Jurnal Matematika dan Statistika serta Aplikasinya, Vol. 14, No. 1, 118-129, 2026
DOI:https://doi.org/10.24252/msa.v13i2.62054

Handling Overdispersion in Data on the Number of Poor People on the Island of Sulawesi Using the Negative Binomial Regression Model

Selvi Sri Ayuning Tyas[1], Ermawati[2]*, Wahidah Alwi[3], Adnan Sauddin[4]
[1,2,3,4]*Mathematics Study Program, Universitas Islam Negeri Alauddin Makassar, Gowa, Indonesia*
***Corresponding author**: ermawati@uin-alauddin.ac.id

***Submission date**: 17 October 2025, **Revision**: 29 December 2025, **Accepted**: 15 April 2026

ABSTRACT
This study focuses on analysing data on the number of poor people in Sulawesi Island in 2023, which is categorised as count data and follows a Negative Binomial distribution. The analysis results identify overdispersion in the data on the number of poor people. To address this issue, a Negative Binomial regression model is used. This study aims to identify factors that significantly influence the number of poor people on the island of Sulawesi in 2023 using the Negative Binomial regression method. The resulting model shows that the variables of population percentage (X_1), human development index (X_4), and average length of schooling (X_5) significantly influence the number of poor people. An increase in population percentage and HDI has the potential to increase the number of poor people, while an increase in average length of schooling can contribute to a decrease in the number of poor people.

KEYWORDS
Poor population, Negative binomial regression, Overdispersion

1. INTRODUCTION

Regression is a statistical method used to predict the relationship between dependent and independent variables. For data with dependent variables in the form of count data, Poisson regression is a commonly used method [1]. One of the basic assumptions of Poisson regression is that the data satisfy the assumption of equidispersion. However, in the application of research data, overdispersion often occurs, which is when the variance exceeds the mean. If it continues to be used, Poisson regression can cause standard errors to become biased and invalidate the model [2]. To overcome this, negative binomial regression was developed as an alternative, as it can handle overdispersion by combining the Poisson and Gamma distributions.

Poverty is one of the social issues that is often highlighted in society. The term poverty refers to a situation in which individuals or groups do not have the ability to meet minimum standards, both in terms of food and non-food items such as education, health, and other aspects of life. These standards are usually measured in terms of expenditure.

According to a report by the Central Statistics Agency, the number of Indonesians living below the poverty line in 2023 reached 25,90 million, with a particularly high proportion on the island of Sulawesi. Unlike other islands in Indonesia, which saw a decline in the percentage of people living in poverty, Sulawesi did not experience such a decline. The poverty rate on the island of Sulawesi in 2023 increased by 0,02% from the poverty rate in September 2022 [3]. In March 2023, the number of poor people on the island of Sulawesi reached 2.04 million, with a percentage of 7,89%.

Previous studies on poverty generally use Poisson regression to model count data, even though this method requires the assumption of equidispersion. In regional poverty data, this assumption is often not met due to overdispersion, which has the potential to produce biased parameter estimates.

Negative binomial regression has been developed to address the problem of overdispersion. However, the application of this method in poverty studies in Indonesia is still limited, particularly at the district/city level and in the Sulawesi region. Therefore, this study fills the research gap by applying negative binomial regression to analyze the factors that significantly affect the data on the number of poor people in Sulawesi Island in 2023 at the district/city level, thus providing a more appropriate methodological approach and stronger empirical evidence.

2. LITERATURE REVIEW

2.1 Poisson Distribution

A discrete random variable Y is said to follow a Poisson distribution with parameter $\lambda > 0$ if it has a probability density function defined according to the following distribution characteristics [3]:

$$f(y,\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \ \ y = 0,1,2,\cdots \tag{1}$$

with,
$\lambda =$ the average number of successes occurring within a specific time interval and area.
$e =$ euler (2.71828)
The mean and variance have the same value, namely $E(Y) = \lambda$ and $Var(Y) = \lambda$ [4].

2.2 Gamma Distribution

The Gamma distribution is a distribution for continuous random variables that have parameters $\theta > 0$ and $\beta > 0$, where θ and β are positive numbers [5]. The Gamma function of θ is denoted by $\Gamma(\theta)$, which is

$$\Gamma(\theta) = \int_0^{\infty} y^{\theta-1}e^{-y}dy \tag{2}$$

Suppose that y in the Gamma function in equation (2) is a variable that depends on the variables λ and β, namely $y = \lambda/\beta$, with $\beta > 0$, then it forms the following equation:

$$\Gamma(\theta) = \int_0^{\infty} \left(\frac{\lambda}{\beta}\right)^{\theta-1} e^{-\lambda/\beta}d\lambda \tag{3}$$

A continuous random variable λ is said to follow a Gamma distribution if it is described by two parameters, namely θ as the shape parameter and β as the scale parameter, and has a probability density function, which is defined as follows:

$$h(\lambda) = \begin{cases} \frac{1}{\Gamma(\theta)\beta^{\theta}}\lambda^{\theta-1}e^{-\lambda/\beta}, & 0 < \lambda < \infty \\ 0 & , other\ \lambda \end{cases} \tag{4}$$

where $\theta > 0$, $\beta > 0$

The mean and variance of the Gamma distribution are $\mu = \theta\beta$ and $\sigma^2 = \theta\beta^2$. From $\mu = \theta\beta$, reparameterization is performed so that $\beta = \mu\theta$. If θ is written as the inverse of, $\theta = 1/\theta$, then $\beta = \mu\theta$. Thus, in equation (4), the scale parameter is changed to $\mu\theta$ and the shape parameter is changed to $\frac{1}{\theta}$. The probability form can be written as follows:

$$h(\lambda) = \frac{1}{\Gamma\left(\frac{1}{\theta}\right)(\mu\theta)^{\frac{1}{\theta}}}\lambda^{\frac{1}{\theta}-1}e^{-\lambda/\mu\theta} \tag{5}$$

2.3 Negative Binomial Distribution

The negative binomial distribution can be analyzed using various approaches. Several approaches to this distribution can be seen as a series of Bernoulli trials and the result of a mixture distribution between Poisson and Gamma [6].

The Gamma distribution was chosen because it has properties as a prior distribution that is naturally conjugate to the Poisson distribution. The relationship between the conditional probability function $f(y|\lambda)$ and the prior function $h(\lambda)$ can be expressed in a general form that describes the integration between the two distributions in the Bayesian approach framework, as follows:

$$\begin{aligned} f(y|\lambda)h(\lambda) &= \frac{e^{-\lambda}\lambda^{y}}{y!}\cdot\frac{1}{\Gamma\left(\frac{1}{\theta}\right)(\mu\theta)^{\frac{1}{\theta}}}\lambda^{\frac{1}{\theta}-1}e^{-\lambda/\mu\theta} \\ &= \frac{1}{\Gamma\left(\frac{1}{\theta}\right)(\mu\theta)^{\frac{1}{\theta}}y!}\cdot\lambda^{y}\lambda^{\frac{1}{\theta}-1}\cdot e^{-\lambda}e^{-\frac{\lambda}{(\mu\theta)}} \\ &= \frac{\lambda^{y+\frac{1}{\theta}-1}e^{-\lambda+\left(-\frac{\lambda}{\mu\theta}\right)}}{\Gamma\left(\frac{1}{\theta}\right)(\mu\theta)^{\frac{1}{\theta}}y!} \end{aligned} \tag{6}$$

Thus, through the Poisson-gamma mixture distribution approach, the random variable Y can be expressed as having the following probability function form [7]:

$$Pr(Y=y) = \frac{\Gamma\left(y+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y!}\left(\frac{\mu}{\mu+\frac{1}{\theta}}\right)^{y}\left(\frac{\frac{1}{\theta}}{\mu+\frac{1}{\theta}}\right)^{\frac{1}{\theta}} \tag{7}$$

where,
$Pr(Y=y)$: The probability that the random variable Y takes the value y, $y=0,1,2,\cdots$
θ: Dispersion parameter
μ: Mean of the distribution

Kolmogorov-Smirnov is a statistical method that aims to evaluate whether empirical data distribution conforms to a specific theoretical distribution, such as the normal distribution, Poisson distribution, or other distributions. This test is based on the following hypothesis:
H_0: data not distributed according to the negative binomial distribution
H_1: negative binomial distributed data.

Following test statistics.

$$KS_{hit} = max_{y}\left|F_n(y) - F_0(y)\right|; y\in\mathbb{R} \tag{8}$$

and

$$F_n(y) = \frac{number\ of\ observations \leq y}{n}$$
$$F_0(y) = \begin{pmatrix} y+k-1 \\ k-1 \end{pmatrix} p^{k}(1-p)^{y},\ y=0,1,2,\cdots$$

Where,
$F_n(y)$: Frequency distribution
$F_0(y)$: Cumulative relative frequency of data
k: Number of failed attempts before the y-th success
p: Chance of success in each attempt
Test criteria: $p-value > \alpha$ with $\alpha = 0.05$

2.4 Overdispersion

The regression model most often applied to count data is Poisson regression. However, this model is only appropriate if the data shows the same distribution between the mean and the variance. In practice, it is common to find values where the variance is greater than the mean, a phenomenon known as overdispersion [8].

Testing for overdispersion can be done using the following formula [9].

$$\theta = \frac{value\ of\ deviance}{df} \tag{9}$$

with,
df: degree of freedom

$$D = 2\sum_{i=1}^{n}\left(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right)$$

where,
D: Deviance
n: Number of observations
i: Observation index i
y_i: Observed value in the i-th data point
$\hat{\mu}_i$: Predicted value for the i-th data point
If the value of $\theta > 1$, then overdispersion occurs, and if $\theta < 1$, then underdispersion occurs [1].

2.5 Multicollinearity

Multicollinearity occurs when two or more independent variables in a regression model have a high linear relationship with each other. To detect this problem, one commonly used diagnostic measure is Variance Inflation Factor (VIF). A VIF value exceeding 10 generally indicates multicollinearity [10]. The basic principle in testing multicollinearity is explained as follows:
H_0: There is no multicollinearity
H_1: There is multicollinearity
One way to identify multicollinearity is by looking at the VIF value, which can be defined as follows [11].

$$VIF = \frac{1}{1 - R_j^2} \tag{10}$$
$$R_j^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

where,
SST: Sum of Squares Total
SSE: Sum of Square Errors
SSR: Sum of Squares for Regression
Reject H_0 if the VIF value of an independent variable is less than 10. This indicates that there is no significant multicollinearity [1].

2.6 Negative Binomial Regression

Negative binomial regression is a form of nonlinear regression model developed through a combination of Poisson and Gamma distributions. This distribution has a dispersion parameter that allows for overdispersion [9]. The following is the probability function of the negative binomial distribution.

$$Pr(Y = y) = \frac{\Gamma\left(y + \frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) y!}\left(\frac{\mu}{\mu + \frac{1}{\theta}}\right)^{y}\left(\frac{\frac{1}{\theta}}{\mu + \frac{1}{\theta}}\right)^{\frac{1}{\theta}} Pr(Y = y) = \frac{\Gamma\left(y + \frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) y!}\left(\frac{\theta\mu}{1 + \theta\mu}\right)^{y}\left(\frac{1}{1 + \theta\mu}\right)^{\frac{1}{\theta}} \tag{11}$$

Referring to the components in Generalized Linear Model (GLM), the negative binomial regression model formulation can be expressed as follows [12].

$$\ln(\mu_i) = \boldsymbol{X}_i^T\boldsymbol{\beta} = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$
$$\mu_i = \exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right) = \exp\left(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}\right), \quad i = 1, 2, \cdots, n \tag{12}$$

with, $\boldsymbol{X}_i = [1, X_{1i}, X_{2i}, \cdots, X_{ki}]^T$ and $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_k]^T$ where,
X_{ki}: Independent variable located at index k in observation unit i in the data set.
β_0 : Constant parameter
$\beta_1,\ \beta_2, \cdots, \beta_k$: Regression parameters for each independent variable
μ_i: The average number of successes occurring in observation i

Parameters in the negative binomial regression model, such as β_0,β_1,$\cdots$, and β_k. The value is unknown, so it needs to be estimated. The estimation of these parameters is done using the maximum likelihood estimation approach [11].

Suppose there is random observation data, consisting of pairs of observation data $(y_i, (x_{1i}, x_{2i}, \cdots, x_{ki}))$, for $i = 1, 2, \cdots, n$. Next, the value μ_i is substituted into the conditional probability function of the random variable Y_i, taking into account the effect of covariates $(x_{1i}, x_{2i}, \cdots, x_{ki})$. Based on the distribution form as shown in equation (11), the complete form of the marginal probability function can be written as follows.

$$f(y_i, \boldsymbol{\beta}, \theta) = \frac{\Gamma\left(y_i + \frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) y_i!} \left(\frac{\theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}\right)^{y_i} \left(\frac{1}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}\right)^{1/\theta}$$

The steps for interpreting the parameters are as follows:

1. Form the likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}, \theta) &= \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta}, \theta) \\ L(\boldsymbol{\beta}, \theta) &= \prod_{i=1}^{n} \frac{\Gamma\left(y_i + \frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) y_i!} \left(\frac{\theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}\right)^{y_i} \left(\frac{1}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}\right)^{1/\theta} \end{aligned} \tag{13}$$

2. Convert the likelihood into a natural logarithm function, taking into account the properties of logarithms

$$\begin{aligned} \ln \prod_{i=1}^{n} f_i &= \sum_{i=1}^{n} \ln f_i \\ &= \sum_{i=1}^{n} \ln \left[\frac{\prod_{r=0}^{y_i - 1} (1 + \theta r)}{\theta^{y_i} y_i!} \left(\frac{\theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}\right)^{y_i} \left(\frac{1}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}\right)^{1/\theta} \right] \end{aligned} \tag{14}$$

3. The estimation process of the parameters, β and θ, is carried out by minimizing the partial derivatives of the log-likelihood function. Then, maximize the log-likelihood function by differentiating the function with respect to each parameter and setting the derivative equal to zero.

   The first derivative of the likelihood function with respect to the parameter

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left(\frac{(y_i - \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})) \boldsymbol{X}_i^T}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})} \right) \tag{15}$$

   The first derivative of the likelihood function with respect to the parameter

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{n} \left(\sum_{r=0}^{y_i - 1} \frac{r}{1 + \theta r} + \theta^{-2} \ln(1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})) - \frac{(y_i + \theta^{-1}) \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})}{1 + \theta \exp(\boldsymbol{X}_i^T \boldsymbol{\beta})} \right) \tag{16}$$

   Equations (15) and (16) cannot be solved directly using analytical methods, so a numerical approach is used. One technique commonly used to solve these equations is the Newton-Raphson method.

2.7 Model Fitness Test

The purpose of model fit testing is to evaluate whether the model obtained is consistent with the observed data. The approach used to test the fit of the negative binomial regression model is deviance [13], which is formulated as follows.

$$D = 2\sum_{i=1}^{n}\left(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i+\theta)\ln\left(\frac{y_i+\theta}{\hat{\mu}_i+\theta}\right)\right) \tag{17}$$

Model suitability testing was conducted based on the following hypotheses:
H_0: The model obtained is not relevant to the research data
H_1: The model obtained is relevant to the research data
With a significance level of $\alpha = 0.05$
The testing decision rejects H_0 if the deviance value exceeds the significance limit.

2.8 Testing the Parameters of the Negative Binomial Regression Model

This testing process can be carried out through simultaneous and partial testing [14]. Simultaneous testing evaluates whether all independent variables collectively have a significant effect on the dependent variable. This test is carried out using the following hypothesis:
H_0:$\beta_1 = \beta_1 = \cdots = \beta_j = 0$, which means that no independent variables significantly affect the dependent variable.
H_1: There is at least one $\beta_j \neq 0$; for $j = 1,2,3,\cdots k$, which shows that at least one independent variable has a significant effect.
To determine the test statistic value, the following is used:
The decision rule is H_0 rejected, if the value of G is greater than $\chi^2_{(k,\alpha)}$, where k is the number of independent variables.
Partial testing was conducted to evaluate the individual influence of each independent variable on the dependent variable.
The hypotheses applied in this test are as follows.
$H_0 : \beta_j = 0$; for $j = 1,2,\cdots,k$
H_0:$\beta_j \neq 0$
Statistics test

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{18}$$

With a significance level of $\alpha = 0.05$, the test decision is that H_0 will be rejected if $|Z| > Z_{\alpha/2}$.

3. METHODOLOGY

This study uses secondary data obtained from BPS publications in each province included in the scope of Sulawesi Island in 2023. The variable used in this study is the number of poor people at the regency/city level in Sulawesi Island in 2023 (Y), Percentage of the population (X_1), the open unemployment rate (X_2), the growth rate of regional gross domestic product at constant prices (X_3), the human development index (X_4), and the average length of schooling (X_5).

3.1 Data Analysis Procedure

The analysis procedures applied in this study were as follows:

1. Identify and explain the relationship patterns between dependent and independent variables.
2. To support the application of the negative binomial regression model, the following assumptions need to be tested:
   (a) Test whether the data on the number of poor people in Sulawesi Island in 2023 follows a negative binomial distribution.
   (b) Examine the assumption of equidispersion.
   (c) Examine the assumption of multicollinearity among independent variables.

3. Developing a negative binomial regression model to analyze the number of poor people on the island of Sulawesi. This process involves several steps, namely:
   (a) Forming a negative binomial regression model based on the parameter estimation results obtained from the data analysis process.
   (b) Model suitability test
   (c) Conducting significant testing of parameters in the model to determine their level of significance, consisting of:
      i. Simultaneous parameter significance test.
      ii. Partial parameter significance test.
   (d) Determining the variables that significantly affect the number of poor people on the island of Sulawesi in 2023 based on the results of the negative binomial regression model analysis.

4. RESULT & DISCUSSION

Table 1 presents information about the average, minimum, maximum, and variance values of each variable that is the object of this study.

Table 1. Descriptive Statistics

Variable	Average	Variance	Min	Max
Y	25222	298496200	4320	80320
X_1	7.403	36.203	1.42	33.45
X_2	3.574	3.554	1.16	10.60
X_3	4.593	13.368	0.55	23.04
X_4	72.519	16.917	66.94	85.51
X_5	8.823	1.314	7.00	12.53

Based on the data presented in **Table 1**, it can be seen that the mean of the dependent variable is 25222 and the variance is 298406200. Since these two values are not equal, the data does not meet the assumption of equidispersion. A condition where the variance exceeds the mean indicates overdispersion in the data.

4.1 Negative Binomial Regression Assumption Test

1. Negative Binomial Distribution Test

   This test is based on the hypothesis that:
   H_0: The data is not negatively binomially distributed H_1: The data is negatively binomially distributed The following test statistics:

   $$KS_{hit} = max_y \left|F_n(y) - F_0(y)\right|; y \in \mathbb{R}$$

   Test criteria:
   $p - value > \alpha$, with $\alpha = 0.05$

Table 2. Negative Binomial Distribution Test

Test	Test Statistic	p-value
Kolmogorov–Smirnov	0.069	0.831

Based on **Table 2**, the Kolmogorov-Smirnov test results show that the p-value is greater than 0.05, which is the basis for rejecting H_0. This statistical decision indicates that the Y variable for the data on the number of poor people on the island of Sulawesi in 2023 follows a negative binomial distribution.

2. Overdispersion Test

   Overdispersion can also be identified by calculating the ratio between deviance and degrees of freedom, with the following hypotheses:
   H_0: Meets the equidispersion condition
   H_1: Meets the overdispersion or underdispersion condition Test statistic:

$$\theta = \frac{deviance\ value}{df}$$

   The following are the values from the equidispersion test results on the data.

$$\theta = \frac{deviance\ value}{df} = \frac{507976.1}{75} = 6773.01$$

   Since the resulting ratio is greater than 1, H_0 it is rejected. Therefore, it can be concluded that the 2023 data on the number of poor people in Sulawesi Island exhibits overdispersion.

3. Multicollinearity Test

   The hypothesis framework used in this test is as follows:
   H_0: There is no multicollinearity
   H_1: There is multicollinearity

   The following test statistics:

$$VIF = \frac{1}{1-R_j^2}$$

   Test criteria:
   Reject H_0 if the independent variable has a VIF<10.

Table 3. Variance Inflation Factor (VIF) of Independent Variables

Variable	VIF
X_1	1.015
X_2	2.132
X_3	1.017
X_4	4.341
X_5	3.461

Based on **Table 3**, all independent variables show VIF values less than 10. This finding indicates that there is no multicollinearity, which means that there is no significant linear relationship between the independent variables [1].

4.2 Negative Binomial Regression Modeling on Data on the Number of Poor People in Sulawesi Island in 2023

Table 4 shows the parameter estimation results from the model analyzed using the negative binomial approach. The following negative binomial regression model is constructed based on the results shown in**Table 4** above, namely:

$$\ln(\hat{\mu}_i) = 5.593 + 0.053X_1 - 0.042X_2 + 0.002X_3 + 0.111X_4 - 0.437X_5$$
$$\hat{\mu}_i = \exp(5.593 + 0.053X_1 - 0.042X_2 + 0.002X_3 + 0.111X_4 - 0.437X_5)$$

Model suitability testing is conducted first to ensure the model's feasibility before it is used in further analysis.

Table 4. Estimated Parameter Values from Negative Binomial Regression

Parameter	Estimate	Standard Error
β_0	5.593	0.000
β_1	0.053	0.000
β_2	-0.042	0.325
β_3	0.002	0.890
β_4	0.111	0.000
β_5	-0.437	0.000

4.3 Negative Binomial Regression Model Fit Test

This test aims to determine whether the estimated model fits the pattern found in the data. The hypothesis for this test is stated as follows:

H_0: The model obtained is not relevant to the research data

H_1: The model obtained is relevant to the research data

Test statistic:

$$D = 2\sum_{i=1}^{n}\left(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i+\theta)\ln\left(\frac{y_i+\theta}{\hat{\mu}_i+\theta}\right)\right)$$

Test criteria:

$deviance > \alpha$, with $\alpha = 0.05$

Table 5. Negative Binomial Regression Model Fit Test

Indicator	df	Deviance	Deviance/df
Deviance	75	84.217	1.123

The ratio between deviance and degrees of freedom listed in **Table 5** shows a value of 1.123, which is greater than $\alpha = 0.05$, so H_1 cannot be rejected. This indicates that the resulting model is consistent with the observed data.

4.4 Significance Testing of Negative Binomial Regression Parameters

1. Simultaneous Test

   The hypothesis applied in this test is as follows:

   H_0:$\beta_0 = \beta_1 = \cdots = \beta_5 = 0$, which means that no independent variables significantly affect the dependent variable.

   H_1: There is at least one $\beta_j \neq 0$; for $j = 1,2,3,\cdots,5$, which shows that at least one independent variable has a significant effect.

   Test statistic:

   $$G = 2(\ln L\left(\hat{\Omega}\right) - \ln L\left(\hat{\omega}\right))$$

   Test criteria:

   $G > \chi^2_{(5,\alpha)}$ or $p - value > \alpha$, $\alpha = 0.05$

Table 6. Simultaneous Test

Likelihood Ratio Chi-Square (G)	$\chi^2_{0.05,5}$
43.275	11.070

Referring to **Table 6**, it is known that the G statistic value is 43.275. This G statistic value is greater than the $\chi^2_{(0.05;5)}$ value of 11.070. This result provides sufficient statistical evidence to reject H_0, indicating that at least one parameter in the negative binomial regression model has a significant effect on the dependent variable.

2. Partial Test

   The testing continued with a partial analysis to test the significance of each parameter, with the following hypotheses:
   H_0: $\beta_j = 0$; $j = 1, 2, \cdots, 5$
   H_1: $\beta_5 \neq 0$

   Test statistic:

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

   Test criteria :
   $Z > Z_{\alpha/2}$ or $p - value < \alpha$

Table 7. Estimation Coefficients in the Negative Binomial Regression Approach

Parameter	Estimate	Standard Error	Z-Statistic	p-value
β_0	5.593	1.459	3.834	0.000
β_1	0.053	0.009	5.763	0.000
β_2	-0.042	0.042	-0.980	0.320
β_3	0.002	0.015	0.139	0.890
β_4	0.111	0.028	4.010	0.000
β_5	-0.437	0.089	-4.920	0.000

Based on the data in **Table 7** and using a significance level of 5%, it was found that the independent variables showed a p-value less than $\alpha = 0,05$. This finding leads to the rejection of H_0. Thus, it can be concluded that the variables of population percentage (X_1), human development index (X_4), and average length of schooling (X_5) have a significant effect on the number of poor people on the island of Sulawesi.

4.5 Analysis of Negative Binomial Regression Models Based on Significant Independent Variables

Based on the results of the partial tests conducted previously, it is known that only X_1, X_4, and X_5 have a significant effect on the dependent variable. Therefore, the negative binomial regression model was re-estimated by only including variables that were proven to be significant. The simplified model is shown in Table 4.8:

Table 8. Testing Significant Parameters

Parameter	Estimate	Standard Error	p-value
β_0	6.515362	1.216	0.000
β_1	0.051862	0.009	0.000
β_4	0.097119	0.025	0.000
β_5	-0.440666	0.089	0.000

The negative binomial regression model obtained based on the results in **Table 8** can be expressed as follows:

$$\ln(\hat{\mu}_i) = 6.515362 + 0.051865X_1 + 0.097119X_4 - 0.440666X_5$$
$$\hat{\mu}_i = \exp(6.515362 + 0.051865X_1 + 0.097119X_4 - 0.440666X_5)$$

Based on the results of the equation estimation, the percentage of population variable (X_1) with a parameter value of $\hat{\beta}_1 = 0.051862$ obtained an exponential value of 1.0532. This value indicates that every increase in the percentage of the population has the potential to increase the average number of poor people, assuming that other variables remain constant.

For the human development index variable (X_4) with a parameter value of $\hat{\beta}_4 = 0.097119$, an exponential value of 1.1020 was obtained. This value indicates that every increase in the human development index can increase the average number of poor people, assuming that other variables remain constant.

For the variable of average length of schooling (X_5) with a parameter value of $\hat{\beta}_5 = -0.440666$, an exponential value of 0.6435. This value indicates a statistically significant negative relationship with the number of poor people. This means that every increase in the average length of schooling has the potential to reduce the average number of poor people, assuming that other independent variables remain constant [15].

This study uses negative binomial regression, which is appropriate for count data with overdispersion, resulting in more accurate parameter estimates than Poisson regression. In addition, this study successfully identified the variables of population percentage, human development index, and average length of schooling as factors that significantly affect the number of poor people on the island of Sulawesi. However, this study only used data from 2023, so it was not able to capture the dynamics of poverty over time. The explanatory variables used are still limited, and the model used does not consider spatial dependence between regions, which could potentially affect the estimation results. Therefore, future studies are advised to include additional independent variables that may be relevant in efforts to reduce the number of poor people. In addition, modeling that considers the spatial characteristics of each district/city on the island of Sulawesi is also worth considering, for example by applying the Geographically Weighted Negative Binomial Regression (GWNBR) method.

5. CONCLUSION

The application of the negative binomial regression model to data on the number of poor people in Sulawesi Island in 2023 shows the following regression model form.

$$\ln(\hat{\mu}_i) = 5,593 + 0,053X_1 - 0,042X_2 + 0,002X_3 + 0,111X_4 - 0,437X_5$$

The results of the analysis using the negative binomial regression approach show that there are three independent variables that have a statistically significant effect on the number of poor people, namely the percentage of population (X_1), the human development index (X_4), and the average length of schooling (X_5).

REFERENCES

[1] R. Ruliana, P. Hendikawati, and A. Agoestanto, "Pemodelan generalized poisson regression (gpr) untuk mengatasi pelanggaran equidispersi pada regresi poisson kasus campak di kota semarang tahun 2013," *UNNES Journal of Mathematics*, vol. 5, no. 1, pp. 39–46, 2016. [Online]. Available: http://journal.unnes.ac.id/sju/index.php/ujm

[2] W. Wang and F. Famoye, "Modeling household fertility decisions with generalized poisson regression," *Journal of Population Economics*, vol. 10, no. 3, pp. 273–283, 1997.

[3] A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data*. New York: Cambridge University Press, 1998.

[4] P. D. Nugroho and Sigit, *Pengantar Statistika Matematika*, 1st ed. Bengkulu: UNIB Press, 2008.

[5] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*. Macmillan Publishing Co., Inc., 1978.

[6] J. M. Hilbe, *Negative Binomial Regression*. New York: Cambridge University Press, 2008. [Online]. Available: https://archive.org/details/negativebinomial0000hilb/mode/1up

[7] W. Greene, "Functional forms for the negative binomial model for count data," *Economics Letters*, vol. 99, no. 3, pp. 585–590, 2008.

[8] B. Durmuş and I. İşçi Güneri, "An application of the generalized poisson model for over dispersion data on the number of strikes between 1984 and 2017," *Alphanumeric Journal*, vol. 8, no. 2, pp. 249–260, 2020.

[9] M. A. Ma'sum, Suparti, and D. Ispriyanti, "Perbandingan model regresi binomial negatif dengan model geographically weighted poisson regression (gwpr): Studi kasus angka kematian ibu di provinsi jawa timur tahun 2011," *Jurnal Gaussian*, vol. 2, pp. 259–267, 2013.

[10] Sihabudin *et al.*, *Ekonometrika Dasar: Teori dan Praktik Berbasis SPSS*. Pena Persada, 2021.

[11] U. Kurniawan, "Penaksiran dan pengujian hipotesis parameter model regresi binomial negatif bivariat: Studi kasus jumlah kematian bayi dan jumlah kematian ibu di provinsi jawa timur tahun 2013," 2015, undergraduate Thesis.

[12] A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd ed. Canada: John Wiley & Sons, 2009.

[13] Irwan and Sari, "Pemodelan regresi poisson dan binomial negatif pada kasus kecelakaan kendaraan bermotor di lalu lintas sumatera barat," no. November, pp. 978–979, 2013.

[14] Fitrial, "Pemodelan jumlah kasus covid-19 di indonesia dengan pendekatan regresi poisson dan regresi binomial negatif studi kasus 34 provinsi di indonesia," pp. 65–72, 2020.

[15] B. Louis, T. Priharyanto, and U. Alifah, "Pemodelan jumlah kekerasan terhadap perempuan di jawa timur dengan regresi poisson dan binomial negatif (modeling the number of violence cases against women in east java using poisson and negative binomial regression)," *JSN: Jurnal Sains Natural*, 2024.