Journal of Applied Data Sciences
Vol. 7, No. 2, May 2026, pp. 1545-1557

On the Use of Zero-Inflated Mixed Models for Count Data: A Simulation and Empirical Evidence

Anang Kurnia[1,*], Zafira Fakhriyah[2], Kusman Sadik[3] Dian Handayani[4,*]

[1,3]*School of Data Science, Mathematics, and Informatics - IPB University, Indonesia*

[2]*WPP Agency - Jakarta, Indonesia*

[4]*Study Program of Statistics - Universitas Negeri Jakarta, Indonesia*

(Received: November 10, 2025; Revised: January 15, 2026; Accepted: April 1, 2026; Available online: May 3, 2026)

Abstract

This paper evaluates the performance of classical count regression models (Poisson, Negative Binomial, Generalized Poisson), zero-inflated models (Zero-Inflated Poisson/ZIP, Zero-Inflated Negative Binomial/ZINB, Zero-Inflated Generalized Poisson/ZIGP), and zero-inflated mixed models (ZIPMM, ZINBMM, ZIGPMM) for over-dispersed count data, particularly due to excess zeros and unobserved heterogeneity. Using simulation and empirical studies, we evaluated the performance of the models based on their predictive capability and their ability to yield valid inferences through hypothesis testing. The simulation, replicated 1000 times, involves 27 scenarios that combine various sample sizes, proportions of zero counts, and response variable distributions. Our findings indicate that ZIGPMM and ZINBMM provide the smallest root mean square error (RMSE) values. Although the Poisson model yields a relatively small RMSE, it does not adequately account for overdispersion, leading to underestimated standard errors and potentially misleading significance tests. The negative binomial model yields dispersion estimates closest to 1, indicating good performance, whereas ZIGP, ZINB, ZIGPMM, and ZINBMM perform better when zero counts are extremely high. Empirical analysis of data on under-five mortality due to pneumonia in Java Island, Indonesia, indicates that ZINB, ZINBMM, and ZIGPMM have the smallest Akaike Information Criterion (AIC), making them the most suitable models. These models show that exclusive breastfeeding and vitamin A have no significant effect on under-five child mortality due to pneumonia, while severe malnutrition has a statistically significant impact ($\alpha = 0.05$).

Keywords: Mixed Model, Pneumonia, Poisson, Negative Binomial, Zero-Inflated

1. Introduction

Count data refers to non-negative integer values that typically represent the frequency of occurrences within a specific time or place [1]. The simplest regression model for modeling count data is the Poisson regression model. Poisson regression relies on the equidispersion assumption, which requires the mean and variance of the response variable to be equal [2]. In practice, count responses often exhibit larger variance than their mean, a phenomenon known as overdispersion [3]. Overdispersion violates the equidispersion assumption of the Poisson model and can arise due to excess zero responses or unobserved heterogeneity [4].

Some count regression models that can accommodate overdispersion include the negative binomial and generalized Poisson regression models, as demonstrated by [5] and [6]. In cases that overdispersion is primarily due to excess or abundance of zero responses, the zero-inflated models are commonly used ([7], [8], [9], [10]). Furthermore, to manage overdispersion resulting from both excess zero responses and unobserved heterogeneity, the zero-inflated mixed model can be employed ([11], [12], [13], [14]).

This paper aims to evaluate various count regression models for handling overdispersion data through simulations and empirical studies. The overdispersion phenomenon observed in our study is caused by excess zeros and unobserved heterogeneity. The best-selected model will be applied to identify significant factors associated with the under-five mortality rate from pneumonia in Java Island, Indonesia, using 2021 data. According to the 2021 Indonesian Health Profile, some provinces in the Java Islands, i.e. East Java, Central Java, and West Java, recorded the highest numbers of under-five child mortality due to pneumonia [15]. It means that identifying significant factors contributing to under-

*Corresponding author: Dian Handayani (dianh@unj.ac.id), Anang Kurnia (anangk@apps.ipb.ac.id)
DOI: https://doi.org/10.47738/jads.v7i2.1303

five child mortality due to pneumonia is crucial. This study is expected to assist policymakers in developing targeted and effective programs to reduce child mortality due to pneumonia on the island of Java.

2. Literature Review

2.1. Poisson Regression Model

Poisson regression is the most commonly used generalized linear model (GLM) for modeling data with a count response variable. In the Poisson regression model, the response variable Y_i is assumed to follow a Poisson distribution with a single parameter mean $\mu > 0$ as follows [16]:

$$f(y_i \mid \mu_i) = \frac{e^{-\mu}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0,1,2,\dots; i = 1,2\dots n \tag{1}$$

n is the number of observations (sample size). Because the response variable Y_i follows a Poisson distribution, it can be shown that the mean and the variance of Y_i are equal [17].

Poisson regression models, as well as other count regression models, are generally GLMs with a logarithm link function, which can be expressed as follows:

$$\log \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} = \boldsymbol{x}_i^T \boldsymbol{\beta}, i = 1,2\dots n. \tag{2}$$

The maximum likelihood estimates (MLE) for β are obtained by maximizing the likelihood (or log-likelihood) function [18]. The solution for the estimator of β that maximizes the likelihood (or log-likelihood) function can be obtained using the Newton–Raphson or Fisher Scoring iterative methods. Based on model (2), the mean μ_i is given by:

$$\mu_i = \exp(x_i^T \beta) = e^{x_i^T \beta} \tag{3}$$

$x_i = (1, x_{i1}, x_{i2}, \dots x_{ip})^T$; x_{ij} is the values of $j -$ explanatory variable in i^{th} observation, $i = 1,2\dots n$; $j = 1,2\dots p$; $\beta = (\beta_0, \beta_1, \beta_2, \dots \beta_p)^T$; β_0 is intercept, $\beta_j, j = 1,2\dots p$ is slope/coefficient model which is corresponding with the explanatory variable x_j.

The log likelihood function for obtaining the MLE of β based on n random sample and the observed response $y_1, y_2, \dots y_n$ from the Poisson distribution is given by:

$$\log L(\beta | y_i) = \sum\left[y_i {x_i}^T \beta - e^{{x_i}^T \beta} - \ln(y_i!)\right], \; i = 1,2,3,\dots n. \tag{4}$$

Overdispersion in count response data can be indicated by the actual variance of the response variable being greater than the mean [19]. Cameron and Trivedi [20] suggest the dispersion statistics $\hat{\phi}$ to detect the overdispersion as follows:

$$\hat{\phi} = \frac{\text{Pearson Chisquare } (\chi^2)}{\text{degree of freedom}} \tag{5}$$

If the dispersion statistics $\hat{\phi}$ is equal to one, it suggests that the equidispersion is hold. The statistic $\hat{\phi}$ greater than one indicates the presence of overdispersion. Some common causes of overdispersion include unobserved heterogeneity, outliers, data clustered, and an excessive number of zero responses [21].

2.2. Negative Binomial Regression Model

Negative Binomial Regression (NBR) model is an alternative regression model used to handle count data with overdispersion. It assumes that the response variable Y_i follows negative binomial distribution. The probability mass function for Y_i which follows negative binomial is defined by [19]:

$$f(y_i \mid \mu, \alpha) = \frac{\Gamma(y_i + 1/\alpha_i)}{\Gamma(y_i + 1)\,\Gamma(1/\alpha_i)} \left(\frac{1}{1 + \alpha_i \mu_i}\right)^{1/\alpha_i} \left(1 - \frac{1}{1 + \alpha_i \mu_i}\right)^{y_i}, y_i = 0,1,2,\dots;\; \mu_i > 0, \alpha_i > 0 \tag{6}$$

μ is rate/mean parameter and α is dispersion parameter.
Based on (6), the mean $E(Y_i)$ and variance $Var(Y_i)$ of the response variable Y_i which follows negative binomial distribution is given by:

$$E(Y_i) = \mu_i \tag{7}$$

$$Var(Y_i) = \mu_i + \alpha_i \mu_i^2$$

Based on (7), if $\alpha_i > 0$, $Var(Y_i)$ will be greater than $E(Y_i)$, which allows the negative binomial distribution to handle data overdispersion. If $\alpha_i = 0$ then $Var\ (Y_i) = E(Y_i)$, which is the same as equidispersion characteristics in Poisson distribution. Similar to the Poisson regression model, the negative binomial regression model is also a GLM with a logarithm link function. The negative binomial regression model is also represented by (2). The MLE for $\boldsymbol{\beta}$ are obtained by maximizing the log-likelihood function [19] as follows:

$$\log L(\boldsymbol{\beta}|\boldsymbol{y},\alpha) = \sum_{i=1}^{n} y_i \ln\left(\frac{\alpha e^{x_i^T\boldsymbol{\beta}}}{1+\alpha e^{x_i^T\boldsymbol{\beta}}}\right) - \frac{1}{\alpha}\ln\left(1+\alpha e^{x_i^T\boldsymbol{\beta}}\right) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha}\right) \quad (8)$$

2.3. Generalized Poisson Regression Model

The Generalized Poisson regression (GPR) model is one of the count regression models that can also be represented by (2). The GPR model assumes that the response variable Y_i follows a generalized Poisson distribution, which its probability function can be written as follows [22], [23]:

$$f(y_i \mid \mu_i, \theta_i) = \left(\frac{\mu_i}{1+\theta_i\mu_i}\right)^{y_i} \frac{(1+\theta_i y_i)^{y_i-1}}{y_i!} exp\left(\frac{-\mu_i(1+\theta_i y_i)}{1+\theta_i\mu_i}\right), y_i = 0,1,2\ldots \quad (9)$$

The mean $E(Y_i)$ and variance $Var(Y_i)$ based on (9) is given by

$$E(Y_i) = \mu_i \quad (10)$$
$$Var(Y_i) = \mu_i(1+\theta_i\mu_i)^2$$

$\mu_i = \exp(x_i^T\beta) = e^{x_i^T\beta}, i = 1,2\ldots n$. Based on (10), if $\theta_i = 0$, then $Var\ (Y_i) = E(Y_i) = \mu_i$. It means that the response variable Y_i will have equidispersion characteristic. In other words, for $\theta_i = 0$, the generalized Poisson distribution will reduce to Poisson distribution. If $\theta_i > 0$ then $Var\ (Y_i) > E(Y_i)$. It means that the Generalized Poisson distribution could overcome overdispersion data. Otherwise, if $\theta_i < 0$, the Generalized Poisson distribution could accomodate underdispersion data. The log-likelihood function for obtaining the MLE of β in the GPR model is given by [22]:

$$\log L(\boldsymbol{\beta},\theta) = \sum_{i=1}^{n} y_i \ln\left[\frac{\mu_i}{1+\theta_i\mu_i}\right] + (y_i - 1)\ln(1+\theta_i\mu_i) - \ln(y_i!) - \left(\frac{\mu_i(1+\theta_i\mu_i)}{1+\theta_i\mu_i}\right) \quad (11)$$

2.4. Zero-Inflated and Zero-Inflated Mixed Regression Models

Zero-inflated distribution can be used for modelling excess zero values on response variable Y. The zeros values can be considered as the results of two possible conditions: from a population with all elements being zero (structural zeros) or from a population with positive discrete values including zeros (sampling zeros). The zero-inflated distribution considers that response variable Y come from two processes [24]. The first process describes zeros values, whereas the second process for non-zero values. The first process is also called binary process because it will determine the zeros value come from structural zeros or sampling zeros. On the other hand, the second process is also called count process. For example, the Zero-inflated Poisson distribution assumes that the probability for a response variable Y equal zero and the probability for Y equal to $r = 1,2,3,\ldots$ follow Poisson distribution. The probability function for the zero-inflated Poisson is given by

$$P(Y=0) = \delta + (1-\delta)\exp(-\mu) \quad (12)$$
$$P(Y=r) = \frac{(1-\delta)\exp(-\mu)\mu^r}{r!}, r = 1,2,3,\ldots$$

δ is the probability that zeros values come from structural zeros, $(1-\delta)$ is the probability that zero values come from sampling zeros.

The binary process (which modeling the zeros value as a mixture from structural zeros or sampling zeros) is generally linked to some covariates using logit link function, while the count process (which modeling non-zeros value) is linked to some covariates by logarithm link function. The model for both processes, which is called zero-inflated regression model, can be written by:

$$\log\left(\frac{p}{1+p}\right) = G\gamma \qquad \log(\mu) = B\beta \tag{13}$$

G is covariate matrix for binary process (logistic component) and B is covariate matrix for count process component. Parameter γ and β can be estimated by MLE.

In the mixed-effects framework, the model (13) could be extended by including fixed effects as well as random effects. The zero-inflated mixed regression model retains the two-component process as well as introduces random effects [25] [13]. The zero-inflated mixed regression model is given by:

$$\log\left(\frac{p}{1+p}\right) = G\gamma + u \qquad \log(\mu) = B\beta + v \tag{14}$$

u and v are vector of random effects assumed to follow a normal distribution [26]. The parameter in the zero-inflated mixed model is typically estimated by the expectation-maximization (EM) algorithm as part of the maximum likelihood approach [27].

3. Methodology

To evaluate the performance of our proposed model, zero-inflated mixed model (zero-inflated Poisson mixed model/ZIPMM, zero-inflated negative binomial mixed model/ZINBMM, and zero-inflated generalized Poisson mixed model/ZIGPMM), against the standard count regression models (Poisson and Negative Binomial regression) and zero-inflated models (Zero-Inflated Poisson/ZIP, Zero-Inflated Negative Binomial/ZINB, Zero-Inflated Generalized Poisson/ZIGP), we conduct a simulation study as well as empirical study. The models which are evaluated in the simulation study are fitted to the data under-five child mortality due to pneumonia in 2021 on Java Island, Indonesia. Figure 1 presents the systematic steps involved in conducting our research. By using simulation and empirical studies, we evaluate the model's predictive ability as well as its hypothesis testing performance in producing valid conclusions. The model's predictive ability is evaluated in terms of Root Mean Square Error (RMSE) and relative bias (RB). The performance of hypothesis testing is evaluated in terms of dispersion value estimates ϕ and standard error estimates. On the other hand, to evaluate the overall goodness of fit of the model, we employ the Akaike Information Criterion (AIC).

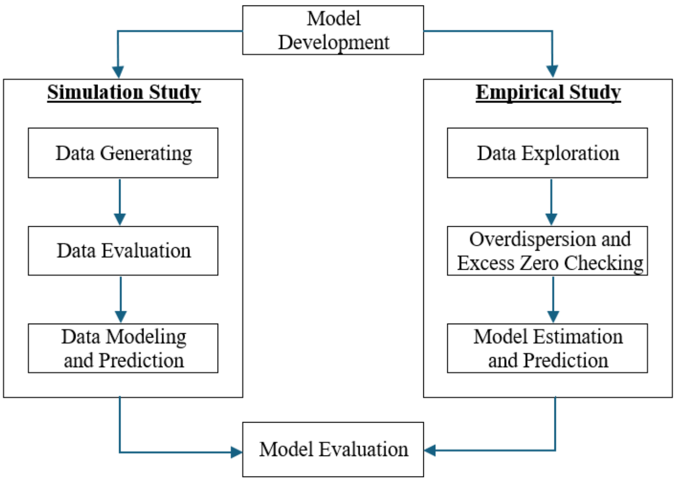

Figure 1. The research workflow

3.1. Simulation Study

In our simulation study, we generate hypothetical data with systematically controlled levels of overdispersion by introducing a zero-inflated data structure and by incorporating random effects to demonstrate the impact of unobserved heterogeneity. The hypothetical data is constructed under 27 different scenarios, based on a combination of three factors: (1) distribution of the response variable Y, i.e Poisson, negative binomial, and generalized Poisson, (2) proportion of zero values (p_0 = 0.2, 0.5, and 0.8), and (3) sample size (n =30,100 and 500).

To construct the hypothetical data, we generate n observations of explanatory variables x_1 and x_2 from a normal distribution: $x_1 \sim N(4,1)$ and $x_2 \sim N(3,1)$. The area-specific random effects u generated from normal distribution $N(0,1)$ to reflect the moderate unobserved heterogeneity across the districts and municipalities in empirical data. We set the regression coefficients $\beta_0 = 1.3$, $\beta_1 = -0.1$, $\beta_2 = 0.5$, representing a baseline mortality risk that decreases with higher x_1 and increases with x_2. Finally, we calculated the mean of Y, denoted by μ, based on the following regression model:

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{15}$$

The values of the response variable Y are generated from a zero-inflated Poisson (μ, p_0), a zero-inflated negative binomial (μ, p_0, α) and zero-inflated generalized Poisson (μ, p_0, α). To generate Y which follows a zero-inflated negative binomial and a zero-inflated generalized Poisson distribution, we set the dispersion parameter $\alpha = 4$. The simulation study was replicated 1,000 times.

We also checked our hypothetical data to ensure that they already satisfy the conditions of overdispersion and excess zeros. Overdispersion is examined by specifying the hypothesis: $H_0: \alpha = 0$ (equidispersion) and $H_1: \alpha > 0$ (overdispersion) [20]. If the test statistic $T_z = \frac{\hat{\alpha}}{\sqrt{Var(\hat{\alpha})}} > z_\alpha$ or the $p - value$ is less than the significance level, then the H_0 is rejected, indicating the presence of overdispersion in the data. The presence of excess zeros is examined using the score test with hypotheses $H_0: p_0 = 0$ (no excess zero) dan $H_1: p_0 > 0$ (excess zero) [28]. If the test statistic $S_p = \frac{(n_0 - n\hat{p}_0)^2}{n\hat{p}_0(1-\hat{p}_0) - n\bar{y}\hat{p}_0{}^2} > X^2_{\alpha,1}$ or the $p - value$ is less than the significance level, then the H_0 is rejected, indicating that the data contain an excess of zeros. If the conditions of overdispersion and excess zeros are satisfied, the analysis proceeds to data modeling. However, if either of these conditions is not met, data generating is repeated until both conditions are met.

Our generated data is modeled using nine regression models: three classical count regression models (Poisson, negative binomial, generalized Poisson), three zero-inflated models (zero-inflated Poisson/ZIP, zero-inflated negative binomial/ZINB, zero-inflated generalized Poisson/ZIGP), and three zero-inflated mixed models (zero-inflated Poisson mixed model/ZIPMM, zero-inflated negative binomial mixed model/ZINBMM, and zero-inflated generalized Poisson mixed model/ZIGPMM.

Model performance is evaluated by assessing the model's predictive ability and hypothesis-testing performance. The model's predictive ability is assessed based on the root mean square error (RMSE) and relative bias (RB). The model's performance to produce valid conclusions based on hypothesis testing is assessed using dispersion statistics ϕ and standard error estimates of the regression coefficient. On the other hand, to select the best model for fitting our data, the Akaike Information Criterion (AIC) is used. The formula for relative bias, dispersion statistics, RMSE, and AIC are given by:

$$RB = \frac{|E(\hat{\beta}_j) - \beta_j|}{\beta_j}, j = 1,2 \ldots p\ ;\ \phi = \frac{deviance}{degree\ of\ freedom};\ RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}};\ AIC = -2\log L + 2p \tag{16}$$

p is the number of estimated parameters in the model, n is the sample size, and L is the value that maximize likelihood function.

3.2. Empirical Study

The nine regression models examined in our simulation study were applied for modeling the empirical data, i.e. the number of under-five mortality cases due to pneumonia in 2021 on Java Island. The observation units comprise 100 districts and municipalities across three provinces on Java Island: West Java, Central Java, and East Java. Our research variables include one response variable Y and four explanatory variables (X_1, X_2, X_3 and X_4). Details for the research variables are shown in table 1.

Table 1. Research Variables

Symbol	Description	Type
Y	Number of under-five children's mortality due to Pneumonia	Random
X_1	Percentage of under-five children receiving exclusive breastfeeding	Fixed
X_2	Percentage of under-five children receiving vitamin A	Fixed

ISSN 2723-6471

X_3	Percentage of under-five children with nutrient-deficient	Fixed
X_4	Province effect	Random

4. Results and Discussion

4.1. Simulation Results

Our simulation studies are based on data generated from 27 scenarios, which are repeated 1000 times. The average RMSE for the predictive response values $\hat{y}$ based on the nine regression models shown in figure 2.

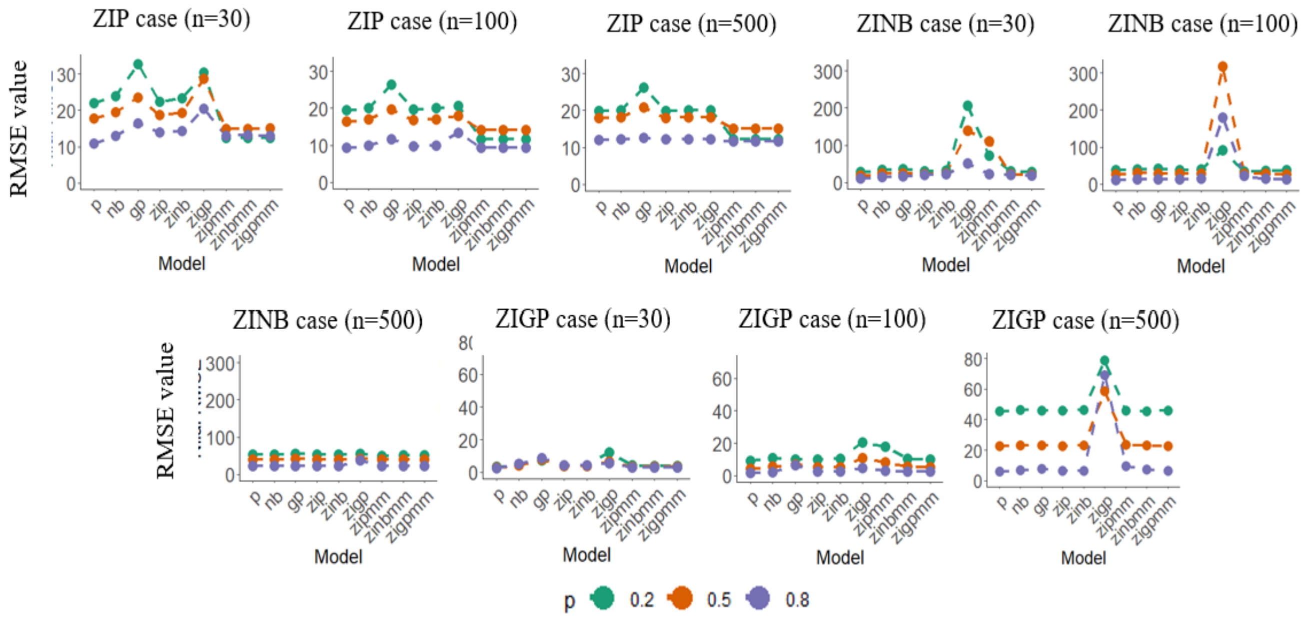

Figure 2. The average RMSE for $\hat{y}$ based on nine regression models in a simulation study

From figure 2, it can be seen the average RMSE for each model is not significantly different in the majority of simulation scenarios, except for the ZIGP regression model, which has the highest RMSE in every scenario. Meanwhile, ZIGPMM, Poison, and ZINBMM have the smallest RMSE in most scenarios.

Table 2. The relative bias for the estimates of β_1 and β_2 based on 27 simulation scenarios

n	Model	p_0= 0.2			p_0 = 0.5			p_0 = 0.8		
		ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP
30	P	0.014[a]0.047[b]	0.047[a]0.173[b]	0.008[a]0.465[b]	0.040[a]0.076[b]	0.033[a]0.174[b]	0.075[a]0.516[b]	0.103[a]0.169[b]	0.127[a]0.295[b]	0.426[a]0.439[b]
	NB	0.021[a]0.019[b]	0.046[a]0.105[b]	0.055[a]0.395[b]	0.027[a]0.080[b]	0.002[a]0.127[b]	0.047[a]0.498[b]	0.101[a]0.214[b]	0.140[a]0.272[b]	0.471[a]0.375[b]
	GP	0.120[a]0.204[b]	0.723[a]0.755[b]	0.567[a]0.077[b]	0.920[a]0.742[b]	1.153[a]1.387[b]	1.169[a]0.428[b]	2.504[a]1.280[b]	2.500[a]1.368[b]	1.712[a]0.929[b]
	ZIP	0.000[a]0.044[b]	0.002[a]0.157[b]	0.190[a]0.407[b]	0.003[a]0.017[b]	0.066[a]0.126[b]	0.161[a]0.435[b]	0.017[a]0.010[b]	0.049[a]0.162[b]	0.499[a]0.111[b]
	ZINB	0.002[a]0.006[b]	0.005[a]0.105[b]	0.176[a]0.416[b]	0.019[a]0.017[b]	0.075[a]0.076[b]	0.192[a]0.426[b]	0.003[a]0.013[b]	0.047[a]0.107[b]	0.527[a]0.165[b]
	ZIGP	0.012[a]0.037[b]	0.017[a]0.011[b]	0.046[a]0.375[b]	0.008[a]0.034[b]	0.027[a]0.113[b]	0.169[a]0.525[b]	0.119[a]0.069[b]	0.058[a]0.016[b]	1.100[a]0.845[b]
	ZIPMM	0.001[a]0.000[b]	0.014[a]0.089[b]	0.197[a]0.340[b]	0.001[a]0.010[b]	0.061[a]0.205[b]	0.164[a]0.337[b]	0.019[a]0.012[b]	0.028[a]0.100[b]	0.322[a]0.519[b]
	ZINBMM	0.001[a]0.002[b]	0.006[a]0.079[b]	0.211[a]0.351[b]	0.001[a]0.016[b]	0.085[a]0.046[b]	0.151[a]0.337[b]	0.015[a]0.036[b]	0.071[a]0.077[b]	0.305[a]0.547[b]
	ZIGPMM	0.001[a]0.000[b]	0.037[a]0.154[b]	0.218[a]0.372[b]	0.002[a]0.012[b]	0.104[a]0.046[b]	0.201[a]0.352[b]	0.020[a]0.031[b]	0.077[a]0.070[b]	0.344[a]0.287[b]
100	P	0.002[a]0.013[b]	0.039[a]0.078[b]	0.121[a]0.492[b]	0.000[a]0.030[b]	0.044[a]0.126[b]	0.106[a]0.503[b]	0.024[a]0.103[b]	0.072[a]0.188[b]	0.198[a]0.504[b]
	NB	0.010[a]0.007[b]	0.013[a]0.017[b]	0.132[a]0.440[b]	0.003[a]0.027[b]	0.039[a]0.038[b]	0.111[a]0.517[b]	0.016[a]0.115[b]	0.049[a]0.131[b]	0.207[a]0.486[b]
	GP	0.032[a]0.039[b]	0.127[a]0.682[b]	0.093[a]0.531[b]	0.089[a]0.277[b]	0.104[a]0.939[b]	0.198[a]0.714[b]	0.082[a]1024[b]	0.127[a]0.792[b]	0.703[a]0.283[b]
	ZIP	0.002[a]0.027[b]	0.047[a]0.136[b]	0.153[a]0.413[b]	0.001[a]0.034[b]	0.047[a]0.135[b]	0.138[a]0.363[b]	0.016[a]0.028[b]	0.058[a]0.081[b]	0.276[a]0.427[b]

ISSN 2723-6471

n	Model	p_0= 0.2			p_0 = 0.5			p_0 = 0.8		
		ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP
	ZINB	0.014[a]0.003[b]	0.005[a]0.054[b]	0.131[a]0.453[b]	0.011[a]0.006[b]	0.037[a]0.065[b]	0.125[a]0.343[b]	0.003[a]0.019[b]	0.065[a]0.028[b]	0.295[a]0.430[b]
	ZIGP	0.020[a]0.011[b]	0.065[a]0.034[b]	0.375[a]0.331[b]	0.007[a]0.015[b]	0.160[a]0.017[b]	0.837[a]0.105[b]	0.011[a]0.012[b]	0.082[a]0.148[b]	0.896[a]0.618[b]
	ZIPMM	0.000[a]0.001[b]	0.046[a]0.097[b]	0.126[a]0.433[b]	0.000[a]0.001[b]	0.050[a]0.108[b]	0.101[a]0.333[b]	0.009[a]0.010[b]	0.014[a]0.044[b]	0.223[a]0.369[b]
	ZINBMM	0.000[a]0.001[b]	0.014[a]0.047[b]	0.169[a]0.415[b]	0.000[a]0.001[b]	0.054[a]0.056[b]	0.161[a]0.322[b]	0.010[a]0.013[b]	0.048[a]0.012[b]	0.271[a]0.375[b]
	ZIGPMM	0.000[a]0.001[b]	0.043[a]0.211[b]	0.132[a]0.398[b]	0.001[a]0.001[b]	0.063[a]0.172[b]	0.183[a]0.354[b]	0.010[a]0.011[b]	0.041[a]0.021[b]	0.265[a]0.339[b]
500	P	0.001[a]0.003[b]	0.000[a]0.012[b]	0.053[a]0.357[b]	0.003[a]0.001[b]	0.013[a]0.036[b]	0.040[a]0.372[b]	0.006[a]0.021[b]	0.019[a]0.076[b]	0.082[a]0.408[b]
	NB	0.001[a]0.000[b]	0.005[a]0.002[b]	0.068[a]0.330[b]	0.004[a]0.000[b]	0.006[a]0.006[b]	0.046[a]0.381[b]	0.009[a]0.021[b]	0.013[a]0.031[b]	0.046[a]0.389[b]
	GP	0.001[a]0.008[b]	0.015[a]0.107[b]	0.025[a]0.230[b]	0.001[a]0.025[b]	0.045[a]0.237[b]	0.017[a]0.243[b]	0.023[a]0.174[b]	0.342[a]0.847[b]	0.127[a]0.247[b]
	ZIP	0.004[a]0.019[b]	0.008[a]0.086[b]	0.058[a]0.323[b]	0.008[a]0.015[b]	0.004[a]0.099[b]	0.032[a]0.351[b]	0.003[a]0.027[b]	0.003[a]0.116[b]	0.080[a]0.370[b]
	ZINB	0.000[a]0.005[b]	0.002[a]0.010[b]	0.127[a]0.287[b]	0.003[a]0.011[b]	0.012[a]0.013[b]	0.094[a]0.303[b]	0.001[a]0.004[b]	0.016[a]0.021[b]	0.081[a]0.385[b]
	ZIGP	0.001[a]0.004[b]	0.000[a]0.000[b]	0.185[a]0.016[b]	0.005[a]0.009[b]	0.000[a]0.020[b]	0.306[a]0.107[b]	0.001[a]0.003[b]	0.025[a]0.025[b]	0.765[a]0.339[b]
	ZIPMM	0.001[a]0.000[b]	0.006[a]0.068[b]	0.053[a]0.313[b]	0.000[a]0.001[b]	0.001[a]0.082[b]	0.038[a]0.360[b]	0.001[a]0.000[b]	0.001[a]0.092[b]	0.118[a]0.361[b]
	ZINBMM	0.001[a]0.000[b]	0.006[a]0.007[b]	0.075[a]0.319[b]	0.000[a]0.001[b]	0.004[a]0.003[b]	0.057[a]0.338[b]	0.001[a]0.000[b]	0.015[a]0.016[b]	0.092[a]0.326[b]
	ZIGPMM	0.001[a]0.000[b]	0.047[a]0.242[b]	0.134[a]0.445[b]	0.000[a]0.001[b]	0.040[a]0.228[b]	0.163[a]0.398[b]	0.001[a]0.000[b]	0.026[a]0.181[b]	0.170[a]0.298[b]

[a]Relative bias for β_1; [b]Relative bias for β_2

Based on table 2, it is evident that the relative bias of the models does not show a significant difference. The ZIP, ZIGPMM, and ZIPMM appear to have good predictive ability, as they exhibit small relative biases. Furthermore, the Poisson regression model retains good predictive performance, as evidenced by its small relative bias, despite violating the equidispersion assumption. Based on the RMSE and relative bias, it can be concluded that overdispersion does not pose serious problems for predicting coefficient regression models. This is consistent with the explanation in the literature review that, relative to the expected value, overdispersion does not greatly affect estimates of the expected value but does greatly affect estimates of the variance and standard error.

Based on table 3, it can be seen the value of ϕ for Poisson regression model is greater than one, aligning with the violation of the equidispersion assumption. On the other hand, in general, the negative binomial regression model yields ϕ close to one. However, for high proportions of zeros, ZIGP, ZINB, ZIGPMM, and ZINBMM have ϕ closer to one.

Table 3. The average of ϕ based on 27 simulation scenarios

n	Model	p_0= 0.2			p_0 = 0.5			p_0 = 0.8		
		ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP
30	P	21.00	30.40	3.74	20.82	40.35	3.34	14.43	12.93	2.65
	NB	1.31	1.07	0.60	3.27	0.90	0.54	8.72	5.18	0.49
	GP	8.75	6.57	2.38	6.80	5.53	2.14	3.61	2.86	1.91
	ZIP	18.24	17.60	2.16	10.16	25.22	1.88	3.32	2.55	1.65
	ZINB	9.09	6.91	2.23	6.67	5.50	1.96	3.21	2.54	1.73
	ZIGP	8.83	6.55	2.06	6.29	5.46	1.85	2.82	2.19	1.67
	ZIPMM	8.43	8.98	2.34	6.58	16.15	2.05	3.34	2.66	1.79
	ZINBMM	8.82	7.52	2.44	6.88	5.59	2.14	3.48	2.77	1.87
	ZIGPMM	8.75	7.52	2.36	6.76	5.61	2.01	3.42	2.70	1.65
100	P	17.73	40.35	7.66	18.95	29.38	4.00	11.82	11.77	1.68
	NB	1.21	0.90	0.42	0.96	0.65	0.31	3.39	0.53	0.19
	GP	7.52	5.53	1.95	5.71	3.89	1.39	2.82	1.91	0.80

ISSN 2723-6471

n	Model	p_0= 0.2			p_0 = 0.5			p_0 = 0.8		
		ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP
	ZIP	15.38	25.22	3.16	9.70	13.44	1.59	3.45	3.12	0.71
	ZINB	7.36	5.50	1.96	5.33	3.84	1.36	2.57	1.83	0.71
	ZIGP	7.31	5.46	1.87	5.28	3.78	1.27	2.49	1.73	0.66
	ZIPMM	5.94	16.15	2.18	4.60	8.21	1.41	2.44	2.08	0.72
	ZINBMM	6.00	5.59	1.97	4.65	3.91	1.37	2.45	1.86	0.72
	ZIGPMM	5.99	5.61	1.98	4.64	3.91	1.37	2.44	1.86	0.70
500	P	17.58	51.57	31.99	19.85	39.52	15.75	14.41	20.47	3.78
	NB	1.18	0.88	0.37	0.93	0.61	0.24	2.45	0.27	0.11
	GP	7.30	5.36	2.00	5.57	3.73	1.35	2.71	1.77	0.60
	ZIP	14.86	33.88	16.15	10.02	20.43	6.43	4.28	7.04	1.08
	ZINB	7.00	5.28	2.05	5.12	3.64	1.36	2.45	1.71	0.60
	ZIGP	6.98	5.27	1.99	5.11	3.64	1.33	2.45	1.71	0.58
	ZIPMM	5.29	25.20	11.38	4.10	15.13	4.26	2.11	4.83	0.74
	ZINBMM	5.29	5.21	2.03	4.10	3.62	1.35	2.10	1.71	0.59
	ZIGPMM	5.29	5.26	2.01	4.10	3.64	1.35	2.10	1.72	0.59

According to table 4, it is also evident that Poisson regression consistently has the smallest standard error estimates of regression coefficients, both for the standard error of β_1 and β_2, in all conditions. This suggests an underestimation of the standard error estimates, indicating they are smaller than they should be. Underestimating the standard error will inflate the resulting test statistic. The larger the test statistic, the greater the tendency to falsely reject H_0, that is, to conclude that an explanatory variable has a significant effect on the response variable, even though its effect may in fact be insignificant. Based on the value of ϕ and the standard error estimates for the Poisson regression model, it is clear that the overdispersion affects the validity of inference under hypothesis-testing.

Table 4. The Mean of Standard Error Estimates of β_1 and β_2 based on 27 simulation scenarios

n	Model	p_0= 0.2			p_0 = 0.5			p_0 = 0.8		
		ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP
30	P	0.047[a]0.040[b]	0.059[a]0.051[b]	0.200[a]0.177[b]	0.062[a]0.053[b]	0.084[a]0.074[b]	0.222[a]0.196[b]	0.100[a]0.087[b]	0.126[a]0.111[b]	0.263[a]0.277[b]
	NB	0.240[a]0.224[b]	0.398[a]0.371[b]	0.550[a]0.507[b]	0.278[a]0.255[b]	0.389[a]0.355[b]	0.591[a]0.544[b]	0.260[a]0.256[b]	0.395[a]0.374[b]	0.669[a]0.689[b]
	GP	0.306[a]0.321[b]	1.231[a]1.474[b]	2.617[a]3.176[b]	1.322[a]1.568[b]	3.201[a]3.663[b]	3.627[a]4.093[b]	6.976[a]7.256[b]	7.347[a]7.693[b]	4.766[a]5.050[b]
	ZIP	0.048[a]0.042[b]	0.064[a]0.059[b]	0.332[a]0.328[b]	0.067[a]0.062[b]	0.109[a]0.104[b]	0.396[a]0.385[b]	0.150[a]0.156[b]	0.237[a]0.247[b]	0.482[a]0.512[b]
	ZINB	0.187[a]0.177[b]	0.348[a]0.336[b]	0.397[a]0.396[b]	0.216[a]0.209[b]	0.365[a]0.355[b]	0.432[a]0.423[b]	0.243[a]0.252[b]	0.373[a]0.385[b]	0.541[a]0.543[b]
	ZIGP	0.168[a]0.151[b]	0.255[a]0.233[b]	0.272[a]0.260[b]	0.157[a]0.142[b]	0.202[a]0.185[b]	0.289[a]0.271[b]	0.132[a]0.123[b]	0.164[a]0.153[b]	0.331[a]0.406[b]
	ZIPMM	0.069[a]0.086[b]	0.144[a]0.183[b]	0.389[a]0.382[b]	0.121[a]0.143[b]	0.268[a]0.312[b]	0.426[a]0.415[b]	0.226[a]0.249[b]	0.375[a]0.394[b]	0.542[a]0.535[b]
	ZINBMM	0.071[a]0.089[b]	0.322[a]0.318[b]	0.399[a]0.398[b]	0.127[a]0.147[b]	0.348[a]0.345[b]	0.434[a]0.422[b]	0.228[a]0.247[b]	0.376[a]0.389[b]	0.508[a]0.521[b]
	ZIGPMM	0.059[a]0.075[b]	0.258[a]0.250[b]	0.312[a]0.312[b]	0.089[a]0.111[b]	0.301[a]0.290[b]	0.318[a]0.312[b]	0.194[a]0.215[b]	0.332[a]0.346[b]	0.271[a]0.277[b]
100	P	0.023[a]0.027[b]	0.024[a]0.029[b]	0.075[a]0.089[b]	0.029[a]0.035[b]	0.032[a]0.039[b]	0.108[a]0.127[b]	0.050[a]0.059[b]	0.060[a]0.072[b]	0.181[a]0.201[b]
	NB	0.118[a]0.145[b]	0.222[a]0.274[b]	0.355[a]0.438[b]	0.178[a]0.211[b]	0.295[a]0.364[b]	0.409[a]0.504[b]	0.222[a]0.272[b]	0.360[a]0.429[b]	0.535[a]0.594[b]
	GP	0.145[a]0.184[b]	0.418[a]0.544[b]	0.741[a]1.028[b]	0.294[a]0.372[b]	0.882[a]1.307[b]	1.098[a]1.706[b]	1.514[a]2.272[b]	2.461[a]3.432[b]	2.497[a]3.328[b]
	ZIP	0.023[a]0.028[b]	0.025[a]0.031[b]	0.093[a]0.119[b]	0.030[a]0.150[b]	0.035[a]0.043[b]	0.152[a]0.197[b]	0.055[a]0.072[b]	0.078[a]0.100[b]	0.359[a]0.437[b]
	ZINB	0.096[a]0.119[b]	0.225[a]0.285[b]	0.350[a]0.446[b]	0.118[a]0.139[b]	0.290[a]0.363[b]	0.371[a]0.478[b]	0.168[a]0.222[b]	0.347[a]0.448[b]	0.437[a]0.530[b]

n	Model	p_0= 0.2			p_0 = 0.5			p_0 = 0.8		
		ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP	ZIP	ZINB	ZIGP
	ZIGP	0.097[a]0.120[b]	0.236[a]0.286[b]	0.305[a]0.383[b]	0.112[a]0.044[b]	0.254[a]0.310[b]	0.245[a]0.305[b]	0.125[a]0.156[b]	0.118[a]0.230[b]	0.251[a]0.294[b]
	ZIPMM	0.026[a]0.031[b]	0.031[a]0.038[b]	0.187[a]0.218[b]	0.037[a]0.044[b]	0.050[a]0.060[b]	0.273[a]0.338[b]	0.087[a]0.111[b]	0.172[a]0.214[b]	0.424[a]0.517[b]
	ZINBMM	0.027[a]0.032[b]	0.223[a]0.274[b]	0.358[a]0.458[b]	0.039[a]0.046[b]	0.284[a]0.350[b]	0.371[a]0.478[b]	0.091[a]0.116[b]	0.336[a]0.430[b]	0.450[a]0.544[b]
	ZIGPMM	0.026[a]0.030[b]	0.164[a]0.197[b]	0.295[a]0.361[b]	0.036[a]0.043[b]	0.219[a]0.262[b]	0.330[a]0.413[b]	0.072[a]0.091[b]	0.299[a]0.378[b]	0.365[a]0.456[b]
500	P	0.011[a]0.012[b]	0.011[a]0.012[b]	0.021[a]0.022[b]	0.014[a]0.015[b]	0.014[a]0.015[b]	0.030[a]0.032[b]	0.023[a]0.024[b]	0.023[a]0.025[b]	0.065[a]0.069[b]
	NB	0.059[a]0.063[b]	0.116[a]0.123[b]	0.221[a]0.235[b]	0.089[a]0.091[b]	0.158[a]0.168[b]	0.274[a]0.291[b]	0.135[a]0.141[b]	0.263[a]0.279[b]	0.390[a]0.414[b]
	GP	0.072[a]0.079[b]	0.209[a]0.222[b]	0.451[a]0.493[b]	0.146[a]0.157[b]	0.309[a]0.330[b]	0.552[a]0.620[b]	0.328[a]0.352[b]	0.649[a]0.783[b]	0.872[a]1.031[b]
	ZIP	0.011[a]0.012[b]	0.011[a]0.012[b]	0.022[a]0.023[b]	0.014[a]0.016[b]	0.014[a]0.016[b]	0.032[a]0.035[b]	0.023[a]0.026[b]	0.024[a]0.027[b]	0.084[a]0.091[b]
	ZINB	0.047[a]0.051[b]	0.120[a]0.128[b]	0.222[a]0.246[b]	0.060[a]0.065[b]	0.162[a]0.174[b]	0.300[a]0.329[b]	0.096[a]0.104[b]	0.260[a]0.280[b]	0.433[a]0.471[b]
	ZIGP	0.050[a]0.055[b]	0.147[a]0.158[b]	0.390[a]0.422[b]	0.063[a]0.069[b]	0.183[a]0.197[b]	0.429[a]0.469[b]	0.096[a]0.105[b]	0.248[a]0.268[b]	0.458[a]0.496[b]
	ZIPMM	0.012[a]0.012[b]	0.012[a]0.013[b]	0.025[a]0.027[b]	0.015[a]0.016[b]	0.015[a]0.017[b]	0.041[a]0.044[b]	0.025[a]0.028[b]	0.028[a]0.031[b]	0.146[a]0.160[b]
	ZINBMM	0.012[a]0.013[b]	0.112[a]0.119[b]	0.279[a]0.302[b]	0.015[a]0.016[b]	0.151[a]0.162[b]	0.346[a]0.374[b]	0.026[a]0.029[b]	0.249[a]0.269[b]	0.470[a]0.508[b]
	ZIGPMM	0.011[a]0.012[b]	0.086[a]0.091[b]	0.205[a]0.219[b]	0.015[a]0.016[b]	0.117[a]0.125[b]	0.277[a]0.289[b]	0.025[a]0.028[b]	0.193[a]0.213[b]	0.389[a]0.410[b]

[a]Mean of standard error for β_1; [b]Mean of standard error for β_2

Figure 3 displays the mean of AIC across models under each simulation condition. In general, for a fixed proportion of zero values, larger sample sizes tend to produce higher AIC across models. Conversely, for a given sample size, higher proportions of zero values are associated with lower AIC. Table 5 summarizes the best regression model for each simulation scenario based on the AIC criterion. As indicated in table 5, the ZINB, ZINBMM, and ZIGPMM models emerge as the most suitable regression approaches for addressing overdispersion attributable to excess zeros and unobserved heterogeneity.

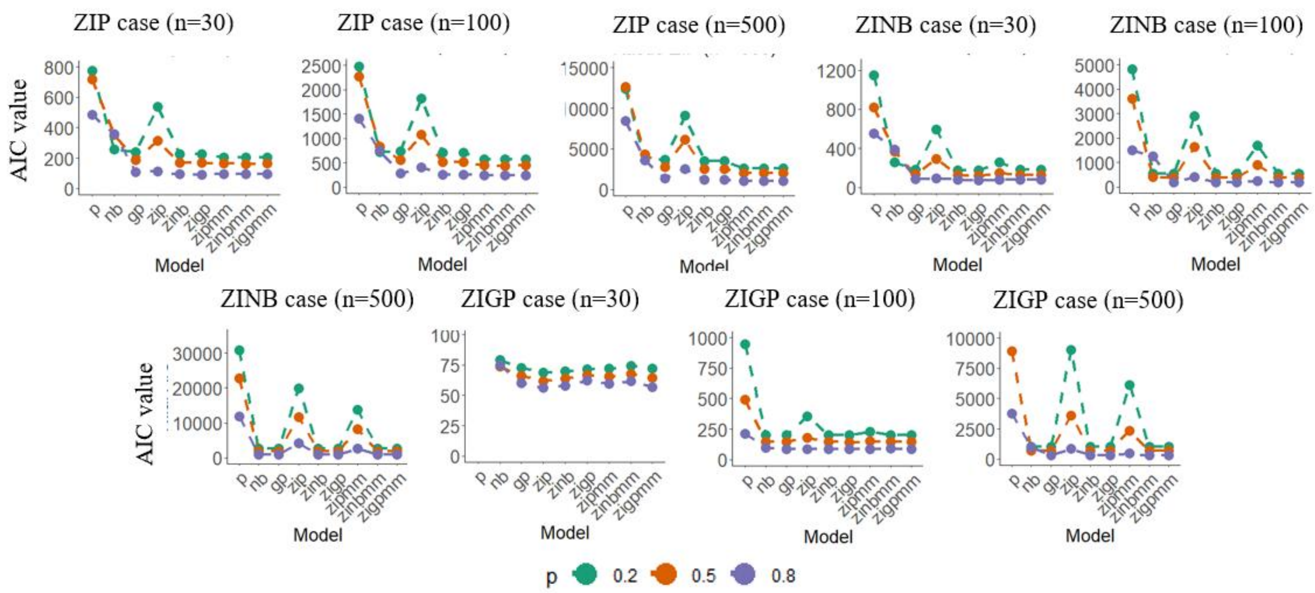

Figure 3. The average of AIC based on 27 simulation scenarios

Table 5. The Best Model based on AIC for each simulation scenario

n	$p_0 = 0.2$	$p_0 = 0.5$	$p_0 = 0.8$
30	ZINB	ZIGPMM	ZINB
100	ZINB	ZINB	ZIGPMM
500	ZINBMM	ZINBMM	ZINBMM

4.2. Results of Empirical Study: Modelling the Number of Under-Five Children Mortality Due to Pneumonia in Java Island

Distribution of the number of under-five children's mortality due to Pneumonia (Y) in Java Island is skewed to the right, with 20% of the observation units having zero under-five children mortality. Table 6 presents all VIF values for X_1, X_2 and X_3 are close to one (less than 10), indicating no evidence of multicollinearity among the explanatory variables. The results of hypothesis testing related to overdispersion and excess zeros yields the p-value less than significance level $\alpha = 0.05$. It indicates that there is an overdispersion and excess zero responses in our empirical data.

Table 6. VIF values for the Fixed Explanatory Variables

Explanatory Variable	VIF
X_1 (exclusive breastfeeding)	1.0111
X_2 (vitamin A)	1.0141
X_3 (nutrient-deficient)	1.0037

Due to the presence of overdispersion in our empirical data, Poisson regression model is no longer appropriate to be fitted. In table 7, it also can be seen the AIC for Poisson regression model is the largest. Consequently, the alternative models were used for modeling our data. From table 7, the three-smallest AIC for modeling under-five children mortality due to Pneumonia in Java Island are ZINB, ZINBMM, and ZIGPMM.

Table 7. AIC values for the nine regression models

Classic/Standard Count Models	AIC
Poisson (P)	568.65
NB	487.51
GP	493.34
Zero-Inflated Models	
ZIP	524.61
ZINB	470.55
ZIGP	552.68
Zero-Inflated Mixed Models	
ZIPMM	531.78
ZINBMM	485.78
ZIGPMM	485.03

As can be seen in table 8, there is a notable difference in the results of hypothesis testing among the four models. In Poisson modeling, the conclusion for the β_2 is to reject H0 ($\alpha = 0.05$). It means that the explanatory variable X_2 (the percentage of under-five children given vitamin A) is significant in affecting the response variable Y (the number of under-five children mortality due to pneumonia in Java Island). In contrast, in the three other models (ZINB, ZINBMM, and ZIGPMM), the conclusion for the β_2 is not to reject H0. Based on fitting these three models, X_2 does not significantly affect Y. The difference in results occurs due to the violation of the equidispersion assumption in Poisson regression, which causes the standard error estimate to be underestimated, which ultimately makes the results of the hypothesis testing tend to conclude that the explanatory variable has a significant influence on the response variable. Based on fitting the models ZINB, ZINBMM, and ZIGPMM, it can be concluded that at significance level $\alpha = 0.05$, the variable X_3 (the percentage of under-five children with nutrient deficiency) significantly affects Y.

Table 8. The results of fitting the regression models of Poisson, ZINB, ZINBMM, and ZIGPMM

Model	Parameter	Estimated Coefficient	Standard Error	p-value	Conclusion
Poisson	β_1 (exclusive breastfeeding)	-0.0065	0.0045	0.0170	Fail reject H0
	β_2 (vitamin A)	0.0197	0.0082	0.0000	Reject H0
	β_3 (nutrient-deficient)	0.0926	0.0184	0.0000	Reject H0
ZINB	β_1 (exclusive breastfeeding)	-0.0058	0.0086	0.5000	Fail reject H0
	β_2 (vitamin A)	0.0180	0.0115	0.1180	Fail reject H0
	β_3 (nutrient-deficient)	0.1140	0.0333	0.0005	Reject H0
ZINBMM	β_1 (exclusive breastfeeding)	-0.0050	0.0072	0.4930	Fail reject H0
	β_2 (vitamin A)	0.0112	0.0101	0.2370	Fail reject H0
	β_3 (nutrient-deficient)	0.1522	0.0273	0.0000	Reject H0
ZIGPMM	β_1 (exclusive breastfeeding)	-0.0053	0.0072	0.4600	Fail reject H0
	β_2 (vitamin A)	0.0117	0.0101	0.2440	Fail reject H0
	β_3 (nutrient-deficient)	0.1525	0.0277	0.0000	Reject H0

5. Conclusion

The impact of overdispersion in analysis of count data is more pronounced in hypothesis-testing of model parameter, particularly in the estimation of the standard error of the regression coefficient which became underestimated. The underestimated standard error leads to more easily rejecting the null hypothesis or more easily to conclude that the explanatory variable has a significant effect on the response. Based on AIC values, it is shown that the three best regression models in handling overdispersion count data due to excess zero responses and unobserved heterogeneity are ZINB, ZINBMM, and ZIGPMM. According to these models, the percentage of under-five children receiving exclusive breastfeeding and the percentage receiving vitamin A supplementation did not have a statistically significant effect on the number of under-five child mortality due to pneumonia in Java Island (α=0.05). On the other hand, the percentage of under-five children with nutrient deficiency has a statistically significant impact. As a result, the government should place greater emphasis on improving the nutritional status of children under five in each district and municipality in Java Island, to minimize the annual number of under-five deaths attributable to pneumonia.

6. Declarations

6.1. Author Contributions

Conceptualization: A.K., K.S., and D.H.; Methodology: A.K., Z.F., D.H., and K.S.; Software: K.S., and Z.F.; Validation: A.K., K.S., and D.H.; Formal Analysis: A.K., K.S., Z.F., and D.H; Investigation: A.K., and K.S.; Resources: Z.F., and D.H.; Data Curation: A.K., K.S., and Z.F..; Writing Original Draft Preparation: A.K., Z.F., D.H., and K.S.; Writing Review and Editing: A.K., Z.F., and D.H.; Visualization: Z.F.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This research was funded by a competitive research grant under the Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi, Republik Indonesia. Contract number: No.7/UN39.14/C3/DT.05.00/PFR/PL/2025.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. E. Ribeiro, W. M. Zeviani, W. H. Bonat, C. G. Demetrio, and J. Hinde, "Reparametrization of COM–Poisson regression models with applications in the analysis of experimental data," *Stat. Model.*, vol. 20, no. 5, pp. 443–466, 2019, doi: 10.1177/1471082X19838651.

[2] D. Handayani, W. Safitri, A. Amanda, W. Rahayu, and V. Santi, "Count regression models for analyzing crime rates in the East Java province," *J. Phys. Conf. Ser.*, vol. 2021, no. Jan., pp. 1–10, 2021, doi: 10.1088/1742-6596/2123/1/012028.

[3] J.-F. Dupuy, *Statistical Methods for Overdispersed Count Data*. Elsevier, 2018. [Online]. Available: https://doi.org/10.1016/B978-1-78548-266-3

[4] D. Saputro, A. Susanti, and N. Pratiwi, "The handling of overdispersion on Poisson regression model with the generalized Poisson regression model," *AIP Conf. Proc.*, vol. 2021, no. Jan., pp. 1–10, 2021, doi: 10.1063/5.0040330.

[5] E. Espinoza, U. Saputri, F. Fadilah, and D. Devianto, "Modeling the count data of public health service visits with overdispersion problem by using negative binomial regression," *J. Phys. Conf. Ser.*, vol. 2021, no. Jan., pp. 1–10, 2021, doi: 10.1088/1742-6596/1940/1/012021.

[6] A. Prahutama, D. Ispriyanti, and B. Warsito, "Modelling generalized Poisson regression in the number of dengue hemorrhagic fever (DHF) in East Nusa Tenggara," *E3S Web Conf.*, vol. 2020, no. Jan., pp. 1–10, 2020, doi: 10.1051/e3sconf/202020212017.

[7] A. Chowdhury and S. Biswas, "Appropriate statistical model for zero-inflated count data: simulation based study," *Comput. Methods Soc. Sci.,* vol. 6, no. 1, pp. 5–12, 2018.

[8] R. D. Kusuma and Y. Purwono, "Zero-inflated Poisson regression analysis on frequency of health insurance claim PT XYZ," *Adv. Econ. Bus. Manag. Res.*, vol. 72, no. Jan., pp. 321–325, 2019, doi: 10.2991/icbmr-18.2019.52.

[9] T. Xu, G. Zhu, and S. Han, "Study of depression influencing factors with zero-inflated regression models in a large-scale population survey," *BMJ Open*, vol. 7, no. 11, pp. 1–6, 2017, doi: 10.1136/bmjopen-2017-016471.

[10] R. F. Hassankiadeh, A. Kazemnejad, M. G. Fesharaki, and S. K. Jahromi, "Efficiency of zero-inflated generalized Poisson regression model on hospital length of stay using real data and simulation study," *Caspian J. Health Res.*, vol. 3, no. 1, pp. 5–9, 2018, doi: 10.29252/cjhr.3.1.5.

[11] X. Harrison, "Using observation-level random effects to model overdispersion in count data in ecology and evolution," *PeerJ,* vol. 2, no. Jan., pp. 1–12, 2014, doi: 10.7717/peerj.616.

[12] H. K. Lim, W. K. Li, and P. L. Yu, "Zero-inflated Poisson regression mixture model," *Comput. Stat. Data Anal.*, vol. 71, no. Jan., pp. 1–12, 2014, doi: 10.1016/j.csda.2013.06.021.

[13] R. Fang, B. Wagner, and J. Harris, "Zero-inflated negative binomial mixed model: an application to two microbial organisms important in oesophagitis," *Epidemiol. Infect.*, vol. 144, no. 11, pp. 2447–2455, 2016, doi: 10.1017/S0950268816000662.

[14] F. C. Xie, J. G. Lin, and B. C. We, "Score test for homogeneity of dispersion in generalized Poisson mixed models with excess zeros," *Commun. Stat. Simul. Comput.,* vol. 46, no. 1, pp. 301–314, 2015, doi: 10.1080/03610918.2014.963610.

[15] Kemenkes RI, "Indonesia's Health Profile 2021", Jakarta: Kementerian Kesehatan RI, 2022.

[16] Y. Mouatassim and E. H. Ezzahid, "Poisson regression and zero-inflated Poisson regression: application to private health insurance data," *Eur. Actuar. J.*, vol. 2, no. 1, pp. 187–204, 2012, doi: 10.1007/s13385-012-0056-2.

[17] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2024. [Online]. Available: https://doi.org/10.1201/9781003456285

[18] R. H. Myers, D. C. Montgomery, G. G. Vining, and T. J. Robinson, “Generalized linear models with application in engineering and the sciences,” *J. Appl. Stat*., vol. 2010, no. Jan., pp. 1–20, 2010.

[19] J. M. Hilbe, “Negative binomial regression,” *Stat. Model. J.*, vol. 2011, no. Jan., pp. 1–20, 2011.

[20] A. C. Cameron and P. K. Trivedi, “Regression analysis of count data,” *Econom. Stat. J.*, vol. 2013, no. Jan., pp. 1–20, 2013.

[21] J. W. Hardin and J. M. Hilbe, “Generalized linear models and extensions,” *Stat. Anal. J.*, vol. 2007, no. Jan., pp. 1–20, 2007.

[22] W. Wang and F. Famoye, “Modeling household fertility decisions with generalized Poisson regression,” *J. Popul. Econ*., vol. 10, no. 3, pp. 273–283, 1997, doi: 10.1007/s001480050043.

[23] F. Famoye and W. Wang, “Censored generalized Poisson regression model,” *Comput. Stat. Data Anal*., vol. 46, no. 3, pp. 547–560, 2004, doi: 10.1016/j.csda.2003.08.007.

[24] T. H. Lin and M. H. Tsai, “Modeling health survey data with excessive zero and K responses,” *Stat. Med*., vol. 32, no. Jan., pp. 1572–1583, 2013, doi: 10.1002/sim.5650.

[25] P. Wang, I. M. Cockburn, and M. L. Puterman, “Analysis of patent data: a mixed Poisson regression model approach,” *J. Bus. Econ. Stat.,* vol. 16, no. 1, pp. 27–41, 1998.

[26] 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,” *Biometrical J.*, vol. 45, no. 4, pp. 437–452, 2003, doi: 10.1002/bimj.200390024.

[27] F. C. Xie, B. C. Wei, and J. G. Lin, “Assessing influence for pharmaceutical data in zero-inflated generalized Poisson mixed models,” *Stat. Med*., vol. 27, no. 18, pp. 3656–3673, 2008, doi: 10.1002/sim.3233.

[28] V. Eminita, A. Kurnia, and K. Sadik, “Handling overdispersion in modeling count data with zero-inflated responses,” *Fibonacci J*., vol. 1, no. Jan., pp. 71–80, 2019.