Available Online at http://www.journalajst.com

ISSN: 0976-3376

ASIAN JOURNAL OF
SCIENCE AND TECHNOLOGY

Asian Journal of Science and Technology
Vol. 11, Issue, 09, pp.11172-11181, September, 2020

RESEARCH ARTICLE

BAYESIAN LINEAR MODEL WITH ZERO-INFLATED COVARIATES

[1,*] Adarabioyo, M. I.,[2]Adejuwon, S. O., [3]Ayemidotun, D. and [4]Adeyemo, O.A.

[1,4]Department of Mathematical Sciences, Afe Babalola University, Ado-Ekiti, Nigeria
[2]Department of Mathematical and Physical Sciences, Afe Babalola University, Ado-Ekiti, Nigeria
[3]Department of Business Administration, Rufus Giwa Polytechnic, Owo, Ondo State. Nigeria

ARTICLE INFO

Article History:
Received 17th June, 2020
Received in revised form
09th July, 2020
Accepted 24th August, 2020
Published online 30th September, 2020

Key words:

Bayesian, Linear, Model, Zero-Inflated, Covariate, Predictors.

ABSTRACT

Covariates sometimes called predictors or explanatory variables are factors that explain or predict the response or dependent variable. The impact of such covariates can be tested using several statistical tools depending on the nature of the data under consideration. This study was carried out to measure the distribution of zero-inflated covariates by fitting Zero-inflated Poisson, Zero-inflated Generalized Poisson and Zero-inflated Negative Binomial at different levels of covariates and sample size. Zero-inflated Poisson outperformed other models when the sample size was high while Zero-Inflated Negative Binomial tends to perform better at smaller sample size. The choice of which model to consider depends on the size of sample, proportion of zeros and the number of covariate to be included in the model. However, ZIP and ZINB look more suitable under any of the aforementioned scenario.

Citation: Adarabioyo, M. Adejuwon, S. O., Ayemidotun, D. and Adeyemo, O.A. 2020. "Bayesian linear model with zero-inflated covariates", *Asian Journal of Science and Technology*, 11, (09), 11172-11181.

INTRODUCTION

Covariates sometimes called predictors or explanatory variables are factors that explain or predict the response or dependent variable. The impact of such covariates can be tested using several statistical tools depending on the nature of the data under consideration. Bayesian statistics has over the years grown to be a very wildly used aspect of statistics because of its flexibility and robustness. Bayesian approaches have on several occasions outperformed the frequentist counterparts in modeling and analyzing important datasets or at least provided an alternative to the frequentist approaches. It generally refers to a theory in the field of statistics which is predominately based on Bayesian interpretation of probability. The degree of belief may be based on prior knowledge about the event, such as the results of previous experiments, or on personal beliefs about the event. This differs from a number of other interpretations of probability, such as the frequentist interpretation that views probability as the limit of the relative frequency of an event after a large number of trials.(1) One frequent manifestation of overdispersion is that the incidence of zero counts is greater than expected for the Poisson distribution and this is of interest because zero counts frequently have special status.(2) Encountering an excess number of zeros is pretty common in some clinical experiments. In such cases, mixed-distribution models like the zero-inflated Poisson and zero inflated Negative Binomial are often employed (3). Several researches have worked on the Bayesian component of zero-inflated models. To mention a few, Neelon et al for example described a practical Bayesian approach which incorporates prior information, has optimal small-sample properties and allows for tractable inference (3). Zero-inflated distributions (ZID) were studied from the Bayesian point of view using the data augmentation algorithm. The zero-inflated Poisson distribution (ZIP) and an illustrative example via MCMC algorithm were considered (4). A zero-inflated Poisson regression model with random effects to evaluate a manual handling injury prevention strategy was discussed in (5). A comparison study of several modeling strategies for vaccine adverse event count data in which the data were characterized by excess zeroes and heteroskedasticity. The compared models the Poisson, Negative Binomial (NB), zero-inflated Poisson (ZIP), zero-inflated Negative Binomial (ZINB), Poisson Hurdle (PH), and Negative Binomial Hurdle (NBH)(6). An explanation of the Bayesian linear regression was the focus of (7). In the Bayesian point of view, linear regression was formulated using distributions rather than point estimates with the response was not estimated as a single value but was assumed to be drawn from a probability distribution. The Bayesian Linear Regression was constructed such that it has the capacity to determine the posterior distribution for the model parameters and not to find a single best value of the model parameters. Markov Chain Monte

***Corresponding author:* Adarabiyo, M. I.,**
Department of Mathematical Sciences, Afe Babalola University, Ado-Ekiti, Nigeria.

Carlo was employed to demonstrate it application. Some many other authors have contributed knowledge on Bayesian Linear Regression models. To mention a few (8,9,10). A flexible class of zero inflated models which includes other familiar models such as the Zero Inflated Poisson (ZIP) models, as special cases was introduced since a Bayesian estimation method is developed as an alternative to traditionally used maximum likelihood based methods to analyze such data. Simulation studies showed that the proposed method has better finite sample performance than the classical method with tighter interval estimates and better coverage probabilities. A real-life data set is analyzed to illustrate the practicability of the proposed method easily implemented using WinBUGS. (11). A study examining spatiotemporal patterns in inpatient hospitalizations was the motivation of proposing an efficient Bayesian approach for fitting zero-inflated negative binomial models. To facilitate posterior sampling, a set of latent variables that are represented as scale mixtures of normals, where the precision terms follow independent Polya-Gamma distributions was introduced. Conditional on the latent variables, inference was made from straightforward Gibbs sampling. (12)
In this paper, we present a Bayesian linear model with zero-inflated covariates. Rigorous simulations were performed to examine the behavior of the Bayesian linear model under variety of situations.

METHODOLOGY

Let $y = (y_i)_{i=1}^n$ be an nx1 vector of independent observations on a dependent variable. Px1 is the vector of regressors, X_i. The linear expression is given by

$$y = X\beta + \varepsilon,$$

Where $X = (X_i^T)_{i=1}^n$ is a nxp matrix of regression with ith row being X_i^T, β is the slope vector of the regression coefficient and $\varepsilon = (\varepsilon)_{i=1}^n$ is the vector of the random error. The observations will be assumed to be independent. Therefore,

$\epsilon \sim MVN(0, \sigma^2 I_n)$ and the column of the matrix are linearly independent, hence the rank of X is p.

We are going to assume a Normal Inverse Gamma (NIG) as conjugate priors for y which is of the form

$$p(\beta, \sigma^2) = P(\beta,\ \sigma^2)P(\sigma^2)P(\sigma^2) = N(\mu,\ \sigma^2 v_\beta \text{x} IG(a, b)$$

$$=NIG(\mu_\beta, V_\beta, a, b)$$

Where $\Gamma(.)$ is the Gamma function and IG(a, b) prior density for σ^2 is given by

$$p(\ \sigma^2) = \frac{b^a}{\Gamma(a)}(\frac{1}{\sigma^2})^{a+1}\exp\left(-\frac{b}{a^2}\right), \sigma^2 > 0$$

Where a, b>0. This referred to as Normal Inverse Gamma (NIG) prior denoted as $NIG(\mu_\beta, V_\beta, a, b)$.

The NIG probability distribution is a joint probability distribution of a vector and a scalar σ^2. If $(\beta, \sigma^2) \sim NIG(\mu, v, a, b)$, then an interesting analytical form results from integrating out σ^2 from the joint density is obtained as

$$\int NIG\ ((\mu, v, a, b)d\sigma^2 = \frac{b^a}{2\pi^{\frac{p}{2}}} \mid v \mid^{\frac{1}{2}}\Gamma(a)\int \frac{1}{\sigma^2}^{a+1} \exp\{-\frac{1}{\sigma^2}(b + \frac{1}{2}(\beta-\mu)^T v^{-1}(\beta-\mu)d\sigma^2$$

$$= \frac{\Gamma(a+\frac{p}{2})}{\pi^{p/2}} \mid \frac{(2a)b}{a}v \mid^{\frac{1}{2}}\Gamma(a)\left[1 + \frac{(\mu-\beta)^T\left[\frac{b}{a}v\right]^{-1}(\beta-\mu)}{2a}\right]^{\frac{-2a/p}{2}}$$

This is equivalent to multivariate t density MVStv

$$MVStv(\mu, \varepsilon) = \frac{\Gamma(\frac{v+p}{2})}{\Gamma(\frac{v}{2})\pi^{p/2} \mid v\Sigma \mid^{\frac{1}{2}}}[1 + (\beta-\mu)^T\Sigma^{-1}(\beta-\mu)]^{-\frac{v+p}{2}}$$

With $v = 2a$ and $(\Sigma = (\frac{b}{a})v$

The likelihood of the model is given by

$$p(y/\beta\ , \sigma^2) = N(X\beta, \sigma^2 I) = (\frac{1}{2\pi\sigma^2}exp\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\}$$

Posterior distribution from NIG prior is given by

$$p(\beta, \sigma^2/y) = \frac{p(\beta, \sigma^2)p(y/\beta, \sigma^2)}{p(y)}$$

Where $p(y)$=$\int(\beta,\sigma^2)p(y/\beta,\sigma^2)d\beta d\sigma^2$ is the marginal distribution of the data.

Simulation Settings

Four types of simulations were performed. The first simulation involved simulation of the covariates X_1, X_2, and X_3 from Zero-Inflated Poisson distributions. X_1 was simulated from Zero-Inflation Poisson with parameters $\lambda = 0.6$ *and* $\omega = 0.8$, X2 was simulated from ZIP with parameters $\lambda = 0.6$ *and* $\omega = 0.2$ and X3 with parameters $\lambda = 0.6$ *and* $\omega = 0.2$. The second type was simulated from Zero-Inflated Generalized Poisson (ZIGP). X_1 was simulated with parameters $\lambda = 0.6$ *and* $\omega = 0.8$, X_2 with parameters $\lambda = 0.6$ *and* $\omega = 0.2$ and X_3 with Parameters $\lambda = 0.6$ *and* $\omega = 0.2$. Likewise the third model was Zero-Inflated Negative Binomial with X_1, X_2 and X_3 simulated with parameters $\lambda = 0.6$ *and* $\omega = 0.8$ for X1, $\lambda = 0.6$ *and* $\omega = 0.2$ for X_2 and X_3. However the error term e was simulated from Normal distribution N(0,1). The prior parameters were set as $\beta_0 = 3.0, \beta_1 = 10.0$, β_2=30.5 and $\beta_3 = 15.0$ and β_4 =5.0

RESULTS AND DISCUSSION

We present the posterior means and posterior standard deviations of the Bayesian linear model with some sets of covariates derived from Zero-inflated Poisson (ZIP), Zero-inflated Negative Binomial (ZINB) and Zero-inflated Generalized Poisson (ZIGP). The first set of the models has four covariates. The precision of the models were obtained in order to determine model fit.

Table 1: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 15, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			Sample size (n) = 15, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0299	0.1758	-0.0604	0.1933	0.0381	0.1846
β_1	11.1222	0.4439	11.3563	0.4325	10.2214	0.2606
β_2	29.8861	0.2624	30.3530	0.4259	30.2052	0.2911
β_3	16.4544	0.4361	14.5991	0.4328	15.0432	0.2923
β_4	6.3183	0.3529	5.9273	0.4535	6.7011	0.2778
Precision	1.92		1.03		0.87	

From table1, the posterior mean for β_0 = 0.0299 and β_2 =0.2624 were the least in ZIP when compared with that of ZIGP and ZINB. Likewise the posterior mean for β_1 =0.2606 and β_4 = 0.2778 were the least in ZINB when compared with others. Under the above simulation techniques ZIP has the highest precision of 1.92. The linear model with ZIP covariates performed better when the sample size is 15.

Table 2: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 15, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$			Sample size (n) = 15, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0131	0.1727	-0.0543	0.1932	0.0707	0.1790
β_1	11.0839	0.4376	11.0436	0.3545	10.2962	0.2392
β_2	29.8720	0.2609	29.9850	0.3106	30.2355	0.2881
β_3	16.4769	0.4339	14.8184	0.3964	15.0639	0.2909
Precision	1.83		1.07		0.84	

In table 2, the posterior mean for β_0 = 0.1727 and β_2 =0.2609 were also the least in ZIP when compared with that of ZIGP and ZINB. Likewise the posterior mean for β_1 =0.2392 and β_4 = 0.2909 were the least in ZINB when compared with others. Under the above simulation techniques ZIP has the highest precision of 1.83. The linear model with ZIP covariates equally performed better when the number of covariates was reduced to three and the sample size is 15.

Table 3: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 15, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2$			Sample size (n) = 15, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0475	0.1724	-0.0609	0.1927	0.0823	0.1710
β_1	10.7487	0.4264	10.9455	0.2826	10.2957	0.2392
β_2	30.3615	0.2177	29.9593	0.3055	30.2335	0.2879
Precision	0.82		0.84		1.03	

From table 3, the posterior mean for β_0 = 0.1710 and β_1 =0.2392 were the least in ZINB when compared with that of ZIGP and ZIP. Likewise the posterior mean for β_1 =0.2177 was the least in ZIP when compared with others.

Under the above simulation techniques ZINB has the highest precision of 1.03. The linear model with ZINB covariates performed better under two covariates linear model when the sample size is 15.

Table 4: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 20, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			Sample size (n) = 20, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0230	0.1752	-0.0390	0.1916	-0.0139	0.0762
β_1	9.9774	0.4264	10.6947	0.3001	10.0140	0.0458
β_2	30.4471	0.1908	30.4079	0.3741	30.5124	0.0452
β_3	15.7794	0.3162	14.9531	0.2993	14.9709	0.0451
β_4	6.4461	0.2969	6.4679	0.3398	6.6232	0.0446
Precision	0.78		0.76		1.40	

From table 4, the posterior standard deviations of all the parameters were minimum under ZINB. That is β_0 = 0.0762, β_1 =0.0458, …, β_4 =0.0446 were the least in ZINB when compared to that of ZIGP and ZIP. Under the above simulation techniques ZINB has the highest precision of 1.40. The linear model with ZINB covariates performed better under four covariates linear model when the sample size is increased to 20.

Table 5: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 20, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$			Sample size (n) = 20, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0214	0.1750	-0.0395	0.1915	0.0426	0.1637
β_1	9.9508	0.4006	10.6855	0.2837	11.4020	0.5174
β_2	30.4278	0.1587	30.3851	0.2856	30.5160	0.2983
β_3	15.8004	0.2942	14.9575	0.2957	14.6921	0.3038
Precision	1.391		0.762		0.734	

From table 5, the posterior standard deviation when β_0=0.1637was minimum under ZINB,β_1 = 0.2837 was minimum under ZIGP and β_2 =0.1587 and β_3 = 0.2942 were the least under ZIP when compared with that of ZIGP and ZINB. However, under the above simulation techniques ZIP has the highest precision of 1.391. The linear model with ZIP covariates performed better under three covariates linear model when the sample size is increased to 20.

Table 6: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 20, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2$			Sample size (n) = 20, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0456	0.1747	-0.0421	0.1907	0.0150	0.1614
β_1	10.3887	0.3669	10.6698	0.2618	11.3001	0.5075
β_2	30.5499	0.1522	30.3665	0.2549	30.4442	0.2898
Precision	0.746		0.751		0.719	

From table 6, the posterior standard deviation when β was minimum under ZINB,β_1 was minimum under ZIGP and β_2 was least under ZIGP .However, under the above simulation techniques ZIP has the highest precision of 0.751. The linear model with ZIP covariates performed better under two covariates linear model when the sample size is increased to 20.

Table 7: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 25, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			Sample size (n) = 25, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
P(Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1073	0.1726	-0.0170	0.1898	-0.0170	0.1898
β_1	11.9137	0.6084	10.7373	0.2918	10.7373	0.2918
β_2	30.4389	0.2091	30.5137	0.3495	30.5137	0.3495
β_3	14.8626	0.2310	14.8961	0.2649	14.8961	0.2649
β_4	6.5432	0.2377	6.3985	0.3303	6.3985	0.3303
Precision	0.891		0.864		0.905	

From table 7, the posterior standard deviation when β_0 was minima under ZIP, β_1 was least under ZIGP and ZINB, β_2, β_3 and β_4 were least under ZIP. However, under the above simulation techniques ZINB has the highest precision of 0.905. The linear model with ZINB covariates performed better under four covariates linear model when the sample size is increased to 25.

Table 8: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 25, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$			Sample size (n) = 25, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1162	0.1655	-0.0173	0.1898	0.1314	0.1581
β_1	11.9013	0.6046	10.7043	0.2714	10.0676	0.2014
β_2	30.4424	0.2083	30.4393	0.2522	30.4675	0.2179
β_3	14.8578	0.2295	14.9128	0.2592	15.4551	0.5635
Precision	0.869		0.867		0.875	

From table 8, the posterior standard deviation when β_0 and β_1 was minima under ZINB and $\beta_2$and β_3 were least under ZIP. However, under the above simulation techniques ZINB has the highest precision of 0.875.

The linear model with ZINB covariates performed better under three covariates linear model when the sample size is increased to 25.

Table 9: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 25, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$			Sample size (n) = 25, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0952	0.1620	-0.0266	0.1878	0.1561	0.1551
β_1	11.9422	0.6010	10.6690	0.2502	10.0525	0.2006
β_2	30.4248	0.2064	30.4080	0.2345	30.5363	0.2005
Precision	0.837		0.858		0.864	

From table 9, the posterior standard deviation for β_1, β_1 and β_2 were least under ZINB. However, under the above simulation techniques ZINB has the highest precision of 0.864.

The linear model with ZINB covariates performed better under two covariates linear model when the sample size is increased to 25.

Table 10: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 40, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$			Sample size (n) = 40, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1856	0.1490	0.0268	0.1851	0.0268	0.1851
β_1	10.2140	0.1873	10.6017	0.2320	10.6017	0.2320
β_2	30.6999	0.1448	30.5435	0.2656	30.5435	0.2656
β_3	14.8198	0.2156	14.8045	0.2114	14.8045	0.2114
β_4	6.4497	0.1673	6.6294	0.2326	6.6294	0.2326
Precision	1.016		1.007		0.972	

From table 10, the posterior standard deviation for β_0, β_1, β_2 and β_4 were least under ZIP while β_3 was minimum under ZIGP. However, under the above simulation techniques ZIP has the highest precision of 1.016. The linear model with ZIP covariates performed better under four covariates linear model when the sample size is increased to 40.

Table 11: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 40, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$			Sample size (n) = 40, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1723	0.1423	0.0357	0.1844	0.1414	0.1449
β_1	10.2215	0.1856	10.6130	0.2311	10.3234	0.2003
β_2	30.6921	0.1424	30.6331	0.2112	30.5780	0.2023
β_3	14.8206	0.2156	14.8088	0.2112	15.0286	0.1755
Precision	1.003		0.9881		0.966	

From table 11, the posterior standard deviation for β_0, β_1 and β_2 were the least under ZIP while β_3 was the least under ZIP while β_3 was least under ZIGP. However, under the above simulation techniques ZIP has the highest precision of 1.003. The linear model with ZIP covariates performed better under three covariates linear model when the sample size is increased to 40.

Table 12: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 40, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$			Sample size (n) = 40, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1554	0.1408	0.0029	0.1808	0.1454	0.1428
β_1	10.2110	0.1852	10.5590	0.2233	10.3344	0.1887
β_2	30.6520	0.1340	30.5580	0.1942	30.5787	0.2023
Precision	1.012		0.978		0.988	

From table 12, the posterior standard deviation for β_0, β_1 and β_2 were the least under ZIP. However, under the above simulation techniques ZIP has the highest precision of 1.003. The linear model with ZIP covariates performed better under two covariates linear model when the sample size is increased to 40.

Table 13: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 80, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$			Sample size (n) = 80, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0466	0.1259	-0.0243	0.1708	-0.0243	0.1708
β_1	10.1520	0.1784	10.1898	0.1767	10.1898	0.1767
β_2	30.8403	0.1441	30.6442	0.1520	30.6442	0.1520
β_3	15.0206	0.1277	14.8892	0.1466	14.8892	0.1466
β_4	6.4847	0.1331	6.6316	0.1432	6.6316	0.1432
Precision	1.035		0.993		0.979	

From table 13, the posterior standard deviation for β_0, β_1, β_2 , β_3 and β_4 were the least under ZIP while ZIGP and ZINB performed the same waywas the least under ZIP while β_3 was least under ZIGP. However, under the above simulation techniques ZIP has the highest precision of 1.003. The linear model with ZIP covariates performed better under four covariates linear model when the sample size is increased to 80.

Table 14: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 80, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$			**Sample size (n) = 80, phi=0.6, omega=0.8**			
Distribution	**ZIP**		**ZIGP**		**ZINB**	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0432	0.1224	0.0146	0.1654	0.1395	0.0985
β_1	10.1505	0.1780	10.1913	0.1767	10.1142	0.0927
β_2	30.8398	0.1441	30.6856	0.1451	30.4496	0.0847
β_3	15.0182	0.1259	14.9171	0.1435	15.0368	0.0652
Precision	**0.969**		**0.985**		**1.034**	

From table 14, the posterior standard deviation for β_0, β_1, β_2 , β_3 and β_4 were the least under ZINB while ZIGP with precision value of 0.985 performed better than ZIP with precision value of 0.969. However, under the above simulation techniques ZIGP has the highest precision of 1.034. The linear model with ZIGP covariates performed better under three covariates linear model when the sample size is increased to 80.

Table 15: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 80, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$			**Sample size (n) = 80, phi=0.6, omega=0.8**			
Distribution	**ZIP**		**ZIGP**		**ZINB**	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0499	0.1134	-0.0185	0.1552	0.1310	0.1147
β_1	10.1514	0.1778	10.1606	0.1685	10.1330	0.1263
β_2	30.8378	0.1434	30.6756	0.1441	30.5030	0.0986
Precision	**1.033**		**0.983**		**0.969**	

From table 14, the posterior standard deviation for β_0 was least under ZIP, $\beta_1$and β_2 were least under ZINB while ZIP with precision value of 1.033 performed better than ZIGP with precision value of 0.983 and ZINB of precision value of 0.969. The linear model with ZIP covariates performed better under two covariates linear model when the sample size is increased to 80.

Table 16: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 150, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			**Sample size (n) = 150, phi=0.6, omega=0.8**			
Distribution	**ZIP**		**ZIGP**		**ZINB**	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1833	0.1015	0.0055	0.1484	0.0055	0.1484
β_1	10.0947	0.1133	10.1031	0.1237	10.1031	0.1237
β_2	30.4040	0.0267	30.6260	0.1146	30.6260	0.1146
β_3	15.0193	0.0808	15.0608	0.1043	15.0608	0.1043
β_4	6.4900	0.0925	6.4880	0.1038	6.4880	0.1038
Precision	**0.942**		**0.938**		**0.933**	

From table 16, the posterior standard deviation for β_0,β_1, β_2 ,β_3 and β_4 were the least under ZIP. The precision value for ZIP was 0.942, ZIGP was 0.938 and ZINB was 0.933. Therefore, ZIP outperformed other models with highest precision. The linear model with ZIP covariates performed better under four covariates linear model when the sample size is increased to 150.

Table 17: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 150, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$			Sample size (n) = 150, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1798	0.0963	-0.0003	0.1397	0.1719	0.0734
β_1	10.0941	0.1132	10.1026	0.1236	10.0144	0.0687
β_2	30.4040	0.1122	30.6238	0.1132	30.4480	0.0623
β_3	15.0190	0.0808	15.0593	0.1035	14.9820	0.0534
Precision	0.932		0.934		0.935	

From table 17, the posterior standard deviation for β_0,β_1, β_2 and β_3 were the least under ZINB. The precision value for ZINB was 0.935, ZIGP was 0.934 and ZIP was 0.932. Therefore, ZINB outperformed other models with highest precision. The linear model with ZINB covariates performed better under three covariates linear model when the sample size is increased to 150.

Table 18: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 150, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2$			Sample size (n) = 150, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1882	0.0894	0.0359	0.1245	0.1528	0.0957
β_1	10.0922	0.1130	10.1103	0.1229	10.1215	0.0918
β_2	30.4019	0.1119	30.6269	0.1130	30.4461	0.0845
Precision	0.932		0.933		0.934	

From table 3.18, the posterior standard deviation for β_0,β_1 and β_2 were the least under ZINB. The precision value for ZINB was 0.9342, ZIGP was 0.933 and ZIP was 0.932. Therefore, ZINB outperformed other models with highest precision. The linear model with ZINB covariates performed better under two covariates linear model when the sample size is increased to 150.

Table 19: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 300, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			Sample size (n) = 300, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.2087	0.0780	-0.0255	0.1248	-0.0255	0.1248
β_1	9.9052	0.0719	10.0304	0.0829	10.0304	0.0829
β_2	30.5110	0.0642	30.5989	0.0834	30.5989	0.0834
β_3	14.9429	0.0761	15.0509	15.0509	15.0509	0.0779
β_4	6.4678	0.0694	6.5478	6.5478	6.5478	0.0755
Precision	1.029		0.978		1.020	

From table 19, the posterior standard deviation for β_0,β_1, β_2 , β_3 and β_4 ,were the least under ZIP. The precision value for ZIP was 1.029, ZIGP was 0.978 and ZINB was 1.020. Therefore, ZIP outperformed other models with highest precision. The linear model with ZIP covariates performed better under four covariates linear model when the sample size is increased to 300.

Table 20: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 300, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$			Sample size (n) = 300, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1951	0.0723	0.0090	0.1123	0.1719	0.0734
β_1	9.9060	0.0719	10.0323	0.0828	10.0144	0.0688
β_2	30.5094	0.0641	30.6019	0.0833	30.4480	0.0623
β_3	14.9445	0.7600	15.0517	0.0779	14.9820	0.0534
Precision	1.027		1.021		1.018	

From table 20, the posterior standard deviation for β_0 was least under ZIPbut β_1, β_2 ,and β_3 ,were the least under ZINB. The precision value for ZIP was 1.027, ZIGP was 1.021 and ZINB was 1.018. Therefore, ZIP outperformed other models with highest precision. The linear model with ZIP covariates performed better under three covariates linear model when the sample size is increased to 300.

Table 21: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 300, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2$			Sample size (n) = 300, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.1751	0.0669	0.0434	0.0996	0.1647	0.0703
β_1	9.9066	0.0719	10.0377	0.0824	10.0120	0.0684
β_2	30.5085	0.0641	30.6055	0.0831	30.4488	0.0623
Precision	1.025		1.019		1.018	

From table 21, the posterior standard deviation for β_0 was least under ZIP but β_1 and β_2 ,were the least under ZINB. The precision value for ZIP was 1.025, ZIGP was 1.019 and ZINB was 1.018. Therefore, ZIP outperformed other models with highest precision.. The linear model with ZIP covariates performed better under two covariates linear model when the sample size is increased to 300.

Table 22: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 500, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			Sample size (n) = 500, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0454	0.0620	-0.1290	0.1031	-0.1290	0.1031
β_1	9.9785	0.0566	10.0517	0.0655	10.0517	0.0655
β_2	30.5322	0.0512	30.6288	0.0647	30.6288	0.0647
β_3	15.0323	0.0634	14.9755	0.0624	14.9755	0.0624
β_4	6.5727	0.0548	6.6275	0.0601	6.6275	0.0601
Precision	0.985		0.972		0.973	

From table 22, the posterior standard deviation for β_0 , β_1, β_2 ,and β_4 ,were the least under ZIP but β_3 was least under ZIGP and ZINB. The precision value for ZIP was 0.985, ZIGP was 0.972 and ZINB was 0.973. Therefore, ZIP outperformed other models with highest precision. The linear model with ZIP covariates performed better under four covariates linear model when the sample size is increased to 500.

Table 23: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 500, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$			Sample size (n) = 500, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0765	0.0574	-0.0288	0.0916	0.1160	0.0587
β_1	9.9774	0.0566	10.0508	0.0655	10.0381	0.0500
β_2	30.5309	0.0512	30.6329	0.0647	30.4853	0.0491
β_3	15.0312	0.0634	14.9773	0.0624	14.9195	0.0524
Precision	0.969		0.972		0.977	

From table 23, the posterior standard deviation for β_0 was least under ZIP. However, β_1, β_2 , β_3 and β_4 ,were the least under ZINB. The precision value for ZIP was 0.969, ZIGP was 0.972 and ZINB was 0.977. Therefore, ZINB outperformed other models with highest precision. The linear model with ZINB covariates performed better under three covariates linear model when the sample size is increased to 500.

Table 24. The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 500, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$			Sample size (n) = 500, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0876	0.0528	-0.0442	0.0813	0.0824	0.0544
β_1	9.9770	0.0566	10.0485	0.0651	10.0369	0.0500
β_2	30.5298	0.0511	30.6322	0.0647	30.4834	0.0491
Precision	0.969		0.968		0.976	

From table 24, the posterior standard deviation for β_0 was least under ZIP. However β_1, β_2 andβ_3,were the least under ZINB. The precision value for ZIP was 0.969, ZIGP was 0.968 and ZINB was 0.976. Therefore, ZINB outperformed other models with highest precision. The linear model with ZINB covariates performed better under two covariates linear model when the sample size is increased to 500.

Table 25: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 1000, phi=0.6, omega=0.8 with four covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$			Sample size (n) = 1000, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0954	0.0437	-0.0139	0.0762	-0.0139	0.0762
β_1	10.0200	0.0385	10.0140	0.0458	10.0140	0.0458
β_2	30.5071	0.0416	30.5124	0.0452	30.5124	0.0452
β_3	14.9515	0.0410	14.9709	0.0451	14.9709	0.0451
β_4	6.4895	0.0419	6.6232	0.0446	6.6232	0.0446
Precision	0.963		0.957		0.960	

From table 25, the posterior standard deviation for β_0, β_1, β_2 and β_4 ,were the least under ZIP. The precision value for ZIP was 0.963, ZIGP was 0.957 and ZINB was 0.960. Therefore, ZINB outperformed other models with highest precision. The linear model with ZIP covariates performed better under four covariates linear model when the sample size is increased to 1000.

Table 26: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 1000, phi=0.6, omega=0.8 with three covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$			Sample size (n) = 1000, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0915	0.0408	0.0818	0.0679	0.0688	0.0419
β_1	10.0199	0.0385	10.0158	0.0458	10.0362	0.0352
β_2	30.5075	0.0416	30.5159	0.0452	30.5376	0.0332
β_3	14.9513	0.0410	14.9696	0.0451	14.9549	0.0340
Precision	0.957		0.956		0.960	

From table 26, the posterior standard deviation for $\beta_0$was least under ZIP but β_1, β_2 ,and β_3 ,were the least under ZINB. The precision value for ZIP was 0.957, ZIGP was 0.956 and ZINB was 0.960. Therefore, ZINB outperformed other models with highest precision. The linear model with ZIP covariates performed better under three covariates linear model when the sample size is increased to 1000.

Table 27: The Posterior means, posterior standard deviations and the Precisions at Sample size (n) = 1000, phi=0.6, omega=0.8 with two covariates

Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$			Sample size (n) = 1000, phi=0.6, omega=0.8			
Distribution	ZIP		ZIGP		ZINB	
Parameter	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev	Posterior Mean	Posterior St.Dev
β_0	0.0740	0.0381	0.0594	0.0591	0.0478	0.0387
β_1	10.0188	0.0385	10.0150	0.0458	10.0353	0.0352
β_2	30.5054	0.0416	30.5154	0.0452	30.5388	0.0332
Precision	0.956		0.956		0.958	

From table 27, the posterior standard deviation for β_0 was least under ZIP but β_1, β_2 ,β_3 ,were the least under ZINB. The precision value for ZIP was 0.956, ZIGP was 0.956 and ZINB was 0.958. Therefore, ZINB outperformed other models with highest precision. The linear model with ZIP covariates performed better under two covariates linear model when the sample size is increased to 1000.

DISCUSSION AND CONCLUSION

Bayesian estimation of the posterior means, the standard deviations and precisions of the models were estimated. The precision was used as model selection criterion. The analysis revealed the important role proportions of zero, sample size and number of covariates play in model selection. From the analysis, at low sample size of 20 for instance, ZINB outperformed when covariates was four, ZIGP when covariate was three and ZINB when covariate was two. Likewise at sample size of 40, ZIP outperformed other models at all levels of covariates. At sample size of 80, ZIP also outperformed when covariate was four and two whereas ZINB outperformed when covariate was three. In like manner, when the sample size was 300 ZIP outperformed at all levels of covariate whereas ZINB outperformed when covariate level was three and two but ZIP outperformed when covariate level was four. However, at sample size of 1000, ZINB outperformed when covariate level was three and two while ZIP outperformed when covariate level was four. In summary, the choice of which model to consider depends on the size of sample, proportion of zeros and the number of covariate to be included in the model. However, ZIP and ZINB look more suitable under any of the aforementioned scenario.

REFERENCES

1. Gelman, A.*, Carlin, J. B., Stern, H. S.,Dunson, D.B., Vehtari A., Rubin, D. B. (2013). Bayesian Data Analysis, Third Edition. Chapman and Hall/CRC.*
2. Ridout M., Demétrio C.G.B, Hinde J. Models for count data with many zeros. Proceedings from the International Biometric Conference; Cape Town. December 1998.1998.
3. Neelon B.H., O'Malley A.J. and Normand S.T. (2010) A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. Stat Modelling 10(4):42
4. Rodrigues J. (2003) Bayesian analysis of zero-inflated distributions. Communications in Statistics, Theory and Methods. 32:281–89.
5. Yau K.K., Lee A.H.(2001) Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. Statistics in Medicine. 20:2907–2920.
6. Rose C.E., Martin S.W., Wannemuehler K.A., PlikaytisB.D.(2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of Biopharm Statistics. 16(4):463-481.
7. Will K. (2018). Introduction to Bayesian Linear Regression. https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7.
8. Thomas P.M. (2010). Bayesian Linear Regression
9. https://tminka.github.io/papers/minka-linear.pdf
10. Lyle B.(1985). Bayesian Analysis of linear models. Marcel Dekker.
11. Elster C., Klauenberg K., Walzel M., Wubbeler G., Harris P., Cox M., Matthews C., Smith I., Wright L., Allard A., Fischer N., Cowen S., Ellison S.,Wilson P, Pennecchi F., Kok G., Van der Veen A., Pendrill L.(2015) A Guide to BayesianInference for Regression Problems, Deliverable of EMRP project NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation", 2015.
12. Ghosh S.K., Mukhopadhyay P, Lu. J.C.(2006). Bayesian analysis of zero-inflated regression models. Journal of Statistical Planning and Inference 136(4):1360-1375.
13. Neelon B. (2018). Bayesian Zero-Inflated Negative Binomial Regression Based onPolya-Gamma Mixtures. Bayesian Analysis. TBA:1-27.

****