Bivariate Zero-Inflated Poisson Inverse Gaussian Regression Model and Its Application

Purhadi [a], Ermawati [a,b,*], Rossy Noviyana [a], Sutikno [b]

[a] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
[b] Department of Mathematics, Universitas Islam Negeri Alauddin Makassar, 92118, Indonesia
Corresponding author: *erma.mat@gmail.com

Abstract— **This study developed a Bivariate Zero-Inflated Poisson Inverse Gaussian Regression (BZIPIGR) model to presents the form of BZIPIGR parameter estimation and modeling of the number of HIV and AIDS cases in each sub-district in Trenggalek and Ponorogo regencies to determine the factors that have a significant effect. This model can be used on data that have overdispersion cases caused by extra zeros in the response variables. The parameter estimation of the BZIPIGR model uses the Maximum Likelihood Estimation (MLE). The first derivative of the BZIPIGR model has obtained not closed form, therefor it has continued with the Berndt Hall Hall Hausman (BHHH) iteration to obtain the maximum likelihood estimators, while the hypothesis testing of the BZIPIGR model is derived using Maximum Likelihood Ratio Test (MLRT) approach. Based on the AICc value obtained, the BZIPIGR model is a feasible model to be applied to data on the number of HIV and AIDS cases in Trenggalek and Ponorogo Districts, East Java Province. The variable that had a significant effect on the increase in the number of HIV and AIDS cases was the percentage of the population with low education (SMA). The variables that had a significant effect on reducing the number of HIV and AIDS cases were the percentage of the population aged 25-29 years, the percentage of reproductive-age couples using condoms, the percentage of health educations activities about HIV and AIDS, and the percentage of community health insurance (Jamkesmas).**

Keywords— **Overdispersion; extra zeros; MLE; Bivariate Zero-Inflated Poisson Inverse Gaussian (BZIPIG); HIV; AIDS.**

I. INTRODUCTION

Count data states the number of occurrences in a certain period and is in the form of non-negative integers. Count data cannot use Ordinary Least Square (OLS) regression because it will violate the assumption that the error value follows a normal distribution and has heteroscedasticity. One of the models that can apply for modeling count data is Poisson regression, but this model must fulfill the Equi dispersion assumption (mean and variance are equal). This assumption is often violated. That is, the variance value is smaller than the mean (under dispersion) or otherwise (overdispersion)[1]. The number of zeros (excess zeros) can cause a large enough variance that is known as overdispersion[2]. Occasionally, an event rarely occurs. Therefore, many responses are zero (excess zero). If the case is disregarded, it will underestimate and wrong decision when testing the hypothesis[3].

Several models have been developed to solve the overdispersion problem. The modeling can use Generalized Poisson[4] or a mixed Poisson distribution both discrete and continuous. Some of the mixed-Poisson distributions that have been developed are Negative Binomial [5], and Poisson Inverse Gaussian (PIG) [6]. Poisson Inverse Gaussian (PIG), a mixed Poisson distribution with random effects, has an Inverse Gaussian distribution. These distributions are not good enough to handle overdispersion and extra zeros cases.

Several other mixed Poisson distributions can manage this problem, they are Zero Inflated Negative Binomial (ZINB)[5], Zero Inflated Generalized Poisson (ZIGP)[7], and Zero Inflated Poisson Inverse Gaussian (ZIPIG)[8]. Based on the results of research by Hilbe that the ZIPIGR model is the best compared to ZIP or ZINB in modeling count data with many significant zero values[9]. Constraints in the parameter estimation process of the ZIPIGR model using Maximum Likelihood Estimation (MLE) is the function of the model is complicated, thereby to overcome this problem in obtaining parameter estimator must be followed by numerical iteration. One of the numerical iterations that can apply is the Berndt Hall Hall Hausman (BHHH) algorithm[10], which in this algorithm only uses the first derivative of its likelihood function.

The ZIPIGR model is still limited to one response variable only. Meanwhile, model cases with two response variables will be developed to the Bivariate Zero Inflated Poisson Inverse Gaussian Regression (BZIPIGR) model. This study will be applying it to modeling the number of HIV and AIDS cases. In this case, there are many zeroes in the response variable, so the BZIPIGR model is suitable for the application. HIV (Human Immunodeficiency Virus) is a virus that attacks white blood cells, which causes a decrease in the body's immune system. AIDS (Acquired Immune Deficiency Syndrome) is a group of diseases that arise due to decreased immunity caused by HIV infection. HIV and AIDS have become an epidemic worldwide because no medicine can cure the sufferer. Also, the symptoms of this disease are not visible, and the course of this disease takes a long time. The fifth highest rankings of HIV infections that occurred in Indonesia from 1987 to 2019 were East Java, DKI Jakarta, Papua, Bali, Riau, and West Java [11]. In opposition to HIV, the three highest rankings of AIDS infections were central Java, Papua, and East Java. In 2019 the number of HIV 8,885 cases and 920 cases of AIDS in East Java, this number increased from the previous year[12]. Many HIV and AIDS infections in East Java, which is extremely high, need to be addressed. Trenggalek and Ponorogo regencies are districts in East Java Province. This study will apply the BZIPIGR model to data on the number of HIV and AIDS cases in Trenggalek and Ponorogo Districts.

II. MATERIALS AND METHOD

A. Poisson Inverse Gaussian Regression (PIGR)

The PIG distribution is a mixed Poisson distribution consisting of two parameters, namely μ (mean) as the location parameter and τ (dispersion parameter) as the shape parameter. Let Y is the response variable that PIG distribution and can be denoted by $Y \sim PIG(\mu, \tau)$. The probability density function of Y is as follows[13].

$$f(y; \mu, \tau) = \left(\frac{2z}{\pi}\right)^{\frac{1}{2}} \frac{\mu^y e^{\frac{1}{\tau}} K_s(z)}{(z\tau)^y y!}, \quad y \geq 0 \qquad (1)$$

with $s = y - \frac{1}{2}$ and $z = \sqrt{\frac{1}{\tau^2} + \frac{2\mu}{\tau}}$. Where according to Wilmot, $K_s(z) = K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{(2\mu\tau + 1)}\right)$ is the third kind of the modified Bessel function [14]. The expected value and variance of the PIG distribution are

$$E(Y) = E\{E(Y \mid \mu v)\} = E(\mu v) = \mu.$$

$$Var(Y) = Var\{E(Y \mid \mu v)\} + E\{Var(Y \mid \mu v)\} = \mu + \tau \mu^2.$$

Y is the number of events in a unit of observation in a certain period, μ is the average of these events, and tau is an overdispersion parameter and is the same as $Var(v)$, which is due to the heterogeneity or diversity associated with the unit of observation with certain characters [12].

Suppose Y_i is the response for the ith observation and $\mathbf{x}_i^T = \begin{bmatrix} 1 & x_{1i} & x_{2i} & ... & x_{pi} \end{bmatrix}^T$ is the explanatory vector for the

ith observation with the dimension (k+1) x1 , so the PIG Regression Model is as follows:

$$\mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) \qquad \text{or} \qquad \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \qquad \text{where}$$

$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & ... & \beta_k \end{bmatrix}^T$.

B. Bivariate Poisson Inverse Gaussian Regression (BPIGR)

The Bivariate Poisson Inverse Gaussian (BPIG) distribution has two count variables that are correlated. Suppose that two random variables have a Poisson distribution and are independent of each other, Y_1 and Y_2 , which have the mean $v\mu_1$ and $v\mu_2$ respectively, and the variance $Var(Y_1) = \mu_1 + \mu_1^2 \tau$ and $Var(Y_2) = \mu_2 + \mu_2^2 \tau$. Variable v is a random variable with Gaussian inverse distribution. Hence Y_1 and Y_2 have mixed Poisson distribution, namely Poisson Inverse Gaussian. The BPIG distribution opportunity density function is as follows [15].

$$f(y_1, y_2; \mu_1, \mu_2, \tau) = \left(\frac{2z}{\pi}\right)^{\frac{1}{2}} \frac{\mu_1^{y_1} \mu_2^{y_2} e^{\frac{1}{\tau}} K_s(z)}{(z\tau)^{y_1 + y_2} y_1! y_2!} \qquad (2)$$

where $s = y_1 + y_2 - \frac{1}{2}$ and $z = \frac{1}{\tau}\left(1 + 2\tau \sum_{h=1}^{2} \mu_h\right)^{\frac{1}{2}}$ so

$$K_s(z(u,v)) = K_{y_1 + y_2 - \frac{1}{2}}\left[\frac{1}{\tau}\left(1 + 2\tau \sum_{h=1}^{2} \mu_h\right)^{\frac{1}{2}}\right]$$

BPIGR is a regression model with two correlated responses. Suppose y_{ih} is the hth response to the ith observation and is give a random sample $(Y_{1i}, Y_{2i}) \sim BPIG(\mu_{hi}, \tau)$ where $i = 1, 2, ..., n$ and $h = 1, 2$. Then the BPIGR model can be written as follows.

$$\log[E(Y_{ih})] = \mathbf{x}_i^T \boldsymbol{\beta}_h \qquad \text{and} \quad E(Y_{ih}) = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}_h\right),$$

where.

$\mathbf{x}_i = \begin{bmatrix} 1 & x_{1i} & x_{2i} & ... & x_{pi} \end{bmatrix}^T$ is the explanatory vector for the ith observation, $i = 1, 2, ..., n$

$\boldsymbol{\beta}_h = \begin{bmatrix} \beta_{h0} & \beta_{h1} & \beta_{h2} & ... & \beta_{hp} \end{bmatrix}^T$ is the vector of the regression coefficient dimension $(p + 1)$ x 1 in the h response, $h=1,2$.

C. Bivariate Zero Inflated Poisson Regression (BZIPR)

Let Y_1 and Y_2 are random variables with a bivariate Poisson distribution $(Y_1, Y_2) \sim BP(\mu_1, \mu_2, \mu_0)$ where the combined probability function is[16]

$$P(Y_1, Y_2) = \begin{cases} p + (1-p)\exp\left(-(\mu_1 + \mu_2 + \mu_0)\right) & y_1 = y_2 = 0 \\ (1-n) \sum_{j=0}^{\min(y_1, y_2)} \frac{\mu_1^{y_1 - j} \mu_2^{y_2 - j} \mu_0^j}{(y_1 - j)!(y_2 - j)! j!} & \\ \qquad \exp\left(-(\mu_1 + \mu_2 + \mu_0)\right) & y_1 \neq 0 \, y_2 \neq 0 \end{cases} \qquad (3)$$

The mean, the variance and the covariance values of the BZIP distribution are.

$$E(Y_l) = (1-p)(\mu_l + \mu_0)$$
$$Var(Y_l) = E(Y_l)[1 + p(\mu_l + \mu_0)], \quad l = 1, 2$$
$$Cov(Y_1, Y_2) = (1-p)[(\mu_0 + p(\mu_1 + \mu_0)(\mu_2 + \mu_0))]$$

The BZIP regression model is as follows.

$$\mu_{hk} + \mu_0 = \exp\left(x_k^T \beta_h\right); h = 1,2 \qquad (4)$$

Method of estimation in the BZIPR model is a MLE. The method for calculating the test statistics on the parameter test is the Maximum Likelihood Ratio Test (MLRT)[17].

D. Zero Inflated Poisson Inverse Gaussian Regression (ZIPIGR)

If $Y=0$ with the probability value v and $Y=Y_1$ with the probability value (1-v), then Y has a Zero-Inflated Poisson Inverse Gaussian distribution, which can be written as withthe probability function follows[18].

$$P(Y = y \mid \mu, \tau, p) = \begin{cases} p + (1-p)P(Y = 0 \mid \mu, \tau) & ; \quad y = 0 \\ (1-p)P(Y = y \mid \mu, \tau) & ; \quad y = 1,2,3,... \end{cases} \quad (5)$$

The ZIPIGR model can be written as follows.

$$p = \frac{\exp\left(x^T \gamma\right)}{1 + \exp\left(x^T \gamma\right)} \text{ and } \mu = \exp\left(x^T \beta\right)$$

E. Bivariate Zero Inflated Poisson Inverse Gaussian Regression (BZIPIGR)

Let Y_1 and Y_2 are random variable which BZIPIG distribution with $\left(Y_{1i}, Y_{2i}\right) \overset{iid}{\sim} BZIPIG\left(\mu_0, \mu_1, \mu_2, \tau_1, \tau_2, p\right)$ then the joint probability function of Y_1, Y_2 as follows [8]

a. If $y_1 = 0$ and $y_2 = 0$, then $P(Y_1 = 0, Y_2 = 0)$.

$$P(y_{1i}, y_{2i}) = p_2(\mathbf{x}_i)p_1(\mathbf{x}_i) + \left[(1 - p_2(\mathbf{x}_i))p_1(\mathbf{x}_i)\left(\frac{2}{\pi\tau}\sqrt{2\lambda_{2i}\tau + 1}\right)^{\frac{1}{2}}\exp\left(\tau^{-1}\right)\right.$$
$$K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\lambda_{2i}\tau + 1}\right) + \left[p_2(\mathbf{x}_i)(1 - p_1(\mathbf{x}_i))\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}\exp\left(\tau^{-1}\right)\right.$$
$$\left(2\lambda_{1i}\tau + 1\right)^{\frac{1}{4}}K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\lambda_{1i}\tau + 1}\right) + \left[(1 - p_1(\mathbf{x}_i))(1 - p_2(\mathbf{x}_i))\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}\right.$$
$$\left.\exp\left(\tau^{-1}\right)\left(2\tau(\lambda_{1i} + \lambda_{2i}) + 1\right)^{\frac{1}{4}}K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\tau(\lambda_{1i} + \lambda_{2i}) + 1}\right)\right]$$

$$(6)$$

b. If $y_1 \neq 0$ and $y_2 \neq 0$, then $P(Y_1 \neq 0, Y_2 \neq 0)$.

$$P(y_{1i}, y_{2i}) = (1 - p_1(\mathbf{x}_i))(1 - p_2(\mathbf{x}_i))e^{\frac{1}{\tau}}\left(\frac{2}{\pi\tau}\sqrt{2\tau(\lambda_{1i} + \lambda_{2i}) + 1}\right)^{\frac{1}{2}}\lambda_{1i}^{y_{1i}}\lambda_{2i}^{y_{2i}}$$
$$\frac{\exp\left(\tau^{-1}\right)K_{y_{1i}+y_{2i}-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\tau(\lambda_{1i} + \lambda_{2i}) + 1}\right)}{\left(\sqrt{1 + 2\tau(\lambda_{1i} + \lambda_{2i})}\right)^{y_{1i}+y_{2i}}y_{1i}!y_{2i}!}$$

$$(7)$$

The BZIPIGR model can be written as follows.

$$p_1(\mathbf{x}_i) = \frac{\exp\left(-\theta_1\mathbf{x}_i^T\boldsymbol{\beta}_1\right)}{1 + \exp\left(-\theta_1\mathbf{x}_i^T\boldsymbol{\beta}_1\right)} \qquad 1 - p_1(\mathbf{x}_i) = \frac{1}{1 + \exp\left(-\theta_1\mathbf{x}_i^T\boldsymbol{\beta}_1\right)}$$

$$p_2(\mathbf{x}_i) = \frac{\exp\left(-\theta_2\mathbf{x}_i^T\boldsymbol{\beta}_2\right)}{1 + \exp\left(-\theta_2\mathbf{x}_i^T\boldsymbol{\beta}_2\right)} \qquad 1 - p_2(\mathbf{x}_i) = \frac{1}{1 + \exp\left(-\theta_2\mathbf{x}_i^T\boldsymbol{\beta}_2\right)}$$

$$\text{logit}(p_{1i}) = \log\left(\frac{p_1(\mathbf{x}_i)}{1 - p_1(\mathbf{x}_i)}\right) = -\theta_1\mathbf{x}_i^T\boldsymbol{\beta}_1 \qquad \lambda_{1i} = \exp\left(\mathbf{x}_i^T\boldsymbol{\beta}_1\right)$$

$$\text{logit}(p_{2i}) = \log\left(\frac{p_2(\mathbf{x}_i)}{1 - p_2(\mathbf{x}_i)}\right) = -\theta_2\mathbf{x}_i^T\boldsymbol{\beta}_2 \qquad \lambda_{2i} = \exp\left(\mathbf{x}_i^T\boldsymbol{\beta}_2\right)$$

$$(8)$$

F. HIV and AIDS and Factors Affecting the Number of HIV and AIDS Cases

HIV is a type of virus that attacks/infects white blood cells, which will cause a decrease in human immunity and impair its function. If HIV infection is not handled immediately, it will develop into serious health called AIDS, which is the final stage of infection caused by HIV, which can infect the body's organ systems, including the brain, thus damaging the immune system. Until now, there is no cure for HIV and AIDS. However, there are drugs to slow the progression of the disease and increase the life expectancy of sufferers.

HIV and AIDS cases were determined by three main factors, namely[19].

1. Predisposing factors. The factors that encourage a person's behavior include knowledge, attitudes, religion, beliefs, values, traditions, etc. The driving factors for HIV exposure are low knowledge of HIV, accepting attitudes about the behaviors that cause HIV, and traditions that can lead to HIV.

2. Enabling factors. Factors that facilitate a person's behavior to become infected with HIV such as facilities and infrastructure for the occurrence of health behaviors, enabling factors that can cause HIV, namely the distance between prostitution and prostitution localization close to home, distance to health services that are far away so that there is less information about HIV.

3. Reinforcing factors are factors that amplifier the behavior of a person infected with HIV. Sometimes even though people know and can behave healthily, they do not. Examples include family, health workers, and community leaders. In cases of HIV exposure, reinforcing factors are family rules that do not prohibit HIV-causing behavior and the enabling environment for HIV-causing behavior.

G. Data Source

This study used secondary data obtained from the Health Service of Trenggalek and Ponorogo Districts Profile in 2012. This data has been used by Wijaya [20] to know the factors that can affect the number of HIV and AIDS cases. The observation units used in this study were 27 districts in Trenggalek and Ponorogo regencies.

Research variables used in this study consisted of two responses (Y) and five predictor variables (X). The variables in this study are given in Table 1:

TABLE I
RESEARCH VARIABLES

Response Variables	
Y_1	The number of HIV
Y_2	The number of AIDS
Predictor Variables	
X_1	The percentage of age group 25-29 years
X_2	The percentage of population with senior high school level
X_3	The percentage of couples of reproductive ages using condoms
X_4	The Percentage of health education activities about HIV and AIDS
X_5	The percentage of community health insurance (Jamkesmas)

H. Steps to determine the factors that influence the number of HIV and AIDS cases using BZIPIGR Model

The steps to determine the factors that influence the number of HIV and AIDS cases are as follows.

1. Make a descriptive analysis of variables.
2. Correlation test between responses using the Pearson test.
3. Detect multicollinearity from predictor variables using the VIF test criteria.
4. Testing of overdispersion using the Variance Test (VT)
5. Perform data analysis using the BZIPIGR model.
   a. Specifies the parameter estimator value using MLE with BHHH algorithm.
   b. Conducting hypothesis testing simultaneously using MLRT on the regression parameters.
   c. Perform partial hypothesis testing on regression parameters.
   d. Interpret the analysis results.
   e. Conclusion.

III. RESULT AND DISCUSSION

A. Parameter Estimation of BZIPIGR

The parameter estimation of the BZIPIGR model uses MLE method. The probability distribution for BZIPIG is in equation (8). The first stage is to estimate the parameters using MLE by forming the likelihood function, which is defined as follows.

$$L(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\theta_1,\theta_2,\tau) = \prod_{i=1}^{n}(A_i) + \prod_{i=1}^{n}(B_i) \qquad (9)$$

where.

$$A_i = \frac{\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)}{1+\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)} \frac{\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)}{1+\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)} + O_1 + O_2 + O_3$$

$$O_1 = \frac{1}{1+\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)} \frac{\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)}{1+\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)} \left(\frac{2}{\pi\tau}\sqrt{2\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\tau+1}\right)^{\frac{1}{2}}$$
$$\exp(\tau^{-1}) K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\tau+1}\right)$$

$$O_2 = \frac{\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)}{1+\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)} \frac{1}{1+\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} \left(2\tau\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+1\right)^{\frac{1}{4}}$$
$$\exp(\tau^{-1}) K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\tau\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+1}\right)$$

$$O_3 = \frac{\left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}}\exp(\tau^{-1})}{\left(1+\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)\right)\left(1+\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)\right)} \left(2\tau\left(\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\right)+1\right)^{\frac{1}{4}}$$
$$K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\tau\left(\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\right)+1}\right)$$

$$B_i = \frac{e^{\frac{1}{\tau}}\left(\frac{2}{\pi\tau}\sqrt{2\tau\left(\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\right)+1}\right)^{\frac{1}{2}}}{\left(1+\exp(-\theta_1 \mathbf{x}_i^T \boldsymbol{\beta}_1)\right)\left(1+\exp(-\theta_2 \mathbf{x}_i^T \boldsymbol{\beta}_2)\right)} P_1$$

$$P_1 = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)^{y_1}\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)^{y_2}\exp(\tau^{-1})}{\left(\sqrt{1+2\tau\left(\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\right)}\right)^{y_1+y_2} y_1!\, y_2!}$$
$$K_{-\frac{1}{2}}\left(\tau^{-1}\sqrt{2\tau\left(\exp(\mathbf{x}_i^T \boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T \boldsymbol{\beta}_2)\right)+1}\right)$$

The second stage is to determine the log from the likelihood function.

$$\ell = \log L(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\theta_1,\theta_2,\tau) = \log\prod_{i=1}^{n}(A_i)^{1-b_i}(B_i)^{b_i} \qquad (10)$$

$$\ell = \sum_{i=1}^{n}(1-b_i)\log(A_i) + \sum_{i=1}^{n}b_i\log(B_i) \qquad (11)$$

b_i is a dummy variable equal to 1. Furthermore, to obtain parameter estimation, the $\ell(\boldsymbol{\beta}_h,\theta_h,\tau; h=1,2)$ function is derived against the parameter $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\theta_1,\theta_2$ and τ then equated with zero. These first derivative does not yield closed form, hence parameter estimation was continued to a numerical method, namely the BHHH iteration[21], where the initial value $\boldsymbol{\gamma}=[\boldsymbol{\beta}_1^T\ \boldsymbol{\beta}_2^T\ \theta_1\ \theta_2\ \tau]^T$ and $m=0$ with $\hat{\boldsymbol{\gamma}}_{(0)}>0$ corresponds to the ZIPIGR function for convergence tolerance limits. The initial values for $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\theta_1,\theta_2$, and τ are obtained from the estimator in the ZIPIGR model.

The iteration starting from $m=0$ in the following equation. $\hat{\boldsymbol{\gamma}}_{m+1} = \hat{\boldsymbol{\gamma}}_m - \mathbf{H}^{-1}(\hat{\boldsymbol{\gamma}}_m)\mathbf{g}(\hat{\boldsymbol{\gamma}}_m)$. Iteration will stop if value $\left\|\hat{\boldsymbol{\gamma}}_{m+1} - \hat{\boldsymbol{\gamma}}_m\right\| \le \varepsilon$, where ε is a very small positive number close to 0, where the gradient vector $\mathbf{g}(\hat{\boldsymbol{\gamma}}_{(m)})$ is

$$\mathbf{g}(\hat{\boldsymbol{\gamma}}_{(m)}) = \left[\left(\frac{\partial L(.)}{\partial \boldsymbol{\beta}_1^T}\right)^T \left(\frac{\partial L(.)}{\partial \boldsymbol{\beta}_2^T}\right)^T \left(\frac{\partial L(.)}{\partial \theta_1}\right) \left(\frac{\partial L(.)}{\partial \theta_2}\right) \frac{\partial L(.)}{\partial \tau}\right]^T$$

with $L(\bullet)=\log L(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\theta_1,\theta_2,\tau)$ and the Hessian matrix is

$$\mathbf{H}(\hat{\boldsymbol{\gamma}}_{(m)}) = -\sum_{i=1}^{n}\mathbf{g}_i(\hat{\boldsymbol{\gamma}}_m)\mathbf{g}_i(\hat{\boldsymbol{\gamma}}_m)^T$$

with $\mathbf{g}_i(\hat{\boldsymbol{\gamma}}_m) = \dfrac{\partial \log L(y_i|X_i,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\theta_1,\theta_2,\tau)}{\partial(\hat{\boldsymbol{\gamma}}_m)}$

B. Hypothesis Testing of Parameter BZIPIG Model

Testing Hypotheses are examining that simultaneously and partially to determine the significance of each parameter.

1) Simultaneous Testing Parameters β_1, β_2

Hypothesis testing of parameters BZIPIGR model, determined by the likelihood function $L(\hat{\Omega})$ and $L(\hat{\omega})$, where $L(\hat{\Omega})$ is the Maximum Likelihood value under the population which involved the predictor variables, and otherwise $L(\hat{\omega})$ is the likelihood function under H_0. Simultaneous testing of the parameters BZIPIGR model is to determine the significance of the parameters β_1, β_2 together, with the following hypothesis.

$H_0 : \beta_{1k} = \beta_{2k} = 0$ with $k = 1,2,\ldots,p$

H_1 : at least one of $\beta_{hk} = 0$ with $h = 1,2$ and $k = 1,2,\ldots,p$

Before managing the test statistic to be used, it will define the parameter set under the population , namely , then from the set of parameters, the likelihood function can be formed under the population like in equation (9).

The maximum likelihood value under population $L(\Omega)$ is $L(\hat{\Omega})$ were obtained from section 3.1. Meanwhile, the set of parameters under H_0 (ω) is $\omega = \{\beta_{10}, \beta_{20}, \theta_{10}, \theta_{20}, \tau\}$. Then from the set parameters, the likelihood function under H_0 $L(\omega)$ can form.

$$L(\beta_{10}, \beta_{20}, \theta_{10}, \theta_{20}, \tau) = \prod_{i=1}^{n} P((y_{1i}, y_{2i}) \mid \beta_{10}, \beta_{20}, \theta_{10}, \theta_{20}, \tau) \quad (12)$$

The probability function of BZIPIG distribution like in equation (6) and (7). The BZIPIGR model without predictor variable.

$$p_1 = \frac{\exp(-\theta_1 \beta_0)}{1 + \exp(-\theta_1 \beta_0)} = \frac{\exp(-\theta_{10})}{1 + \exp(-\theta_{10})} \qquad 1 - p_1 = \frac{1}{1 + \exp(-\theta_{10})}$$

$$p_2 = \frac{\exp(-\theta_2 \beta_0)}{1 + \exp(-\theta_2 \beta_0)} = \frac{\exp(-\theta_{20})}{1 + \exp(-\theta_{20})} \qquad 1 - p_2 = \frac{1}{1 + \exp(-\theta_{20})}$$

$$\lambda_1 = \exp(\beta_{10}) \qquad\qquad\qquad\qquad \lambda_2 = \exp(\beta_{20})$$

$$(13)$$

the maximum likelihood value under H_0 $L(\omega)$ is $L(\hat{\omega})$ obtained in the following way. Substitute equation (14) in equation (6) and (7) to obtain

$$P(y_{1i}, y_{2i}) = \begin{cases} C_i, (y_{1i}, y_{2i}) = (0,0) \\ D_i, (y_{1i}, y_{2i}) \neq (0,0) \end{cases} \quad (14)$$

The first step is to estimate the parameters using MLE by forming the likelihood function, which is defined as follows.

$$L(\beta_{10}, \beta_{20}, \theta_{10}, \theta_{20}, \tau) = \prod_{i=1}^{n} (C_i) + \prod_{i=1}^{n} (D_i) \quad (15)$$

where.

$$C_i = \frac{\exp(-\theta_{20})}{1 + \exp(-\theta_{20})} \frac{\exp(-\theta_{10})}{1 + \exp(-\theta_{10})} + O_1 + O_2 + O_3$$

$$O_1 = \frac{1}{1 + \exp(-\theta_{20})} \frac{\exp(-\theta_{10})}{1 + \exp(-\theta_{10})} \left(\frac{2}{\pi\tau} \sqrt{2 \exp(\beta_{20})\tau + 1} \right)^{\frac{1}{2}}$$
$$\exp(\tau^{-1}) K_{-\frac{1}{2}} \left(\tau^{-1} \sqrt{2 \exp(\beta_{20})\tau + 1} \right)$$

$$O_2 = \frac{\exp(-\theta_{20})}{1 + \exp(-\theta_{20})} \frac{1}{1 + \exp(-\theta_{10})} \left(\frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\tau \exp(\beta_{10}) + 1)^{\frac{1}{4}}$$
$$\exp(\tau^{-1}) K_{-\frac{1}{2}} \left(\tau^{-1} \sqrt{2\tau \exp(\beta_{10}) + 1} \right)$$

$$O_3 = \frac{\left(\frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\tau(\exp(\beta_{10}) + \exp(\beta_{20})) + 1)^{\frac{1}{4}} \exp(\tau^{-1})}{(1 + \exp(-\theta_{10}))(1 + \exp(-\theta_{20}))}$$
$$K_{-\frac{1}{2}} \left(\tau^{-1} \sqrt{2\tau(\exp(\beta_{10}) + \exp(\beta_{20})) + 1} \right)$$

$$D_i = \frac{e^{\frac{1}{\tau}} \left(\frac{2}{\pi\tau} \sqrt{2\tau(\exp(\beta_{10}) + \exp(\beta_{20})) + 1} \right)^{\frac{1}{2}} P_1}{(1 + \exp(-\theta_1))(1 + \exp(-\theta_2))}$$

$$P_1 = \frac{\exp(\beta_{20})^{y_1} \exp(\beta_{20})^{y_2} K_{-\frac{1}{2}} \left(\tau^{-1} \sqrt{2\tau(\exp(\beta_{10}) + \exp(\beta_{20})) + 1} \right)}{\left(\sqrt{1 + 2\tau(\exp(\beta_{10}) + \exp(\beta_{20}))} \right)^{y_1 + y_2} y_1! y_2!}$$
$$\exp(\tau^{-1})$$

The second stage is to determine the log from the likelihood function.

$$I = \log L(\beta_{10}, \beta_{20}, \theta_{10}, \theta_{20}, \tau) = \log \prod_{i=1}^{n} (C_i)^{1-b_i} (D_i)^{b_i} \quad (16)$$

$$I = \sum_{i=1}^{n} (1 - b_i) \log(C_i) + \sum_{i=1}^{n} b_i \log(D_i) \quad (17)$$

b_i is a dummy variable with a value of 1. Furthermore, to obtain parameter estimation, the function $\ell(\beta_{10}, \beta_{20}, \theta_{10}, \theta_{20}, \tau)$ defines a derivative for parameters then equates to zero. This first derivative for each parameter does not yield close form. Then for getting $L(\hat{\omega})$ is using BHHH algorithm.

From these two maximum likelihood values, a ratio is formed between the maximum likelihood value under population $L(\Omega)$ and the maximum likelihood value under H_0 $L(\omega)$, which is called the odds ratio, as follows:

$$\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

with the rejection area of H_0 $\Lambda < \Lambda_0$ where $0 < \Lambda_0 < 1$.

The next is to determine the distribution of test statistics with the n big sample approach as follows.

$$G^2 = -\log(\Lambda^2)$$

with

$$G^2 > c \text{ and } \alpha = P(G^2 > c \mid H_0 \text{ is true})$$

c will obtain if G^2 is known, so this simultaneous test uses the test statistics.

$$G^2 = -\log\left(\Lambda^2\right) = -2\log\left(\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right) = 2\left(\log\left(L(\hat{\Omega})\right) - \log\left(L(\hat{\omega})\right)\right).$$

will compare with $\chi^2_{\alpha,(a-b)}$, where a is the number of parameters under the population, and b is the number of parameters under H_0. The rejection area of H_0 if $\left|G^2\right| > \chi^2_{\alpha,(a-b)}$.

2) Partial Parameter Testing

If the test results produce a rejecting conclusion, then the test can be continued by partially testing with the Wald test statistics. This test, to find out the effect of the predictor variable partially on the response, with hypothesis:

i) $H_0 : \beta_{hk} = 0$

       $H_1 : \beta_{hk} \neq 0$ with $h = 1, 2$ and $k = 1, 2, \ldots, p$

ii) $H_0 : \theta_h = 0$

       $H_1 : \theta_h \neq 0$ with $h = 1, 2$

iii) $H_0 : \tau = 0$

       $H_1 : \tau \neq 0$

with the test statistics which used are:

$$Z_\beta = \frac{\hat{\beta}_{hk}}{se(\hat{\beta}_{hk})} \text{ where } se(\hat{\beta}_{hk}) = \sqrt{\hat{\text{var}}(\hat{\beta}_{hk})}$$

$$Z_\theta = \frac{\hat{\theta}_h - 1}{se(\hat{\theta}_{hk})} \text{ where } se(\hat{\theta}_{hk}) = \sqrt{\hat{\text{var}}(\hat{\theta}_{hk})}$$

$$Z_\tau = \frac{\hat{\tau}}{se(\hat{\tau})} \text{ where } se(\hat{\tau}) = \sqrt{\hat{\text{var}}(\hat{\tau})}$$

Where the value of $\sqrt{\hat{\text{var}}(\hat{\beta}_{hk})}$, $\sqrt{\hat{\text{var}}(\hat{\theta}_{hk})}$, and $\sqrt{\hat{\text{var}}(\hat{\tau})}$ are obtained from the main diagonal elements of the matrix $-\mathbf{H}^{-1}(\hat{\gamma})$. H_0 reject if $|Z_{score}| > Z_{\frac{\alpha}{2}}$.

C. Descriptive statistics of Response and predictor variable

In each sub-district in Trenggalek and Ponorogo Districts, the highest number of HIV and AIDS cases was 4 cases and almost 50% of the sub-districts had no cases or more than 50% of the sub-districts had cases. The highest mean and variance from the variable of percentage public health insurance and the lowest in the percentage health extension activities variable. This shows that there is still low awareness of the importance of health education, which is followed by a low percentage of reproductive-age couples (PUS) using condoms.

The average percentage for the population of the lowest level of education in a senior high school is 15.43%, with the highest in the Pudak sub-district. The average percentage of couples of reproductive ages using condoms is 3.283%, with the highest in the Tugu sub-district and the least in the Ngebel sub-district. The percentage of health education activities is 0.6407%, with the highest in the Pudak sub-district and the least in the Gandusari sub-district. The percentage of public health insurance is 42.75%, with the highest in the Pudak sub-district and the least in the Trenggalek sub-district.

Based on the value of the correlation coefficient that there are positive relationship patterns of the number cases of HIV and AIDS variable with the percentage of the age group 25-29 years, the percentage of the lowest level of education (SMA), and percentage of couples of reproductive ages using condoms. Meanwhile, the percentage of health education activities and the percentage of public health insurance are negative relationship pattern to the response variable.

1) Pearson Correlation Test

The two responses in the bivariate regression analysis must correlate. To determine whether there is a relationship is tested using the Pearson correlation coefficient with the following hypothesis[10].

$$H_0 : \rho_{y_1, y_2} = 0$$

$$H_1 : \rho_{y_1, y_2} \neq 0$$

Test Statistics.

$$T = \frac{r_{y_1, y_2}\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.851\sqrt{27-2}}{\sqrt{1-0.851^2}} = 8.102$$

The T value (8.102) is greater than $t_{0.025,25} = 2.059$ therefore H_0 is rejecting or the number cases of HIV and AIDS are correlated, so this research can continue with the bivariate model.

2) Multicollinearity Test

There are many ways to detect multicollinearity, one of these is the Variance Inflation Factor (VIF) value or a value that describes the increase in the variance of the estimated parameters between predictor variables. If the $VIF > 10$, it can be said that there is multicollinearity, on the other hand, if the $VIF < 10$, there is no multicollinearity[22]. The VIF value of predictor variables are less than 10 (X_1=1.22, X_2=2.32, X_3=1.93, X_4=1.06, and X_5=1.06). Therefore, the predictor variables do not experience multicollinearity, or there is no correlation between the predictor variables.

3) Overdispersion Test

This study conducted an overdispersion test to determine whether the data used violates the assumptions or not. One way that can do to detect the presence or absence of overdispersion is using the variance test (VT), where overdispersion occurs if VT>1. The value of VT can be calculated using[10].

$$VT_1 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\overline{y}} = \frac{47.4074074}{1.148148} = 41.29$$

$$VT_2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\overline{y}} = \frac{24.0740741}{0.814815} = 29.545$$

According to VT value for Y_1 and Y_2 (HIV and AIDS cases) have VT_1 and VT_2 are more than 1, therefore the data had

overdispersion problems, hence this study continued using the BZIPIGR model.

D. Regression Modeling of the Number of HIV and AIDS Cases in Trenggalek and Ponorogo Districts using the BZIPIG method.

The number of HIV and AIDS cases in Trenggalek and Ponorogo Districts is count data with overdispersion problems. Therefore, this research modeling using BZIPIGR. Simultaneous testing of parameters is carried out to determine whether the significant β_1, β_2 parameter jointly affects the response. Simultaneous testing of the parameters of the BZIPIGR model with a hypothesis

$H_0 : \beta_{1k} = \beta_{2k} = 0$ with $k = 1,2,\ldots,5$

H_1 : at least one of $\beta_{hk} = 0$ with $h = 1,2$ and $k = 1,2,\ldots,5$

The value of test statistics G (73.18336) is greater than $\chi^2_{(0.05;10)} = 18.30704$, hence rejecting H_0. The conclusion is predictor variables together affect the response.

After testing simultaneously, the next step is testing the parameters BZIPIGR model partially. The hypothesis for this testing is:

i) For β

$H_0 : \beta_{hk} = 0$; $h = 1,2$; $k = 1,2,3,4,5$

$H_1 : \beta_{hk} \neq 0$

ii) For θ

$H_0 : \theta_h = 0$; $h = 1,2$

$H_1 : \theta_h \neq 0$

iii) For τ

$H_0 : \tau = 0$

$H_1 : \tau \neq 0$

The partial hypothesis testing on the parameters of the BZIPIGR model can be seen in Table 2.

TABLE II
PARAMETER ESTIMATION OF BZIPIGR MODEL

Parameter	Estimate	Se	Z
$\beta_{1.0}$	-1.117	0.001417051	-788.257
$\beta_{1.1}$	-0.097	0.001422267	-68.201
$\beta_{1.2}$	0.064	0.001409661	45.401
$\beta_{1.3}$	-0.042	0.001405293	-29.887
$\beta_{1.4}$	-0.27	0.001415502	-190.745
$\beta_{1.5}$	-0.015	0.001417234	-10.584
$\beta_{2.0}$	-1.963	0.001515899	-1294.941
$\beta_{2.1}$	-0.019	0.001539334	-12.343
$\beta_{2.2}$	0.056	0.001515931	36.941
$\beta_{2.3}$	-0.014	0.001507484	-9.287
$\beta_{2.4}$	-0.01	0.001469292	-6.806
$\beta_{2.5}$	-0.022	0.001510367	-14.566
θ_1	0.05	1.3956E-07	358267.700
θ_1	0.057	2.59258E-07	219858.600
τ	0.006	9.81499E-05	61.131

$P-value$ of the all parameters are less than the $\alpha = 0.05$ (P-Value < 0.05), hence H_0 is rejecting. Therefore, all the predictor variables partially influence the number of HIV and AIDS cases in Trenggalek and Ponorogo Districts. They are the percentage of the age group 25-29 years, the percentage of the lowest level of education is the senior high school variables, the percentage of couples of reproductive ages using condoms, the percentage of health education activities, and the percentage of public health insurance.

The equation of the BZIPIGR model according to table 2 for the number of HIV cases is divided into 2 models, such as the model for Poisson state and the model for zero state. The Poisson state regression model for HIV as follows.

$$\log\left(\hat{\lambda}_{1i}\right) = -1.1169 - 0.0966X_{1i} + 0.0643X_{2i} - 0.0423X_{3i} + \\ -0.2703X_{4i} - 0.0150X_{5i}$$

The estimation of parameters in the equation above can be interpreted based on the independent variables' coefficient and coefficient sign. For instance, every 1% increase of the population aged 25-29 years $\left(X_1\right)$ will reduce the average number of HIV cases in each sub-district in Trenggalek and Ponorogo districts equal to $\exp(0.0966) = 1.1011$ times, assuming the other variables are constant. The interpretation of the other independent variables can carry out in the same way. The Poisson state regression model for AIDS as follows.

$$\log\left(\hat{\lambda}_{2i}\right) = -1.9625 - 0.0187X_{1i} + 0.0560X_{2i} - 0.0141X_{3i} + \\ -0.0103X_{4i} - 0.0221X_{5i}$$

Like the the first model above, can interpret every 1% increase of population aged 25-29 years $\left(X_1\right)$ will reduce the average number of AIDS cases in each sub-district in Trenggalek and Ponorogo districts equal to $\exp(-0.187) = 0.9815$ times, which assuming the other variables are constant. The interpretation of the other independent variables can carry out in the same way.

There are differences influence of variables by Wijaya, where in Wijaya's research there are 3 variables that have a positive effect on the number of HIV and AIDS, namely the percentage of age group 25-29 years, The percentage of couples of reproductive age using condoms, and The percentage of community health insurance (Jamkesmas), while the other 2 variables have a negative effect. In this study only the percentage of population with senior high school level has a positive effect (it can increase the number of HIV and AIDS as the percentage of this variable increases), while other variables have a negative effect.

The second model of BZIPIGR is zero state regression model for HIV and AIDS. The zero-state regression model for HIV is as follows:

$$\text{logit}\left(\hat{p}_{1i}\right) = -0.05\left(-1.1169 - 0.0966X_{1i} + 0.0643X_{2i} - 0.0423X_{3i} + \\ -0.2703X_{4i} - 0.0150X_{5i}\right) \\ = 0.0558 + 0.0048X_{1i} - 0.0032X_{2i} + 0.0021X_{3i} + 0.0135X_{4i} + \\ 0.0008X_{5i}$$

The logit model equation above can interpret the Opportunities in every district in Trenggalek and Ponorogo did not have HIV cases increased equal to exp (0,0048)=1.0048 times if there is an increase of 1% of the population aged 25-29 years (X_1), assuming other variables constant. Similarly, to the variable X_3, X_4, and X_5. However, if there is an increase of 1% of percentage of the population with senior high school level X2 with the assumption that the other variables are constant, then the chance of not having an HIV case will decrease by exp (0.0032)= 1.0032 times. This is related with the Poisson state regression model, which can increase the number of HIV is variable X_2.

The zero-state regression model for AIDS cases is as follows.

$$\begin{aligned}
\text{logit}\left(\hat{p}_{2i}\right) &= -0.057\left(-1.9625 - 0.0187 X_{1i} + 0.0560 X_{2i} - 0.0141 X_{3i} + \right.\\
&\qquad \left. -0.0103 X_{4i} - 0.0221 X_{5i}\right)\\
&= 0.1119 + 0.0011 X_{1i} - 0.0032 X_{2i} + 0.0008 X_{3i} + 0.0006 X_{4i}\\
&\qquad + 0.0013 X_{5i}
\end{aligned}$$

The logit model equation above can interpret the chance that each sub-district in Trenggalek and Ponorogo districts does not have AIDS cases increases by $\exp(0.0011) = 1.0011$ times if there is an increase of 1% of the population aged 25-29 years, assuming the other variables are constant. Similarly, to the variable X_3, X_4, and X_5. However, if there is an increase of 1% of percentage of the population with senior high school level X_2 with the assumption that the other variables are constant, then the chance of not having an AIDS case will decrease by exp (0.0032)= 1.0032 times. This is related to the Poisson state regression model, which can increase the number of AIDS variables X_2. Based on these two regression models (Poisson state and zero state), an increase in the number of HIV and AIDS can be caused by an increase in the percentage of the population aged 25-29 years (X_2).

According to Akaike Information Criterion Corrected (AICc) value. The AICc value uses if the objective of regression modeling is to identify the influencing factors. The AICc value using the BZIPIGR model obtained is 317.96. Previous research with the same data using the BZIPR method on the number of HIV and AIDS cases got an AICc value of 340.6977. Therefore, the data on the number of HIV and AIDS cases in Trenggalek and Ponorogo Districts is better using the BZIPIGR model.

IV. CONCLUSION

Based on the results and discussion above, it can conclude that: The estimation parameters of the BZIPIGR model using the MLE method has obtained that the first derivative does not yield closed form. Therefore, it is followed by numerical iteration using the Berndt Hall Hall Hausman (BHHH) algorithm. For Simultaneous hypothesis testing of the BZIPIGR model using the MLRT method has obtained test statistics that follow Chi-square distribution. Based on the AICc value were obtained from the BZIPIGR model, it has shown that the model is feasible to apply for data on the number of HIV and AIDS cases. All predictor variables significantly affected the number of HIV and AIDS cases.

Based on the Poisson state regression model, the variable whose effect is to increase the number of HIV and AIDS cases

is the percentage of the population with low education (SMA). The variables that can reduce the number of HIV and AIDS cases are the percentage of the population aged 25-29 years, the percentage of reproductive-age couples (PUS) using condoms, the percentage of health education activities, and the percentage of community health insurance (Jamkesmas). Based on the Zero state regression model,

REFERENCES

[1] R. Fitriani, L. N. Chrisdiana, and A. Efendi, "Simulation on the Zero Inflated Negative Binomial (ZINB) to Model Overdispersed, Poisson Distributed Data," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052025.

[2] R. D. Kusuma and Y. Purwono, "Zero-Inflated Poisson Regression Analysis On Frequency Of Health Insurance Claim PT. XYZ," in *Proceedings of the 12th International Conference on Business and Management Research (ICBMR 2018)*, 2019, vol. 72, no. Icbmr 2018, pp. 321–325, doi: 10.2991/icbmr-18.2019.52.

[3] D. Karlis and E. Xekalaki, "Mixed Poisson distributions," *Int. Stat. Rev.*, vol. 73, no. 1, pp. 35–58, 2005, doi: 10.1111/j.1751-5823.2005.tb00250.x.

[4] Purhadi, Sutikno, S. M. Berliana, and D. I. Setiawan, "Geographically weighted bivariate generalized Poisson regression: application to infant and maternal mortality data," *Lett. Spat. Resour. Sci.*, vol. 14, no. 1, pp. 79–99, 2021, doi: 10.1007/s12076-021-00266-5.

[5] L. Amaliana, U. Sa'Adah, and N. W. Surya Wardhani, "Modeling Tetanus Neonatorum case using the regression of negative binomial and zero-inflated negative binomial," *J. Phys. Conf. Ser.*, vol. 943, no. 1, 2018, doi: 10.1088/1742-6596/943/1/012051.

[6] S. Mardalena, Purhadi, J. T. D. Purnomo, and D. D. Prastyo, "Bivariate Poisson Inverse Gaussian Regression Model with Exposure Variable: Infant and Maternal Death Case Study," *J. Phys. Conf. Ser.*, vol. 1752, no. 1, 2021, doi: 10.1088/1742-6596/1752/1/012016.

[7] X. F. Huang, G. L. Tian, C. Zhang, and X. Jiang, "Type I multivariate zero-inflated generalized Poisson distribution with applications," *Stat. Interface*, vol. 10, no. 2, pp. 291–311, 2017, doi: 10.4310/SII.2017.v10.n2.a12.

[8] M. S. Nur, Purhadi, and A. Choiruddin, "Parameter Estimation and Hypothesis Testing of Geographically Weighted Bivariate Zero Inflated Poisson Inverse Gaussian Regression Models," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1115, no. 1, p. 012043, 2021, doi: 10.1088/1757-899x/1115/1/012043.

[9] J. M. Hilbe, *Cambridge Books Online*, vol. 48, no. 04. 2010.

[10] Q. Aini, Purhadi, and Irhamah, "Bivariate zero inflated generalized Poisson regression model in the number of pregnant maternal mortality and the number of postpartum maternal mortality in the Central Java Province in 2017," *J. Phys. Conf. Ser.*, vol. 1511, no. 1, 2020, doi: 10.1088/1742-6596/1511/1/012055.

[11] Dinas kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur Tahun 2018*. Jawa Timur, Indonesia: Dinas Kesehatan Provinsi Jawa Timur, 2019.

[12] Kementerian Kesehatan Republik Indonesia, "Infodatin (Situasi HIV dan AIDS di Indonesia)," *Pusdatin*, 2020, [Online]. Available: https://pusdatin.kemkes.go.id/resources/download/pusdatin/infodatin/infodatin-2020-HIV.pdf.

[13] I. Purnamasari and I. N. Latra, "Parameter Estimation and Statistical Test in Modeling Geographically Weighted Poisson Inverse Gaussian Regression," in *International Conference Research:Implementation and Education of Mathematics and Science*, 2016, no. May, pp. 16–17.

[14] S. Mardalena, P. Purhadi, J. D. T. Purnomo, and D. D. Prastyo, "Parameter Estimation and Hypothesis Testing of Multivariate Poisson Inverse Gaussian Regression," *Symmetry (Basel).*, vol. 12, no. 10, p. 1738, Oct. 2020, doi: 10.3390/sym12101738.

[15] J. Amalia, Purhadi, and B. W. Otok, "Application of geographically weighted bivariate poisson inverse Gaussian regression," *AIP Conf. Proc.*, vol. 2268, no. September 2020, doi: 10.1063/5.0017393.

[16] J. Pangulimang, "Parameter Estimation and Hypothesis Testing Geographically Weighted Bivariate Zero-Inflated Poisson," in *International Conference Research:Implementation and Education of Mathematics and Science*, 2016, no. May, pp. 16–17, [Online]. Available: http://seminar.uny.ac.id/icriems/sites/seminar.uny.ac.id.icriems/files/prosiding/M-12.pdf.

[17] S. M. Berliana, Purhadi, Sutikno, and S. P. Rahayu, "Parameter estimation and hypothesis testing of geographically weighted multivariate generalized poisson regression," *Mathematics*, vol. 8, no. 9, pp. 1–14, 2020, doi: 10.3390/math8091523.

[18] J. N. Gonçalves and W. Barreto-Souza, "Flexible regression models for counts with high-inflation of zeros," *Metron*, vol. 78, no. 1, pp. 71–95, 2020, doi: 10.1007/s40300-020-00163-9.

[19] H. Cho, T. Porras, D. Baik, M. Beauchemin, and R. Schnall, "Understanding the predisposing, enabling, and reinforcing factors influencing the use of a mobile-based HIV management app: A real-world usability evaluation," *Int. J. Med. Inform.*, vol. 117, pp. 88–95, 2018, doi: 10.1016/j.ijmedinf.2018.06.007.

[20] S. U. Wijaya, *Pendugaan Parameter dan Pengujian Hipotesis pada Bivariate Poisson Inverse Gaussian Regression (Studi Kasus: Jumlah Kasus Baru HIV dan AIDS di Kabupaten Trenggalek dan Ponorogo Tahun 2012)*. Indonesia: Repository ITS, 2017.

[21] S. M. Berliana, Purhadi, Sutikno, and S. P. Rahayu, "Multivariate generalized Poisson regression model with exposure and correlation as a function of covariates: Parameter estimation and hypothesis testing," *AIP Conf. Proc.*, vol. 2192, 2019, doi: 10.1063/1.5139171.

[22] N. Shrestha, "Detecting Multicollinearity in Regression Analysis," *Am. J. Appl. Math. Stat.*, vol. 8, no. 2, pp. 39–42, 2020, doi: 10.12691/ajams-8-2-1.