CAUCHY: Jurnal Matematika Murni dan Aplikasi
Volume 10(1) (2025), Pages 213-223
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Simulation Study of Bayesian Zero Inflated Poisson Regression

Candra RWS Weni Utomo*, **Achmad Efendi, Ni Wayan S. Wardhani**

Departement of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Indonesia

Email: rezziningcandra@gmail.com

ABSTRACT

The purpose of this research is to evaluate the performance of ZIP regression analysis using Bayesian and MLE. The research data is secondary data obtained from the Health Office and the Indonesian Bureau of Statistics (BPS) in East Java with a total of 38 data samples and simulation data. Simulation studies in this study were conducted to see how the performance of ZIP regression analysis using MLE and Bayesian methods. The scenario aspects used are the sample sizes and the proportion of zero values. The results show that the development of ZIP Bayesian regression produces a model that is in accordance with the data conditions, that is excess zero and can handle overdispersion due to excess zero. Bayesian ZIP regression analysis is better than MLE at small sample sizes because it produces relatively stable parameter estimates in each simulation scenario and has the smallest RMSE value compared to MLE ZIP regression analysis. Based on DIC, the ZIP model works better on data with a higher proportion of zeros so that this model is more effectively used on data that has high zero-inflation.

Keywords: Bayesian; Measles; MLE; Overdispersion; Zero Inflation

INTRODUCTION

The Poisson model is commonly employed for analyzing count data [1]. The Poisson distribution is a probability distribution for the number of events that occur in a given time or space interval, when the events occur with a fixed success rate, and the events are mutually independent [2]. The Poisson distribution assumes that the mean is equal to the variance [3]. However, in its application it is often found that the variance is greater than the mean or what is called overdispersion and the variance is smaller than the mean which is called underdispersion [4]. One of the causes of this overdispersion is the presence of excess zeros in the data [5].

In the health sector, especially in measles deaths, zero-valued data are often found. Therefore, it is difficult for the Poisson or Negative Binomial models to accurately represent measles death data due to excessive dispersion due to the number of zeros in the data exceeding expectations based on these models. To consider the excess zeros in order to obtain unbiased estimation results, the Hurdle model and Zero Inflated model were developed. In research conducted by [6], first introduced the Zero Inflated Poisson model using a mixed model approach, by separating into two components, namely zero state and Poisson state. In the study conducted by [7], with the title A comparison of zero-

Submitted: November 03, 2024 Reviewed: January 08, 2025 Accepted: March 17, 2025
DOI: https://doi.org/10.18860/cauchy.v10i1.30207

inflated and hurdle models for modeling zero-inflated count data explains the performance of the Zero Inflated and Hurdle models in modeling data on the Poisson and Negative Binomial distributions. While research conducted by [8], with the title A test of inflated zeros for Poisson regression models explains the testing of the existence of excess zeros by comparing with the Vuong Test, the results of research on simulation studies show that the hypothesis test used is better than the Vuong test. Furthermore, research conducted by [9], with the title Estimation of Poisson Regression Parameters with Maximum Likelihood explains that the Maximum Likelihood Estimation (MLE) process is completed using the Fisher Scoring algorithm, an iterative approach that is utilized to update parameters repeatedly to approach the maximum likelihood value. This research was applied to the number of accident data on Minnesota state highways, and the obtained Poisson Regression model is represented.

A commonly used parameter estimation method in ZIP regression is Maximum Likelihood Estimation (MLE) [10]. MLE is an efficient technique in many cases, allowing for optimal parameter estimation based on the available data. However, although MLE is very reliable in the context of large sample sizes, it faces challenges when applied to small sample sizes. According to research conducted by [11], parameter estimation using MLE on small samples tends to be less stable and may result in biased estimation. This instability is due to the inability of MLE to fully capture the variations and patterns present in a limited sample, which in turn can lead to model misinterpretation [12]. Bayesian methods have proven effective in dealing with overdispersion, there are still some challenges in their implementation. Determining the appropriate prior distribution is often a problem, especially when initial information is limited. Besides, research on the performance of the ZIP model using the Bayesian approach on data with various sample sizes and varying proportions of zeros is still limited, especially in the context of health data such as measles cases in Indonesia.

Based on the background described above, the purpose of this study is to evaluate the performance of the Zero Inflated Poisson (ZIP) model using the Bayesian and MLE approaches in estimating the parameters of count data with a high proportion of zeros. This research is expected to contribute in developing a more accurate parameter estimation method for count data with excess of zero, especially in the field of public health.

METHODS

Data

The data used in this study is secondary and simulation study. The secondary data used is the number of measles cases in East Java in 2022. There are 38 cities/districts with four predictor variables, population size (X_1), percentage of vaccination (X_2), percentage of poor people (X_3), and percentage of proper sanitation (X_4) with the response variable being the number of measles cases in East Java.

Overdispersion and Excess Zero

The overdispersion is a condition that occurs in data that has a variance value greater than the average[13]. Overdispersion occurs when there is a positive correlation between responses or *excess variation* between probabilities or number of responses and there is a violation of assumptions in the data for example when previous events affect the existence of current events [14]. To see whether the data is overdispersed or not, you can use the *Chi-Square* value, if the *Chi-square* value divided by the degree of freedom is greater than 1, it is said that the model has overdispersion [15]. Mathematically written in Equation (1).

$$\chi^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \sim \chi^2_{(n-m)} \tag{1}$$

where, $\hat{\mu}_i = \exp(\hat{\beta}_0 + \sum_{j=1}^{k} x_{ij}\hat{\beta}_j)$, n is number of observation, m is number of parameters $(k+1)$, x_{ij} the value of the j^{th} predictor variable on the i^{th} observation. If $\chi^2/\,(n-p) > 1$ then it can be concluded that observations contain overdispersion.

Excess zero is a condition of many zero values in the response variable [16]. The zero value here has a meaning that cannot be eliminated and must be included in the analysis because they often carry significant information about the underlying data-generating process, particularly in count data scenarios. These zeros can represent two distinct phenomena: structural zeros, which indicate an inherent impossibility of an event occurring, and sampling zeros, which result from the randomness of the process. Neglecting zero values can lead to biased parameter estimates and misinterpretations of the data, especially when zero inflation is present. By incorporating zero values through appropriate statistical models, such as Zero Inflated Poisson or Hurdle models, researchers can better capture the dual processes generating the zeros, leading to more accurate and robust inferences. Excess zero occurs when the proportion of zero values in the response variable is more than 50% [17]. This can cause overdispersion in the regression model.

Zero Inflated Poisson Regression

According to [18], ZIP distribution function as follows.

$$P(Y = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu_i}, & y_i = 0 \\ \dfrac{(1-\omega_i)e_i^{-\mu_i}\mu_i^{y_i}}{y_i!}, & y_i > 0 \end{cases} \tag{2}$$

where ω_i is parameter of zero inflation $0 < \omega_i < 1$ and μ_i is mean value $\mu_i > 0$

According to Lambert in [7], with $Y_i \sim ZIP(\mu, \omega)$ to model generally uses a logit model, i.e. :

$$\omega_i = \frac{\exp x_i^T\gamma}{1+\exp(x_i^T\gamma)} \qquad \mu_i = \exp x_i^T\beta \tag{3}$$

where x_i is matrix for i-th explanatory variable and γ are additional parameter. The model of the relation between μ and ω is as follows [6]:

$$\log(\mu_i) = X_i^T\beta$$

$$\mu_i = \exp X_i^T\beta \tag{4}$$

$$logit\ (\omega_i) = \ln\left(\frac{\omega_i}{1-\omega_i}\right) = X_i^T\gamma \tag{5}$$

Bayesian Zero Inflated Poisson Regression

The ZIP Likelihood function is :

$$L(\boldsymbol{\beta},\boldsymbol{\gamma}|y_i) = \prod_{i:y_i=0}\left[\frac{\exp\,(\boldsymbol{x}_i^T\boldsymbol{\gamma})}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})} + \left(\frac{e^{-\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})}\right)\right] + \prod_{i:y_i>0}\left[\left(\frac{e^{-(\exp\,(\boldsymbol{x}_i^T\boldsymbol{\beta})}\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})^{y_i}}{y_i!(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma}))}\right)\right] \tag{6}$$

The normal distribution for $\boldsymbol{\beta}$ while the beta distribution for the prior $\boldsymbol{\gamma}$. Suppose a random variable $\boldsymbol{\beta}$ with normal distribution with mean μ nd variance σ^2 is chosen, it can be expressed as follows:

$$\pi(\beta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\beta-\mu^2)}{2\sigma^2}\right) \tag{7}$$

Meanwhile $\boldsymbol{\gamma}$ the hyper parameters are and can be expressed as follows:

$$\pi(\gamma) = \frac{\Gamma(\delta_0+\eta_0)}{\Gamma(\delta_0)\Gamma(\eta_0)}\left(\frac{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}\right)}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})}\right)^{(\delta_0-1)}\left(\frac{1}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})}\right)^{(\eta_0-1)} \tag{8}$$

Therefore, the product of the likelihood function with the prior distribution can be written as follows:

$$\begin{aligned}\pi(\beta|y) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\beta-\mu^2)}{2\sigma^2}\right) \times \prod_{i:y_i=0}\left[\frac{\exp\,(\boldsymbol{x}_i^T\boldsymbol{\gamma})}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})} + \left(\frac{e^{-\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})}\right)\right] + \\ &\quad \prod_{i:y_i>0}\left[\left(\frac{e^{-(\exp\,(\boldsymbol{x}_i^T\boldsymbol{\beta})}\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})^{y_i}}{y_i!(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma}))}\right)\right] \\ \pi(\gamma|y) &= \frac{\Gamma(\delta_0+\eta_0)}{\Gamma(\delta_0)\Gamma(\eta_0)}\left(\frac{(\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma}))^{(\delta_0-1)}}{(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma}))^{(\delta_0+\eta_0-2)}}\right) \times \prod_{i:y_i=0}\left[\frac{\exp\,(\boldsymbol{x}_i^T\boldsymbol{\gamma})}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})} + \left(\frac{e^{-\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma})}\right)\right] + \\ &\quad \prod_{i:y_i>0}\left[\left(\frac{e^{-(\exp\,(\boldsymbol{x}_i^T\boldsymbol{\beta})}\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})^{y_i}}{y_i!(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\gamma}))}\right)\right]\end{aligned} \tag{9}$$

Bayesian Model Convergence Test

In the Bayes method there are several ways to see the convergence of model parameters, namely Trace plot, Autocorrelation plot, Quantiles plot, Density plot and Monte Carlo Error (MC Error) value [19]. The trace plot is a plot of iterations against the generated value. Convergence is achieved when the trace plot shows a horizontal pattern. The MC Error can be calculated using the following formula [18].

$$MCE[G(\theta)] = \sqrt{\frac{1}{K(K-1)}\sum_{b=1}^{K}\left(\bar{G}(\theta)_b - G(\theta)\right)^2} \tag{11}$$

where, $\bar{G}(\theta)_b$ is the sample mean of each batch, $G(\theta)$ is the general sample mean, K is the number of batches.

Best Selection Model

The criterion used to measure the goodness of the model after obtaining a model is the Root Mean Square Error (RMSE). RMSE is used based on the estimation error. The error shows how much the secondary data estimation results differ from the simulated data estimation values. This value is used to determine which model is the best. The RMSE formula is as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i^{(0)}\right)^2} \quad (12)$$

where, n is the number of simulations or observations, $\hat{\theta}_i$ is Estimation of parameters in the i-th simulation, and $\theta_i^{(0)}$ is the true value of the parameter in the i-th simulation.

Simulation Study

This simulation data refers to processed data obtained from secondary data obtained from the 2022 Health Office profile. Data generation using *R software* carried out on this simulation data consists of various sample sizes, namely 38, 100, and 300 and the proportion of zero values, namely 0.6, 0.7, and 0.8. Simulations in various sample sizes variations are used to determine the performance of *Zero Inflated Poisson* Bayesian regression in handling data with *overdispersion* properties at various sample sizes and zero value proportion levels.

RESULTS AND DISCUSSION

Overdispersion and Excess Zero

Based on the results of the overdipersion test obtained that $\chi^2/\,(n-p) > 1$ is $5.08 > 1$ therefore it can be said that the data has overdispersion. While the percentage of zero values in the response variable is 60.2% $> 50\%$ then it concluded that the data has excess zero. because the data experiences overdispersion and excess zero so that the Poisson regression model is not suitable for use in modeling, so the ZIP regression model can be used to overcome these problems.

Zero Inflation Poisson Regression with MLE

The results of the parameter estimates from the analysis of the data number of measles used MLE are as the following:

$$\widehat{\mu_i} = \exp(-7{,}331 + 1{,}584X_{1i}) \quad (13)$$

From Equations (12), it can be seen if every 1 person increase in population in East Java will increase the average number of measles cases in Est Java by $\exp(1{,}584) = 4.874 \approx 5$ people.

Zero Inflated Poisson Regression with Bayesian

The results of the parameter estimates from the analysis of the dta number of measles used Bayesian are as the following:

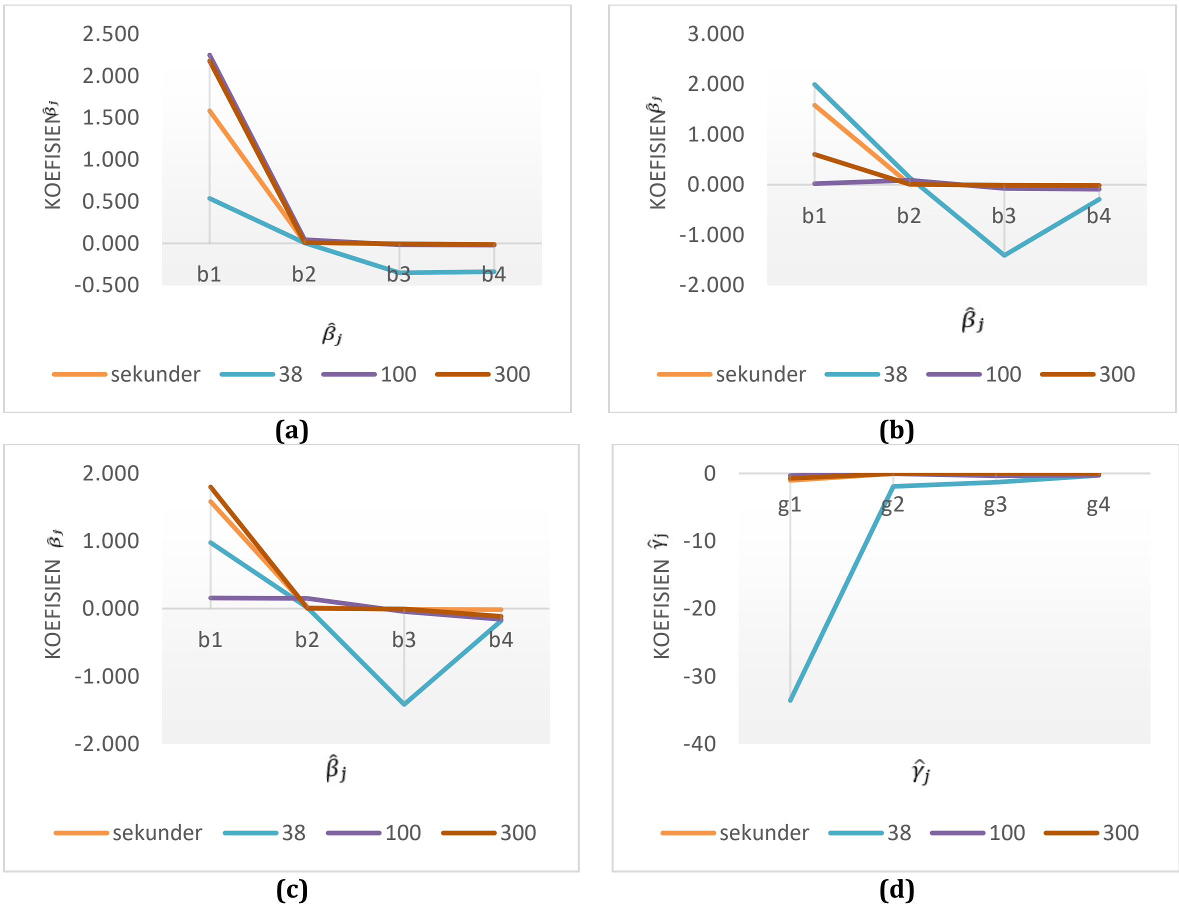
$$\widehat{\mu_i} = \exp(2{,}168X_{1i} - 0{,}092X_{2i} + 0{,}191X_{3i}) \quad (14)$$

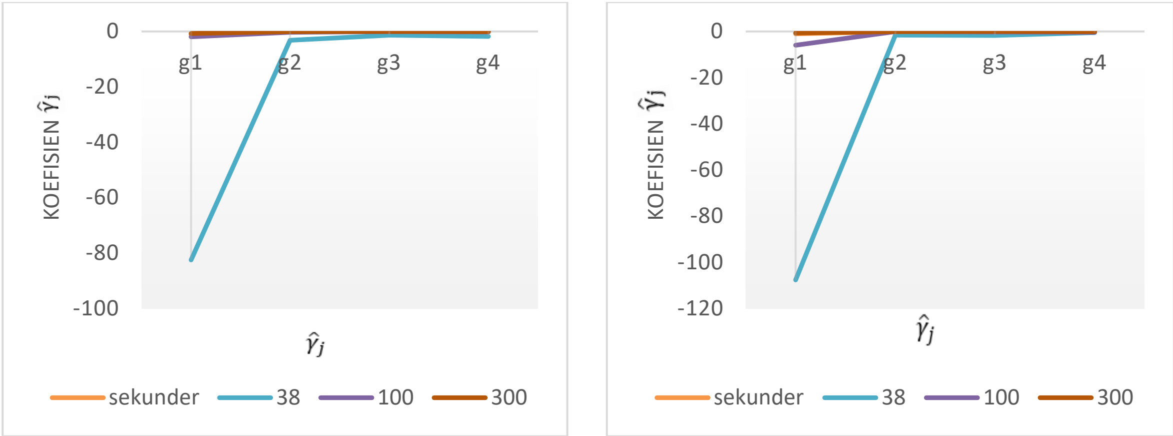
$$\widehat{\omega_i} = \frac{\exp(0{,}5006X_{2i})}{1 + \exp(0{,}5006X_{2i})} \quad (15)$$

From Equation (13), it can be seen if every 1 person increase in population in East Java will increase the average number of measles cases in Est Java by $\exp(2.168) = 8.746 \approx 9$ people, for each 1% increase in percentage of vaccination will decrease the average number of measles cases in East Java by $\exp(-0{,}159) = 0{,}912 \approx 1$ people, and if every 1% increase in percentage of poor people will increase the average number of measles case in Est Java by $\exp(0{,}191) = 1{,}211 \approx 1$ people. While in Equation (14), it can be seen if the vaccination percentage increases by 1%, the log odds of zero inflation probability will increase by 0.5006. This means that every 1% increase in the percentage of poor people will increase the probability of zero inflation by (Odds Ratio-1)×100%=65%. This shows that when the percentage of vaccination increases, it will reduce the probability of measles cases in East Java because measles vaccination can prevent measles from occurring [20].

Simulation Results of Zero Inflated Poisson with MLE

The simulated data is generated based on the original data Zero Inflated Poisson MLE regression parameters. The estimation results of the $\hat{\beta}_j$ and $\hat{\gamma}_j$ parameter can be seen in Figure 1.





(e) **(f)**

Figure 1. Plot of $\hat{\beta}_j$ Parameters Estimation at Various Sample and Proportion of Zero Value (a) p=0.6 (b) p=0.7 (c) p=0.8 and $\hat{\gamma}_j$ Parameters Estimation at Various Sample and Proportion of Zero Value (d) p=0.6 (e) p=0.7 (f) p=0.8

From Figure 1 it can be seen if the estimation results of the parameters $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, and $\hat{\beta}_4$ are relatively more stable at $n = 100$ and 300. The parameter $\hat{\beta}_1$ shows large fluctuations when the sample sizes is small, namely $\hat{\beta}_1 = 0.537$ for $p = 0.6$ while at large sample size, the parameter $\hat{\beta}_1$ tends to stabilize near the initial value of the parameter. This is due to data imbalance in certain scenarios. Parameter estimates at large sample sizes (n=300) become much more stable with consistent coefficients across different scenarios p. This is consistent with statistical theory, where large sample sizes result in more precise parameter estimates. The effect of Proportion zero (p) on parameter estimation when p=0.6 has a relationship that tends to be still clearly visible between predictor variables and responses. However, for p=0.8 it can be seen that the effect of zero inflation is more common compared to the Poisson distribution, as the effect between the predictor and the response is more difficult to detect.

The estimation results of the parameters $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ and $\hat{\gamma}_4$ shows that sample sizes (n) and proportion of zeros (p) have a significant influence on the probability of zero inflation and the stability of the parameter estimates. At small sample sizes (n=38), the parameter estimates show higher fluctuations with the probability of zero inflation varying depending on p. Furthermore, the proportion of zero-inflation (p) also matters, especially at small sample sizes, where higher p (0.7 and 0.8) magnifies the effect of zero-inflation on parameter estimates. However, this effect decreases at larger sample sizes, suggesting that the effect of zero-inflation can be captured more moderately in large sample data such as n=300. Overall, these results suggest that the MLE method requires a sufficiently large sample size to produce stable and precise parameter estimates in the ZIP model.

Simulation Results Zero Inflated Poisson with Bayesian

The simulated data is generated based on the original data Zero Inflated Poisson Bayesian regression parameters. The estimation results of the $\hat{\beta}_j$ and $\hat{\gamma}_j$ parameter can be seen in Figure 2.

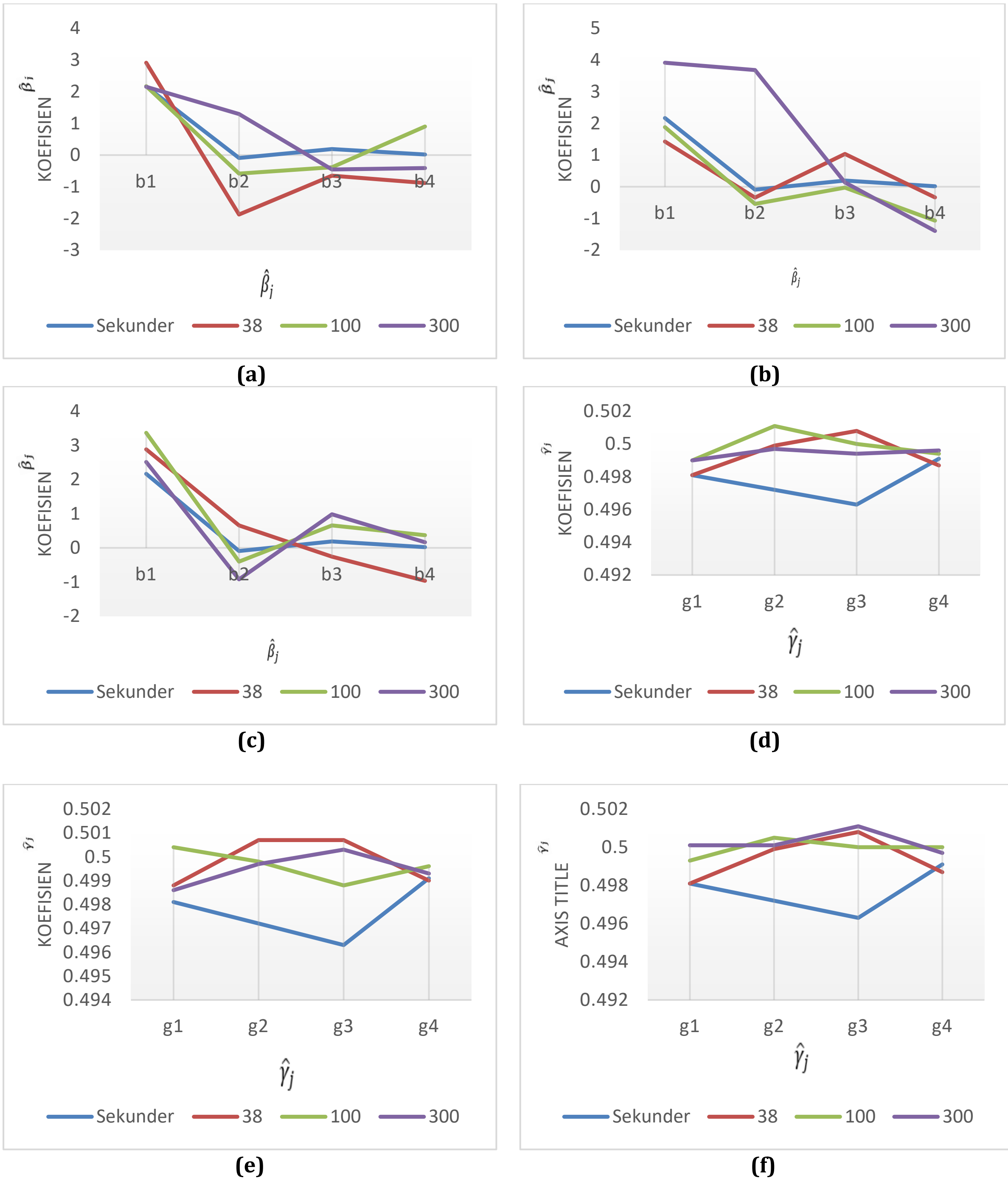


Figure 2. Plot of $\hat{\beta}_j$ Parameters Estimation at Various Sample and Proportion of Zero Value (a) p=0.6 (b) p=0.7 (c) p=0.8 and $\hat{\gamma}_j$ Parameters Estimation at Various Sample and Proportion of Zero Value (d) p=0.6 (e) p=0.7 (f) p=0.8

From Figure 2, It can be seen that at small sample sizes (n=38) the parameter values $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, and $\hat{\beta}_4$ tend to fluctuate more. For example, at p=0.7, the value of $\hat{\beta}_1$=1.427 is lower than p=0.6 and p=0.8. This reflects the instability of the estimate in small samples. While at the medium sample sizes (n=100) and large sample sizes (n=300) the parameters become more consistent. The effect of a low proportion of zero (p=0.6) tends to be more stable on each parameter by showing small variations. At a zero proportion of p=0.7 there is an increase in the value of parameters such as $\hat{\beta}_1$ as well as the variability of parameters $\hat{\beta}_3$ and $\hat{\beta}_4$ for example at n=300, $\hat{\beta}_1$=3.915 it shows that there is a significant increase.

The parameter values $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ and $\hat{\gamma}_4$ shows consistent stability close to 0.5 in all combinations of zero proportion and sample size. For small sample sizes (n=38), the value of $\hat{\gamma}_j$ has a very small variance as shown in Figures 2d, 2e, and 2f, indicating that the estimation for the zero-inflation component is relatively accurate even though the sample size is small. In the medium sample size (n=100), the value of $\hat{\gamma}_j$ becomes more stable and in the large sample size (n=300) the stability of the parameter is getting more visible. This shows that increasing the sample size will increase the precision of the estimation. Moreover, at low proportion of zeros (p=0.6), the parameter value shows consistency close to 0.5 across all sample sizes. This indicates that the model can detect low proportion of zeros well, while at medium proportion of zeros (p=0.7) and high proportion of zeros (p=0.8) the parameter value $\hat{\gamma}$ remains stable which indicates that the model can accommodate data sparsity with high proportion of zeros.

Best Selection Models

To determine the goodness between one method and another, Root Mean Square Error (RMSE) is used. The method that has the smallest RMSE is the best method. The following is the RMSE of each method.

Table 1. RMSE Results of Each Model

n	Proporsi Nol	RMSE	
		MLE	Bayesian
38	0,6	111,90	0,54
	0,7	191,22	0,79
	0,8	188,96	0,74
100	0,6	0,91	0,83
	0,7	5,91	0,68
	0,8	24,70	0,85
300	0,6	0,53	0,79
	0,7	0,49	0,65
	0,8	0,56	0,73

Based on Table 1, it can be seen that at small and medium sample sizes the Bayesian method has a small RMSE value than the MLE, while at large sample sizes the MLE method has a small RMSE value (0.49 to 0.56) than the Bayesian method. this indicates that the MLE method becomes more accurate with increasing sample size and the Bayesian method still produces small RMSE values (0.65 to 0.79) which indicates that this method maintains its performance even though the sample sizes increases.

CONCLUSIONS

Based on the research results, it can be concluded that the performance of the Zero Inflated Poisson (ZIP) regression estimation analysis shows that the Bayesian method has the advantage of producing more accurate parameter estimates on small sample sizes and data conditions with a high proportion of zero values. While the MLE method tends to be sensitive to sample size and its performance increases significantly at large sample sizes. Therefore, for ZIP analysis on data with a small sample sizes or a high proportion of zeros, the Bayesian method is more recommended due to its better consistency.

REFERENCES

[1] A. Fitrianto, "A Study of Count Regression Models for Mortality Rate," *CAUCHY*, vol. 7, no. 1, pp. 142–151, Nov. 2021, doi: https://doi.org/10.18860/ca.v7i1.13642.

[2] Tendriyawati, "Pemodelan Regresi Poisson terhadap Faktor-Faktor yang Mempengaruhi Terjadinya Hipertensi di Kota Kendari," *Jurnal Matematika, Komputasi dan statistika*, Vol. 3 No. 1, 2023, https://doi.org/10.33772/jmks.v3i1.35.

[3] A. Rahayu, "Model-Model Regresi untuk Mengatasi Masalah Overdipersi pada Regresi Poisson," *Peqguruang*, vol. 2, no. 1, p. 1, Jun. 2021, doi: https://doi.org/10.35329/jp.v2i1.1866.

[4] A. D. Chaniago and S. P. Wulandari, "Pemodelan Generalized Poisson Regression (GPR) dan Negative Binomial Regression (NBR) untuk Mengatasi Overdispersi pada Jumlah Kematian Bayi di Kabupaten Probolinggo," *Jurnal Sains dan Seni ITS*, vol. Vol. 11 No. 6, 2022, http://dx.doi.org/10.12962/j23373520.v11i6.93240.

[5] Cahyandari, "Pengujian Overdispersi pada Model Regresi Poisson," *EJournal Unisba*, vol. Vol. 14 No.2, pp. 69–76, Nov. 2014, https://doi.org/10.29313/jstat.v14i2.1204.

[6] D. Lambert, "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing," *Technometrics*, vol. 34, no. 1, p. 1, Feb. 1992, doi: https://doi.org/10.2307/1269547.

[7] C. X. Feng, "A comparison of zero-inflated and hurdle models for modeling zero-inflated count data," *Feng Journal of Statistical Distributions and Applications*, vol. Vol. 8 No. 8, 2021, https://doi.org/10.1186/s40488-021-00121-4.

[8] H. He, H. Zhang, P. Ye, and W. Tang, "A test of inflated zeros for Poisson regression models," *Stat Methods Med Res*, vol. 28, no. 4, pp. 1157–1169, Apr. 2019, doi: https://doi.org/10.1177/0962280217749991.

[9] E. D. Ginting and Sutarman, "Penaksiran Parameter Regresi Poisson Dengan Maximum Likelihood," *Indonesian Journal of Multidiciplinary*, vol. Vol 1 No. 6, 2023, doi: https://journal.csspublishing.com/index.php/ijm/article/view/513.

[10] D. A. Pradana and T. E. Lestari, "Estimasi Parameter Regresi Zero-Inflated Negative Binomial dengan Metode Algoritma Expectation Maximization (EM) (Studi Kasus: Penyakit Difteri di Jawa Barat Tahun 2016)," *JKMA*, vol. 1, no. 1, p. 18, Jun. 2020, doi: http://dx.doi.org/10.17977/um055v1i12020p18-26.

[11] J. Albert, "Bayesian Computation with R". New York, NY: Springer New York, 2009. doi: https://doi.org/10.1007/978-0-387-92298-0.

[12] S. E. Fienberg and A. Rinaldo, "Maximum likelihood estimation in log-linear models," *Ann. Statist.*, vol. 40, no. 2, Apr. 2012, doi: http://dx.doi.org/10.1214/12-AOS986.

[13] J. M. Hilbe, *Modeling Count Data*. England: Cambridge University Press, 2014.

[14] V. Landsman, D. Landsman, C. S. Li, and H. Bang, "Overdispersion models for correlated multinomial data: Applications to blinding assessment," *Statistics in Medicine*, vol. 38, no. 25, pp. 4963–4976, Nov. 2019, doi: https://doi.org/10.1002/sim.8344.

[15] J. R. Wilson, "Chi-Square Tests for Overdispersion with Multiparameter Estimates," *Applied Statistics*, vol. 38, no. 3, p. 441, 1989, doi: https://doi.org/10.2307/2347732.

[16] J. Haslett, A. C. Parnell, J. Hinde, and R. De Andrade Moral, "Modelling Excess Zeros in Count Data: A New Perspective on Modelling Approaches," *Int Statistical Rev*, vol. 90, no. 2, pp. 216–236, Aug. 2022, doi: https://doi.org/10.1111/insr.12479.

[17] F. Famoye and J. S. Preisser, "Marginalized zero-inflated generalized Poisson regression," *Journal of Applied Statistics*, vol. 45, no. 7, pp. 1247–1259, May 2018, doi: http://dx.doi.org/10.1080/02664763.2017.1364717

[18] I. Ntzoufras, A. Katsis, and D. Karlis, "Bayesian Assessment of the Distribution of Insurance Claim Counts Using Reversible Jump MCMC," *North American Actuarial Journal*, vol. 9, no. 3, pp. 90–108, Jul. 2005, doi: https://doi.org/10.1080/10920277.2005.10596213.
[19] A. C. Delima, F. Yanuar, and H. Yozza, "Penerapan Metode Regresi Logistik Ordinal Bayesian untuk Menentukan Tingkat Partisipasi Politik Masyarakat Kota Padang," *Jurnal Matematika UNAND*, vol. Vol 8 No. 3, pp. 1–8, 2019, doi: http://dx.doi.org/10.25077/jmu.8.3.1-8.2019.
[20] H. Hamzah and L. Y. Hendrati, "Kasus Campak pada Kasus Campak yang divaksinasi Menurut Provinsi di Indonesia," *Jurnal Ilmiah Permas: Jurnal Ilmiah STIKES Kendal*, vol. Vol. 13 No.1, Jan. 2023, doi: https://doi.org/10.32583/pskm.v13i1.487