first step and then it needs to evaluate inflation at this point using a proper hypothesis test. In this context, several hypothesis tests were suggested for evaluating inflation at a point in a parametric distribution. Some of these tests have just proposed to evaluate zero-inflation, but the others can be used at any point for any non-degenerate parametric distributions[13,26,28,30]. In this study, a score test (suggested by Todem *et al.*) was used to evaluate inflation (heterogeneity at a point in the Poisson distribution) at point k [7]. Using this test, one can adjust the covariates effect which improves the efficiency.

In the next section, the regression form of the DIP model with a random effect term will be introduced to show systematic and random components and link function of the proposed model for the correlated data.

2.2.2. DIP model with a random effect

Let y_{ij} is the discrete response variable for $(i = 1, \ldots, n \,\text{and}\, j = 1, \ldots, T_i)$. y_{ij} has DIP (p_1, p_2, μ_{ij}) distribution and there is a correlation structure between the response variables. Thus, a DIP model with a random effect is proposed to control the effect of correlation structure as follows:

$$\log\left(\frac{p_{1ij}}{p_{3ij}}\right) = A'_{ij}\alpha,$$

$$\log\left(\frac{p_{2ij}}{p_{3ij}}\right) = Z'_{ij}\gamma,$$

$$\log(\mu_{ij}) = X'_{ij}\beta + \sigma b_i$$

where $b_1, \cdots, b_n \overset{iid}{\sim} N(0, 1)$ are considered as the random effects and A_{ij}, Z_{ij} and X_{ij} represent the design matrices in multinomial and Poisson models, respectively. The random effects can be added in each three parts of the model. In mixed models with random intercept, the covariance of within subjects is σ^2 and the correlation equals $\frac{\sigma^2}{\sigma^2+\sigma_\varepsilon{}^2}$ which $\sigma_\varepsilon{}^2$ is the variance of errors [20].

Therefore, the log likelihood is represented below:

$$l(\alpha,\gamma,\beta,\sigma;y)=\sum_{i=1}^{n}\log\left\{\int_{-\infty}^{+\infty}\left[\prod_{j=1}^{T_i}P(Y_{ij}=y_{ij}|b_i)\right]\phi(b_i)db_i\right\} \tag{1}$$

where

$$P(Y_{ij}=y_{ij}|b_i)=[p_{1ij}+p_{3ij}\exp(-\mu_{ij})]^{u_{1ij}}\times\left[p_{2ij}+p_{3ij}\frac{\exp(-\mu_{ij})\mu_{ij}{}^{k}}{k!}\right]^{u_{2ij}}$$
$$\times\left[p_{3ij}\frac{\exp(-\mu_{ij})\mu_{ij}{}^{y_{ij}}}{y_{ij}!}\right]^{1-u_{1ij}-u_{2ij}}$$

and, $u_{2ij}=\begin{cases}1, y_{ij}=k\\ 0, o.w\end{cases}$ and $u_{1ij}=\begin{cases}1, y_{ij}=0\\ 0, o.w\end{cases}$ Represented the maximum likelihood via Gauss-Hermite quadrature use for fitting the model.

Gauss-Hermite quadrature is a numerical approximation method which approximates an integral given by

$$\int_{-\infty}^{+\infty}f(x)e^{-x^2}dx\approx\sum_{q=1}^{Q}f(x_q)\times w_q \tag{2}$$

where x_q and w_q are quadrature node and corresponding weight, respectively. x_q and w_q can obtain from 'fastGHQuad' R package[1] or tables of mathematical books [33].

We considered $b_i=\sqrt{2}b_i^*$ so b_i^* has $\frac{1}{\sqrt{\pi}}exp(-b_i^{*2})$ as probability density function and $\log(\mu_{ij})$ equals $X_{ij}'\beta+\sigma\sqrt{2}b_i^*$. By implementing Gauss-Hermite quadrature method in log-likelihood function, the approximation of $l(\alpha,\gamma,\beta,\sigma;y)$ from Equation (1) is given by

$$l(\alpha,\gamma,\beta,\sigma;y)\approx\sum_{i=1}^{n}\log\left\{\frac{1}{\sqrt{\pi}}\sum_{q=1}^{Q}w_q\times\left[\prod_{j=1}^{T_i}P(Y_{ij}=y_{ij}|b_i^*)\right]\right\}$$

The approximation of log-likelihood function was maximized by Newton–Raphson method.

3. Applications

3.1. Real data

The models were implemented on 'The Iranian National Oral Survey (INOHS) (2012)' in which doubly inflated occurred. The data included the total numbers of decayed, missed and filled tooth (DMFT) among 12-year-old children as the participants in INOHS. A self-report questionnaire and oral examination were used for data collection. The study aimed

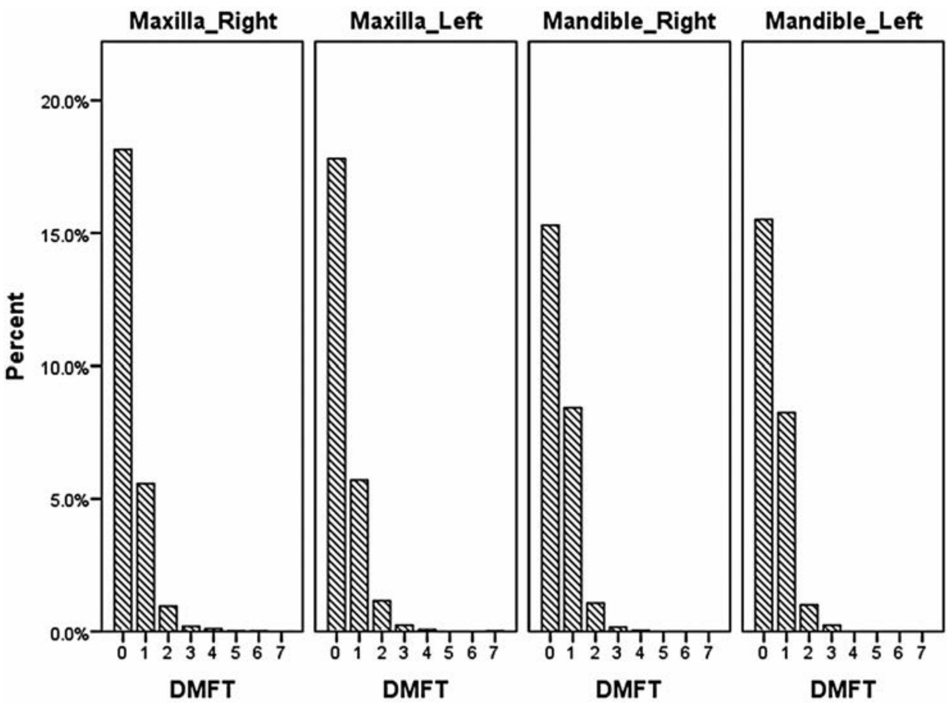

Figure 1. Percentage frequency of DMFT in four sides of oral area.

Table 1. Model fit comparison: −2log-likelihood, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC).

Method	$-2\times$log-likelihood	AIC	BIC
ZIP[a] with random effect in part zero	8358.9	8374.9	8416.6
ZIP with random effect in Poisson part	8182.7	8198.7	8240.5
DIP[b] with random effect in part zero	7747.6	7771.6	7834.2
DIP with random effect in part one	8235.9	8259.9	8322.5
DIP with random effect in Poisson part	8042.4	8066.4	8129.1
DIP without random effect (Fixed effect)	8828.3	8850.3	8923.0

[a]Zero-Inflated Poisson.
[b]Doubly-Inflated Poisson.

to investigate the effect of social-economic characteristics and oral health habits on DMFT. The maxilla and mandible were divided into two right and left parts. The DMFT was calculated in all four areas and an inflation took place in zero and one values. Figure 1 illustrates the DMFT frequencies in the four areas.

The independent variables included sex (female/male), Social-Economic Status score (SES Score) and Sugar Score. A score test was conducted on the data and the results showed that the inflation exists at points 0 (statistic $= 247.11$, df $= 4$, $p < 0.001$) and 1 (statistic $= 16.05$, df $= 4$, $p = 0.003$).

The doubly-inflated Poisson with random effects model was fitted to the dental data. The random effect term was used in different parts of the DIP and ZIP models. The Akaike Information Criteria (AIC) and the Bayesian Information Criteria are shown in Table 1 and the results of the model with the lowest AIC are shown in Table 2.

Table 2. Estimated regression coefficient of the mixed-effects DIP model for INOHS[a] data.

variable	Part zero of model			Part one of model			Poisson part of model		
	α	SE	*P*-value	γ	SE	*P*-value	β	SE	*P*-value
Intercept	1.239	0.460	0.007	−0.452	0.401	0.260	−0.301	0.215	0.162
Sex	0.154	0.283	0. 586	0.171	0.214	0.422	0.165	0.110	0.133
SES[b] Score	0.552	0.141	< 0.001	0.204	0.104	0.050	0.088	0.055	0.109
Sugar Score[c]	–	–	–	0.436	0.167	0.009	0.108	0.107	0.316
σ^2	7.88	0.760	< 0.001	–	–	–	–	–	–
		$R^2{}_{\mathrm{GLMM(c)}} = 0.86$				MSE = 0.948			

[a]Iranian National Oral Health Survey.
[b]Socio-Economic Status.
[c]The coefficient was omitted from part zero because of the converge problem.

As shown in Table 2, the effect of SES score was significant in part zero of model and Sugar score has a significant effect in part one of model. This means that the odds of caries-free in higher socioeconomic levels are more than lower socioeconomic levels. Also the conditional R-square proposed by Nakagawa and Schielzeth for generalized linear mixed models [21] (i.e. $R^2{}_{\mathrm{GLMM(c)}}$) was calculated as the goodness-of-fit statistic and it is reported in Table 2.

3.2. Simulation study

3.2.1. Method and design

In order to investigate the properties of the proposed model, our approach was explained based on simulation studies. As for the validity of the proposed method, three general simulation scenarios were presented, based on Poisson, ZIP and DIP models. In each context, two samples in the size of 50 and 400 were generated with 1000 replications. In addition, $p_1 = 0.3$ was considered as the probability for zero inflation probability in ZIP when $p_2 = 0$, and $p_1 = 0.2$ and $p_2 = 0.3$ for the doubly-inflated in zero and one were regarded for the DIP model. In addition to investigate the influence of p_1 and p_2, we conducted some simulation studies and results are shown in appendix.

First, random variables (X_1) were generated from binary distribution with 0.4 probability of success. Second, another set of random variables (X_2) were independently generated from discrete rectangular distribution in the interval (1, 20). The third variables in our simulated model were the time variable and three times were considered for the purpose of this study ($X_3 = 1, 2, 3$).

Two different sets of the true value including $\beta_0 = 1$, $\beta_1 = 0.2$, $\beta_2 = 0.05$, $\beta_3 = 0.3$ and $\beta_0 = 1$, $\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 0.6$ were emphasized in this study.

On the other hand, for generating a multivariate Poisson distribution, we have used a method that was proposed by Yahav and Shmueli [29] we have generated samples in the four following steps:

1. We have considered the mean of Poisson distribution as follows:

$$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$$

2. We have generated the three variables from multivariate normal with

$$\mathrm{MVN}\left(\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}1 & 0.4 & 0.4\\0.4 & 1 & 0.4\\0.4 & 0.4 & 1\end{bmatrix}\right)$$

3. The cumulative probability of each variable under the normal standard distribution was investigated.
4. In the last step, the quantile of Poisson (μ) distribution according to the cumulative probability of each variable computed in the previous step was considered.

Finally, the generated variables are the multivariate Poisson (μ) with the covariance structure that was considered.

For generating multivariate ZIP and DIP, we have used the same method. For generating the inflation (zero and 1) we have used uniform distribution.

To show the accuracy of our model, use relative bias and the root of mean square error (RMSE) between the estimates and their true values where

$$\mathrm{Re}l.Bias=\frac{1}{\theta}\left(\frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_j\right)-1,\ RMSE=\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{\theta}_j-\theta)^2}.$$

In addition, in order to compare the validity of the two models DIP-Mixed and DIP-Fixed, another scenario was defined. By taking into account the exchangeable correlation structure as given below, the samples in the size 200 were generated from multivariate DIP Distribution with 1000 replications.

$$G=\begin{bmatrix}1 & \rho & \rho\\ \rho & 1 & \rho\\ \rho & \rho & 1\end{bmatrix}$$

where ρ was considered equal to 0.7, 0.4, 0.1 in data simulation. The accuracy of the two models was compared using the relative bias and coverage probability (CP). Coverage probability is the proportion of times that the 95% confidence interval of the estimated mean contains the true value. It is desirable to have coverage near 95%.

3.2.2. Results of simulation study

In the first scenario, we generated the samples from the DIP model. In this model, the inflation took place in zero and one. The real β, the estimated mean of coefficient, standard error, relative bias and RMSE are shown in Table 3. As shown, the relative bias and RMSE in the DIP model are less than the relative bias and RMSE in other two models. Furthermore, an increase in the sample size leads the RMSE to be zero, which indicates the consistency of the estimation coefficient of this model. In addition, the negative amount for relative bias in ZIP and Poisson models proves the underestimation of these two models. On the other hand, bigger coefficients lead to more accurate results. As indicated in Table 3, the RMSE of the DIP model had the lowest amount.

In the second scenario, the sample from the ZIP model was generated. Probability of $p_1 = 0.3$ was considered for zero inflation. Based on the results in Table 4, both DIP and

Table 3. Estimated regression coefficients for the DIP, ZIP and Poisson models with random effects for simulated data from DIP distribution.

		Mixed-Effect Doubly-inflated Poisson Model				Mixed-Effect Zero-Inflated Poisson Model				Mixed-Effect Poisson Model			
						$n = 400$							
	Real	Est.	S.E.	Rel. Bias*	RMSE*	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.4695	0.1976	−0.5305	0.5661	0.2806	0.1130	−0.7194	0.7283	0.2503	0.1095	−0.7497	0.7577
$\beta 1$	**0.2**	0.1937	0.0934	−0.0315	0.0936	0.1630	0.0869	−0.1849	0.0944	0.1558	0.0859	−0.2209	0.0965
$\beta 2$	**−0.05**	−0.0484	0.0082	−0.0328	0.0084	−0.0406	0.0072	−0.1885	0.0119	−0.0391	0.0071	−0.2189	0.0130
$\beta 3$	**0.3**	0.2932	0.0276	−0.0228	0.0284	0.2617	0.0271	−0.1278	0.0469	0.2555	0.0277	−0.1484	0.0524
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.9638	0.1018	−0.0362	0.1080	0.1943	0.1923	−0.8057	0.8284	0.0935	0.1863	−0.9065	0.9254
$\beta 1$	**0.4**	0.4056	0.0664	0.0139	0.0666	0.3249	0.1618	−0.1878	0.1783	0.3056	0.1537	−0.2360	0.1803
$\beta 2$	**−0.1**	−0.1025	0.0068	0.0249	0.0072	−0.0815	0.0128	−0.1848	0.0225	−0.0767	0.0125	−0.2335	0.0265
$\beta 3$	**0.6**	0.6062	0.0312	0.0104	0.0318	0.5414	0.0408	−0.0977	0.0714	0.5311	0.0449	−0.1148	0.0822
						$n = 50$							
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.6039	0.4138	−0.3961	0.5727	0.2769	0.3072	−0.7231	0.7856	0.2425	0.3009	−0.7575	0.8150
$\beta 1$	**0.2**	0.1889	0.2641	−0.0557	0.2642	0.1648	0.2609	−0.1761	0.2632	0.1591	0.2552	−0.2043	0.2583
$\beta 2$	**−0.05**	−0.0503	0.0208	0.0062	0.0208	−0.0399	0.0206	−0.2022	0.0229	−0.0382	0.0200	−0.2351	0.0232
$\beta 3$	**0.3**	0.3013	0.0767	0.0044	0.0767	0.2635	0.0772	−0.1215	0.0854	0.2568	0.0784	−0.1441	0.0895
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.9542	0.2652	−0.0458	0.2690	0.1862	0.3877	−0.8138	0.9014	0.0797	0.3735	−0.9203	0.9932
$\beta 1$	**0.4**	0.3956	0.1716	−0.0109	0.1715	0.3324	0.3104	−0.1690	0.3175	0.3147	0.3056	−0.2133	0.3171
$\beta 2$	**−0.1**	−0.1057	0.0169	0.0569	0.0179	−0.0820	0.0261	−0.1796	0.0317	−0.0770	0.0258	−0.2300	0.0346
$\beta 3$	**0.6**	0.6066	0.0645	0.0110	0.0648	0.5428	0.0879	−0.0954	0.1048	0.5319	0.0951	−0.1134	0.1169

Note: *Rel. Bias: Relative Bias, RMSE: Root of Mean Square Error.
Inflation values are generated with $p_{1i} = 0.2$ and $p_{2i} = 0.3$ for zero and one, respectively.

Table 4. Estimated regression coefficients for the DIP, ZIP and Poisson models with random effects for simulated data from ZIP distribution.

		Mixed-Effect Doubly-inflated Poisson Model				Mixed-Effect Zero-Inflated Poisson Model				Mixed-Effect Poisson Model			
		$n = 400$											
C	**Real**	C Est.	C S.E.	C Rel. Bias*	C RMSE*	C Est.	C S.E.	C Rel. Bias	C RMSE	C Est.	C S.E.	C Rel. Bias	C RMSE
$\beta 0$	**1**	0.9295	0.0711	−0.0705	0.1001	0.9278	0.0695	−0.0722	0.1002	0.3583	0.1196	−0.6417	0.6527
$\beta 1$	**0.2**	0.2110	0.0522	0.0551	0.0534	0.2118	0.0521	0.0590	0.0534	0.2015	0.0935	0.0077	0.0934
$\beta 2$	**−0.05**	−0.0529	0.0044	0.0584	0.0053	−0.0529	0.0044	0.0575	0.0053	−0.0503	0.0079	0.0066	0.0079
$\beta 3$	**0.3**	0.3071	0.0200	0.0237	0.0212	0.3074	0.0198	0.0248	0.0211	0.2998	0.0265	−0.0006	0.0265
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.9723	0.0593	−0.0277	0.0654	0.9716	0.0586	−0.0284	0.0651	0.2904	0.1375	−0.7096	0.7228
$\beta 1$	**0.4**	0.4064	0.0419	0.0161	0.0423	0.4069	0.0411	0.0172	0.0417	0.3951	0.1076	−0.0122	0.1077
$\beta 2$	**−0.1**	−0.1016	0.0040	0.0159	0.0043	−0.1015	0.0037	0.0149	0.0040	−0.0987	0.0090	−0.0134	0.0091
$\beta 3$	**0.6**	0.6038	0.0173	0.0064	0.0177	0.6039	0.0172	0.0065	0.0176	0.5989	0.0256	−0.0018	0.0256
		$n = 50$											
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.9253	0.2522	−0.0747	0.2629	0.8956	0.2664	−0.1044	0.2860	0.3382	0.3652	−0.6618	0.7558
$\beta 1$	**0.2**	0.1974	0.1736	−0.0129	0.1736	0.2088	0.1826	0.0442	0.1828	0.2053	0.2900	0.0267	0.2899
$\beta 2$	**−0.05**	−0.0539	0.0154	0.0788	0.0159	−0.0536	0.0155	0.0713	0.0159	−0.0498	0.0234	−0.0032	0.0233
$\beta 3$	**0.3**	0.3006	0.0602	0.0020	0.0601	0.3052	0.0577	0.0175	0.0579	0.2978	0.0745	−0.0074	0.0745
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	1.0028	0.1724	0.0028	0.1723	0.9784	0.1716	−0.0216	0.1729	0.2770	0.4159	−0.7230	0.8340
$\beta 1$	**0.4**	0.3832	0.1295	−0.0420	0.1305	0.3977	0.1241	−0.0058	0.1241	0.3899	0.3367	−0.0251	0.3367
$\beta 2$	**−0.1**	−0.1026	0.0128	0.0264	0.0131	−0.1041	0.0123	0.0408	0.0130	−0.0992	0.0259	−0.0076	0.0259
$\beta 3$	**0.6**	0.5948	0.0568	−0.0087	0.0570	0.6055	0.0500	0.0092	0.0503	0.6017	0.0778	0.0029	0.0778

Note: *Rel. Bias: Relative Bias, RMSE: Root of Mean Square Error.
Zero inflation values are generated with $p_{1i} = 0.3$.

Table 5. Estimated regression coefficients for the DIP, ZIP and Poisson models with random effects for simulated data from Poisson distribution.

		Mixed-Effect Doubly-inflated Poisson Model				Mixed-Effect Zero-Inflated Poisson Model				Mixed-Effect Poisson Model			
		n = 400											
	Real	Est.	S.E.	Rel. Bias*	RMSE*	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	0.9643	0.0555	−0.0357	0.0659	0.9671	0.0549	−0.0329	0.0640	0.9672	0.0547	−0.0328	0.0638
$\beta1$	**0.2**	0.1932	0.0480	−0.0340	0.0485	0.2001	0.0416	0.0005	0.0416	0.2002	0.0415	0.0009	0.0415
$\beta2$	**−0.05**	−0.0507	0.0039	0.0133	0.0040	−0.0505	0.0036	0.0100	0.0036	−0.0505	0.0036	0.0102	0.0036
$\beta3$	**0.3**	0.3053	0.0244	0.0178	0.0250	0.2999	0.0154	−0.0004	0.0154	0.2998	0.0150	−0.0008	0.0150
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	0.9781	0.0494	−0.0219	0.0540	0.9793	0.0506	−0.0207	0.0547	0.9782	0.0493	−0.0218	0.0539
$\beta1$	**0.4**	0.4019	0.0354	0.0047	0.0355	0.4044	0.0381	0.0109	0.0383	0.4022	0.0348	0.0054	0.0349
$\beta2$	**−0.1**	−0.1002	0.0032	0.0023	0.0032	−0.1004	0.0036	0.0041	0.0036	−0.1002	0.0031	0.0018	0.0031
$\beta3$	**0.6**	0.6006	0.0139	0.0010	0.0139	0.6003	0.0134	0.0006	0.0134	0.6004	0.0132	0.0006	0.0132
		n = 50											
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	0.9704	0.1433	−0.0296	0.1463	0.9658	0.1476	−0.0342	0.1514	0.9631	0.1506	−0.0369	0.1550
$\beta1$	**0.2**	0.1829	0.1266	−0.0856	0.1277	0.1978	0.1268	−0.0110	0.1268	0.2026	0.1281	0.0132	0.1280
$\beta2$	**−0.05**	−0.0521	0.0115	0.0416	0.0117	−0.0507	0.0100	0.0149	0.0101	−0.0504	0.0100	0.0087	0.0100
$\beta3$	**0.3**	0.3132	0.0476	0.0439	0.0494	0.3000	0.0428	0.0001	0.0428	0.3003	0.0414	0.0010	0.0414
	Real	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	0.9979	0.1457	−0.0021	0.1457	0.9866	0.1405	−0.0134	0.1410	0.9809	0.1401	−0.0191	0.1414
$\beta1$	**0.4**	0.3912	0.1033	−0.0220	0.1036	0.3955	0.1021	−0.0113	0.1021	0.4007	0.1000	0.0018	0.1000
$\beta2$	**−0.1**	−0.1008	0.0105	0.0079	0.0105	−0.1011	0.0099	0.0112	0.0100	−0.1005	0.0091	0.0050	0.0091
$\beta3$	**0.6**	0.5976	0.0419	−0.0041	0.0419	0.6005	0.0392	0.0008	0.0392	0.6004	0.0382	0.0006	0.0382

*Note: Rel. Bias: Relative Bias, RMSE: Root of Mean Square Error.

Table 6. Estimated regression coefficients for mixed-effects DIP and fixed-effects DIP models for simulated data from DIP distribution with different correlations ($n = 200$).

		Mixed-Effect Doubly-Inflated Poisson Model				Fixed-Effect Doubly-Inflated Poisson Model			
		Exchangeable structure correlation with $\rho = 0.7$							
	Real	Est.	S.E.	Rel. Bias*	CP*	Est.	S.E.	Rel. Bias	CP
$\beta 0$	**1**	0.1938	0.2533	−0.8062	0.034	1.2096	0.1529	0.2096	0.515
$\beta 1$	**0.2**	0.1790	0.2139	−0.1052	0.856	0.1800	0.1125	−0.0998	0.751
$\beta 2$	**−0.05**	−0.0461	0.0172	−0.0774	0.839	−0.0461	0.0099	−0.0785	0.711
$\beta 3$	**0.3**	0.2819	0.0408	−0.0604	0.921	0.2751	0.0470	−0.0829	0.886
		Exchangeable structure correlation with $\rho = 0.4$							
	Real	Est.	S.E.	Rel. Bias.	CP	Est.	S.E.	Rel. Bias	CP
$\beta 0$	**1**	0.8647	0.2442	−0.1353	0.819	1.2049	0.1576	0.2049	0.532
$\beta 1$	**0.2**	0.2116	0.1313	0.0579	0.925	0.1882	0.1073	0.0591	0.775
$\beta 2$	**−0.05**	−0.0524	0.0105	0.0481	0.940	−0.0459	0.0087	−0.0826	0.758
$\beta 3$	**0.3**	0.2927	0.0517	−0.0243	0.9180	0.2739	0.0525	−0.0869	0.8280
		Exchangeable structure correlation with $\rho = 0.1$							
	Real	Est.	S.E.	Rel. Bias.	CP	Est.	S.E.	Rel. Bias	CP
$\beta 0$	**1**	1.0810	0.1499	0.0810	0.878	1.1651	0.1407	0.1651	0.650
$\beta 1$	**0.2**	0.1891	0.0923	−0.0544	0.942	0.1809	0.0906	−0.0954	0.845
$\beta 2$	**−0.05**	−0.0489	0.0076	−0.0227	0.944	−0.0466	0.0075	−0.0671	0.837
$\beta 3$	**0.3**	0.2901	0.0557	−0.0331	0.896	0.2817	0.0540	−0.0611	0.854

Note: *Rel. Bias: Relative Bias, CP: Coverage probability.

ZIP model are substantially similar. Therefore, we can conclude that the DIP model is regarded as an appropriate model for fitting with those data having no doubly inflation. The RMSE for ZIP and DIP models are less than that of the Poisson model. When $p_1 = 0.3$, the Poisson model includes less bias and bigger RMSE than the two other models.

In the third scenario, we generated the samples from the Poisson model. As illustrated in Table 5, the results of the Poisson and ZIP model are almost the same. Finally, the RMSE modesls are considerably similar when we deal with bigger β coefficient.

In the last scenario, we generated the samples from the DIP model with various correlation structures and the accuracy of two models (DIP-Mixed and DIP-Fixed) was compared. The relative bias in DIP-Mixed is almost similar to DIP-Fixed. However, in the mixed effects DIP model the CP is closer to 0.95 than the fixed effects DIP model. Coverage probability (CP) less than 0.95 indicates an inaccurate estimator (Table 6).

4. Conclusion and discussion

In the present study, we have introduced a new model for the correlated count data, when we are dealing with the data having inflation in two points. In this case, some other point (k) exhibits inflated frequencies, in addition to the inflated frequencies at zero. The classic Poisson and ZIP regression models are special cases of DIP regression models. Moreover, this model is regarded as a direct extension of the DIP model by Chandra [24], for adjusting the effect of correlation structure among the data.

For this purpose, four general simulation scenarios were used to compare the performance of the proposed method with two other current and simpler models (ZIP and Poisson model). In the first scenario, the correlated data were generated from the DIP model. The proposed DIP model had better performance than the ZIP and Poisson model. As for the second scenario, the data generation was done based on the ZIP model [12]. In this case, the performance of the ZIP and DIP model was substantially the same and better than the Poisson model. In other words, the DIP model can be still regarded as an appropriate model if inflation occurs just in one point (zero). The Poisson model is not a suitable model for analysing such inflated data. Thus, when doubly inflation can be observed in the available data, using a model that does not address doubly inflation may lead to less accurate results.

Regarding the third scenario, a Poisson model as a general liner model was used for generating the correlated data. The results of the DIP, ZIP and Poisson models were similar. Furthermore, the performance of the ZIP and DIP model as the same as the Poisson model in such situation.

Using the simple model can lead to misleading results in some specific situations such as zero or doubly-inflated data although it is more common among researchers. These misleading results can play a significant role in making the clinical decision and imposing major problems in the related studies. However, based on the results of the present study, we can have suitable and consistent results like those in the simpler model by selecting the proposed method in this study although the data were not really doubly-inflated. However, this may lead to loss of efficiency.

In the last simulation scenario, the correlated data were generated from the DIP model with different correlation structures and the accuracy of estimations of two models (DIP-mixed and DIP-fixed) was compared and two models were almost the same at estimation bias. But the mixed-effect model was more accurate than the fixed-effects model. So, when response variables are correlated, use fixed-effects doubly-inflated model may lead to inaccurate inferences.

Furthermore, our proposed model was implemented for real dental data as the pilot study. The results of the DIP model with random effect in one part showed that Sex, SES score and Sugar score had a significant effect on DMFT. The results were in congruent with those of other studies [10,11].

In this study for mathematical simplicity, normal distribution was used for random effects. As a recommendation for future studies, other distribution such as t for random effects may lead to interesting results. Furthermore, the DIP with three random effects in each separated parts did not converge in the present study. Finally, other studies are suggested to solve the converging problems.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] A.J.R.p.v.-. Blocker, URL http://CRAN.R-project.org/package...fastGHQuad, *fastGHQuad: Fast Rcpp implementation of Gauss-Hermite quadrature*, (2011).

[2] D. Böhning, E. Dietz, P. Schlattmann, L. Mendonca, and U. Kirchner, *The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology*. J Royal Stat Soc Ser. A 162 (1999), pp. 195–209. Available at http://onlinelibrary.wiley.com/doi/10.1111/1467-985X.00130/full.

[3] A.C. Cameron, P.K. Trivedi, F. Milne, and J. Piggott, *A microeconometric model of the demand for health care and health insurance in Australia*. Rev. Econ. Stud. 55 (1988), pp. 85–106.

[4] M.J. Cannon, L. Warner, J.A. Taddei, and D.G. Kleinbaum, *What can go wrong when you assume that correlated data are independent: an illustration from the evaluation of a childhood health intervention in Brazil*. Stat. Med. 20 (2001), pp. 1461–1467.

[5] S.A. Carlin, C.D. Marchant, P.A. Shurin, C.E. Johnson, D. Murdell-Panek, and S.J. Barenkamp, *Early recurrences of otitis media: reinfection or relapse?*[Erratum appears in J Pediatr 1987 Apr;110(4):668]. J. Pediatr. 110 (1987), pp. 20–25. Available at http://www.sciencedirect.com/science/article/pii/S0022347687802810.

[6] M.L. Dalrymple, I.L. Hudson, and R.P.K. Ford, *Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS*. Comput. Stat. Data. Anal. 41 (2003), pp. 491–504. Available at http://www.sciencedirect.com/science/article/pii/S0167947302001871.

[7] P. Diggle, P.J. Diggle, P. Heagerty, K.Y. Liang, P.J. Heagerty, and S. Zeger, *Analysis of Longitudinal Data*, ed, Oxford University Press, Oxford, 2002.

[8] G.M. Fitzmaurice, N.M. Laird, and J.H. Ware, *Applied Longitudinal Analysis*, John Wiley and Sons, Hoboken, NJ, 2004.

[9] T.P. Garcia, and K. Marder, *Statistical approaches to longitudinal data analysis in neurodegenerative diseases: huntington's disease as a model*. Curr. Neurol. Neurosci. Rep. 17 (2017), pp. 14–23.

[10] Z. Ghorbani, A.E. Ahmady, E. Ghasemi, and A. Zwi, *Socioeconomic inequalities in oral health among adults in Tehran, Iran*. Community Dent. Health 32 (2015), pp. 26–31. Available at http://ovidsp.ovid.com/ovidweb.cgi?T = JS&PAGE = reference&D = medl&NEWS = N&AN = 26263589.

[11] Z. Ghorbani, and K.G. Peres, *Is the association between socioeconomic status and nonreplaced extracted teeth mediated by dental care behaviours in adults?* Community Dent. Oral Epidemiol. 43 (2015), pp. 532–539. Available at https://scholar.google.com/citations?user = WBRdPGQAAAAJ&hl = en&oi = ao.

[12] D.B. Hall, *Zero-inflated Poisson and binomial regression with random effects: a case study*, Biometrics (2000), pp. 1030–1039. Available at http://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.2000.01030.x/full.

[13] W.W. Hsu, D. Todem, K. Kim, and W. Sohn, *A Wald test for zero inflation and deflation for correlated count data from dental caries research*. Stat. Modelling. 14 (2014), pp. 471–488.

[14] M.-C. Hu, M. Pavlicova, and E.V. Nunes, *Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial*. Am. J. Drug Alcohol Abuse 37 (2011), pp. 367–375. Available at http://www.tandfonline.com/doi/full/10.3109/00952990.2011.597280.

[15] S. Koda, and H. Ohara, *Preventive effects an low back pain and occupational injuries by providing the participatory occupational safety and health program*. J. Occup. Health 41 (1999), pp. 160–165. Available at https://www.jstage.jst.go.jp/article/joh1996/41/3/41_3_160/_article/-char/ja/.

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

[17] J.F. Lawless, *Negative binomial and mixed Poisson regression*. Can J Stat 15 (1987), pp. 209–225. Available at http://onlinelibrary.wiley.com/doi/10.2307/3314912/full.

[18] A.H. Lee, K. Wang, and K.K.W. Yau, *Analysis of zero-inflated poisson data incorporating extent of exposure*. Biom. J. 43 (2001), pp. 963–975. Available at https://www.researchgate.net/profile/Kui_Wang3/publication/230284700_Analysis_of_ZeroInflated_Poisson_Data_Incorporating_Extent_of_Exposure/links/53d099280cf2fd75bc5d33f1.pdf.

[19] D. McNeish, and K. Kelley, *Fixed effects models versus mixed effects models for clustered data: reviewing the approaches, disentangling the differences, and making recommendations*, (2018).

[20] A. Moghimbeigi, M.R. Eshraghian, K. Mohammad, and B. Mcardle, *Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros*. J. Appl. Stat. 35 (2008), pp. 1193–1202.

[21] S. Nakagawa, and H. Schielzeth, *A general and simple method for obtaining R2 from generalized linear mixed-effects models*. Methods Ecol Evol 4 (2013), pp. 133–142.

[22] S. Sen, P. Sengupta, and N. Diawara, *Doubly inflated Poisson model using Gaussian copula*. Commun Stat Theory Methods 47 (2017), pp. 1–11. Available at https://www.tandfonline.com/doi/full/10.1080/03610926.2017.1342831.

[23] P. Sengupta, N. Rao Chaganty, and R.T. Sabo, *Bivariate doubly inflated Poisson models with applications*. J. Stat. Theory. Pract. 10 (2016), pp. 202–215. Available at http://www.tandfonline.com/doi/full/10.1080/15598608.2015.1105166.

[24] M. Sheth-Chandra, *The Doubly Inflated Poisson and Related Regression Models*, ed, Old Dominion University, Norfolk, VA, 2011.

[25] D.M. Titterington, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, Chichester/New York, 1985.

[26] D. Todem, W.W. Hsu, and K.J.B. Kim, *On the efficiency of score tests for homogeneity in two-component parametric models for discrete data*. Biometrics 68 (2012), pp. 975–982.

[27] A.H. Welsh, R.B. Cunningham, C.F. Donnelly, and D.B. Lindenmayer, *Modelling the abundance of rare species: statistical models for counts with extra zeros*. Ecol. Modell 88 (1996), pp. 297–308. Available at http://www.sciencedirect.com/science/article/pii/0304380095001131.

[28] L. Xiang, A.H. Lee, K.K. Yau, and G.J. McLachlan, *A score test for zero-inflation in correlated count data*. Stat. Med. 25 (2006), pp. 1660–1671.

[29] I. Yahav, and G. Shmueli, *On generating multivariate Poisson data in management science applications*. Appl. Stoch. Models. Bus. Ind. 28 (2012), pp. 91–102. Available at http://doi.wiley.com/10.1002/asmb.901.

[30] Z. Yang, J.W. Hardin, and C.L. Addy, *Score tests for zero-inflation in overdispersed count data*. Commun Stat Theory Methods 39 (2010), pp. 2008–2030.

[31] K.K.W. Yau, K. Wang, and A.H. Lee, *Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros*. Biom. J. 45 (2003), pp. 437–452. Available at http://onlinelibrary.wiley.com/doi/10.1002/bimj.200390024/full.

[32] H. Zhu, S. Luo, and S.M. DeSantis, *Zero-inflated count models for longitudinal measurements with heterogeneous random effects*. Stat. Methods Med. Res. 0 (2015), pp. 1–6. Available at http://journals.sagepub.com/doi/10.1177/0962280215588224.

[33] D. Zwillinger, CRC Standard Mathematical Tables and Formulae, (2011).

Appendices

Table A1. Estimated regression coefficients for the DIP, ZIP and Poisson mixed models for simulated data from DIP distribution with different inflation probabilities ($n = 400$).

		Mixed-Effect Doubly-inflated Poisson Model				Mixed-Effect Zero-Inflated Poisson Model				Mixed-Effect Poisson Model			
		$n = 400, p_{1i} = 0.5$ and $p_{2i} = 0.3$											
	Real	Est.	S.E.	Rel. Bias.*	RMSE*	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	−0.3318	0.8524	−1.3318	−0.6901	−0.6901	0.2034	−1.6901	1.7023	−0.8074	0.1720	−1.8074	1.8156
$\beta1$	**0.2**	0.1700	0.2226	−0.1499	0.2245	0.1068	0.1548	−0.4659	0.1806	0.0960	0.1448	−0.5198	0.1782
$\beta2$	**−0.05**	−0.0422	0.0177	−0.1560	0.0193	−0.0274	0.0119	−0.4519	0.0255	−0.0247	0.0110	−0.5063	0.0276
$\beta3$	**0.3**	0.2750	0.0547	−0.0834	0.0601	0.2119	0.0430	−0.2938	0.0981	0.2103	0.0449	−0.2989	0.1003
	Real	Est.	S.E.	Rel. Bias.	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	0.9579	0.1302	−0.0421	0.1368	−0.9096	0.2411	−1.9096	1.9247	−1.1701	0.2022	−2.1701	2.1795
$\beta1$	**0.4**	0.4083	0.0753	0.0208	0.0757	0.1663	0.2387	−0.5842	0.3339	0.1855	0.1691	−0.5362	0.2731
$\beta2$	**−0.1**	−0.1032	0.0078	0.0323	0.0084	−0.0541	0.0167	−0.4591	0.0488	−0.0476	0.0132	−0.5242	0.0540
$\beta3$	**0.6**	0.6107	0.0396	0.0179	0.0410	0.4551	0.0458	−0.2414	0.1519	0.4608	0.0534	−0.2320	0.1491
		$n = 400, p_{1i} = 0.3$ and $p_{2i} = 0.5$											
	Real	Est.	S.E.	Rel. Bias.	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	−0.1828	0.5839	−1.1828	1.3189	−0.2768	0.1239	−1.2768	1.2828	−0.2793	0.1233	−1.2793	1.2853
$\beta1$	**0.2**	0.1698	0.1912	−0.1510	0.1935	0.0894	0.1019	−0.5532	0.1504	0.0879	0.1006	−0.5607	0.1506
$\beta2$	**−0.05**	−0.0426	0.0158	−0.1477	0.0174	−0.0224	0.0083	−0.5529	0.0289	−0.0222	0.0082	−0.5569	0.0290
$\beta3$	**0.3**	0.2773	0.0516	−0.0758	0.0564	0.1750	0.0364	−0.4166	0.1302	0.1749	0.0364	−0.4170	0.1303
	Real	Est.	S.E.	Rel. Bias.	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta0$	**1**	0.9551	0.1159	−0.0449	0.1242	−0.5033	0.1541	−1.5033	1.5111	−0.5623	0.1461	−1.5623	1.5691
$\beta1$	**0.4**	0.4092	0.0780	0.0229	0.0785	0.1843	0.1305	−0.5391	0.2520	0.1698	0.1197	−0.5755	0.2595
$\beta2$	**−0.1**	−0.1026	0.0079	0.0264	0.0083	−0.0452	0.0104	−0.5481	0.0558	−0.0422	0.0094	−0.5784	0.0586
$\beta3$	**0.6**	0.6116	0.0374	0.0194	0.0391	0.3960	0.0463	−0.3400	0.2092	0.3964	0.0477	−0.3393	0.2091

*Rel. Bias: Relative Bias, RMSE: Root of Mean Square Error.

Table A2. Estimated regression coefficients for the DIP, ZIP and Poisson mixed models for simulated data from DIP distribution with different inflation probabilities ($n = 50$).

		Mixed-Effect Doubly-inflated Poisson Model				Mixed-Effect Zero-Inflated Poisson Model				Mixed-Effect Poisson Model			
						$n = 50, p_{1i} = 0.5$ and $p_{2i} = 0.3$							
	Real	Est.	S.E.	Rel. Bias.*	RMSE*	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	-0.1342	1.0699	−1.1342	1.5588	−0.6878	0.5457	−1.6878	1.7737	−0.8214	0.5079	−1.8214	1.8908
$\beta 1$	**0.2**	0.1537	0.5576	−0.2317	0.5592	0.0759	0.4099	−0.6206	0.4281	0.0800	0.3948	−0.6000	0.4125
$\beta 2$	**−0.05**	−0.0481	0.0495	−0.0384	0.0495	−0.0278	0.0372	−0.4436	0.0433	−0.0244	0.0351	−0.5118	0.0434
$\beta 3$	**0.3**	0.2924	0.1627	−0.0254	0.1627	0.2066	0.1319	−0.3115	0.1616	0.2083	0.1335	−0.3057	0.1619
	Real	Est.	S.E.	Rel. Bias.	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.7248	0.7248	−0.2752	0.7750	−0.8856	0.6348	−1.8856	1.9894	−1.1489	0.6026	−2.1489	2.2317
$\beta 1$	**0.4**	0.3348	0.5102	−0.1629	0.5140	0.1041	0.5494	−0.7396	0.6237	0.1512	0.4971	−0.6221	0.5557
$\beta 2$	**−0.1**	−0.1121	0.0444	0.1208	0.0460	−0.0548	0.0445	−0.4522	0.0634	−0.0467	0.0396	−0.5327	0.0663
$\beta 3$	**0.6**	0.6218	0.1362	0.0364	0.1379	0.4386	0.1456	−0.2689	0.2173	0.4440	0.1621	−0.2600	0.2249
						$n = 50, p_{1i} = 0.3$ and $p_{2i} = 0.5$							
	Real	Est.	S.E.	Rel. Bias.	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	−0.0019	0.9018	−1.0019	1.3477	−0.2584	0.3777	−1.2584	1.3138	−0.2711	0.3705	−1.2711	1.3239
$\beta 1$	**0.2**	0.1580	0.5468	−0.2098	0.5481	0.0841	0.3117	−0.5794	0.3324	0.0833	0.3077	−0.5833	0.3290
$\beta 2$	**−0.05**	−0.0485	0.0472	−0.0301	0.0472	−0.0228	0.0243	−0.5449	0.0365	−0.0222	0.0240	−0.5557	0.0367
$\beta 3$	**0.3**	0.2908	0.1520	−0.0306	0.1522	0.1722	0.1026	−0.4261	0.1639	0.1728	0.1019	−0.4241	0.1630
	Real	Est.	S.E.	Rel. Bias.	RMSE	Est.	S.E.	Rel. Bias	RMSE	Est.	S.E.	Rel. Bias	RMSE
$\beta 0$	**1**	0.7721	0.5714	−0.2279	0.6149	−0.4877	0.4230	−1.4877	1.5466	−0.5643	0.4093	−1.5643	1.6169
$\beta 1$	**0.4**	0.3769	0.3952	−0.0579	0.3957	0.1698	0.3552	−0.5756	0.4232	0.1644	0.3368	−0.5889	0.4109
$\beta 2$	**−0.1**	−0.1111	0.0390	0.1110	0.0405	−0.0458	0.0290	−0.5421	0.0615	−0.0418	0.0265	−0.5824	0.0640
$\beta 3$	**0.6**	0.6227	0.1207	0.0379	0.1227	0.3884	0.1246	−0.3527	0.2455	0.3895	0.1317	−0.3508	0.2482

*Rel. Bias: Relative Bias, RMSE: Root of Mean Square Error.