Journal of Mathematics and Scientific Computing with Applications

E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 7, Number 1, 2026
DOI: 10.53806/jmscowa.v7i1.1443
Page: 40–52

APPLICATION OF JACKKNIFE NEGATIVE BINOMIAL RIDGE REGRESSION IN MODELING TODDLER STUNTING IN NORTH SUMATRA

Elma Dwi Ariana Aprilia Zam[1], **Ismail Husein**[2], **Rina Widyasari**[3]

[1,2,3]Department of Mathematics, Universitas Islam Negeri Sumatera Utara

Article Info

Article history:
Received : April 20, 2026
Revised : May 21, 2026
Accepted : June 17, 2026

Keywords:
Binomial Negative Regression;
Jackknife Ridge Regression;
Multicollinearity;
Stunting in Toddlers.

ABSTRACT

This study aims to model stunting cases using cross-sectional count data from 33 districts/cities in North Sumatra Province in 2023. To account for differences in population size across regions, a log(Balita) offset was incorporated into the model, allowing the outcome to be interpreted in terms of stunting incidence rates rather than absolute counts. The presence of overdispersion and multicollinearity among explanatory variables creates challenges for standard count regression models and may lead to unstable parameter estimates. Therefore, the Jackknife Negative Binomial Ridge Regression (JNBR) approach was employed to improve estimation reliability under correlated predictors and overdispersed count data. This study specifically focuses on determining the optimal ridge parameter within the JNBR framework based on the Mean Squared Error (MSE) criterion. The results show that the optimal JNBR estimator with ridge parameter $k = 0.0435$ produced the smallest MSE value (48.8816) compared with the standard Negative Binomial Maximum Likelihood Estimator (94.1042), indicating improved estimation stability under multicollinearity conditions. The findings suggest that the JNBR approach provides a more stable and efficient estimation framework for modeling stunting incidence rates in the presence of multicollinearity and overdispersion. These results support the suitability of the proposed method for complex count data structures in public health applications.

Corresponding Author:
Elma Dwi Ariana Aprilia Zam,
Department of Mathematics,
Universitas Islam Negeri Sumatera Utara,
Email: elma0703212028@uinsu.ac.id

1. INTRODUCTION

Stunting is a major obstacle to human capital development because it is associated with impaired physical growth, cognitive development, and long-term health outcomes. According to the child growth standards established by the World Health Organization in 2006, stunting in children under five is defined as height-for-age below the median standard [1]. In 2020, the prevalence of stunting in Indonesia reached 27.6%, exceeding international targets and indicating persistent disparities in nutrition, sanitation, and access to basic health services [2]. In North Sumatra Province, the prevalence of stunting decreased from 25.8% in 2021 to 21.1% in 2022 and 18.9% in 2023, based on data from the Indonesian Nutritional Status Survey (SSGI). In 2024, the Government of North Sumatra Province targeted a further reduction in stunting prevalence to 14%.

Stunting is recognized as a serious public health problem because it has substantial adverse effects on child development and long-term productivity [3]. Previous studies have shown that stunting is associated with limited developmental progress, lower cognitive performance, reduced school participation, increased mortality risk, and higher susceptibility to infectious diseases [4]. Another study conducted in Grobogan Regency reported that nutritional status, child health problems, processed food consumption, and maternal height were significantly associated with stunting among children under five years of age [5]. Factors associated with stunting are multifactorial, involving maternal health conditions, early-life nutrition, immunization coverage, environmental sanitation, and socioeconomic factors [3]. Therefore, identifying factors associated with stunting is important for supporting effective intervention strategies and evidence-based public health policies. Because stunting is influenced by multiple interrelated maternal and child health indicators, statistical modeling often involves correlated predictors and complex count data structures.

In the GLM framework, count data are modeled using a Poisson distribution with a log link. In this study, the model incorporates a population offset term defined as log(number of under-five children) to account for differences in exposure across districts/cities, allowing the model to estimate the incidence rate of stunting rather than raw counts. The Poisson regression model assumes equidispersion, meaning that the mean is equal to the variance. However, this assumption is frequently violated because empirical count data often exhibit overdispersion. Under overdispersion, standard Poisson regression may underestimate the variance of parameter estimates, leading to unreliable statistical inference. In relatively small samples, overdispersion and multicollinearity may further increase the instability of the Maximum Likelihood Estimation in count regression models. Therefore, the Negative Binomial regression model is often considered a more appropriate alternative for overdispersed count data [6],[7].

In addition to overdispersion, multicollinearity among socioeconomic and maternal-child health predictors may lead to unstable parameter estimation and inflated variances in regression coefficients. Ridge regression, originally introduced by Hoerl and Kennard [8], was developed to address multicollinearity by incorporating a shrinkage parameter into the estimation process. Subsequent developments extended estimation of the ridge-type to generalized linear models, including Poisson and Negative Binomial regression models, to improve estimation stability under correlated predictors and overdispersed count data [9],[10]. Furthermore, jackknife bias correction methods have been widely studied to reduce finite-sample bias in regression estimators [11],[12]. The combination of ridge shrinkage and jackknife correction may improve estimation stability and reduce estimator variance.

Several modern regularization techniques have been proposed to address multicollinearity, including LASSO, Elastic Net, and Bayesian shrinkage methods. LASSO performs variable selection by shrinking some coefficients to zero, while Elastic Net combines L1 and L2 penalties to improve prediction performance in correlated predictors [13]. However, these approaches are primarily designed for prediction and variable selection, and may excessively shrink or eliminate correlated predictors that remain substantively important in epidemiological studies. In studies of maternal and child health, retaining all predictors is often preferable for interpretation purposes because each variable may represent an important public health indicator. In addition, previous studies have shown that the performance of ridge based estimators strongly depends on the selection of the ridge parameter and the characteristics of the observed data [14]. Several studies comparing ridge regression and regularization approaches such as LASSO reported that ridge-based estimators tend to provide more stable coefficient estimates under severe multicollinearity conditions, particularly when retaining all predictors is important for interpretation purposes [15],[16]. Therefore, determining the optimal ridge parameter becomes an important step in obtaining stable parameter estimates within the JNBR framework.

Although ridge-type estimators and jackknife bias correction methods have been extensively studied in generalized linear models, studies specifically investigating the performance of Jackknife Negative Binomial Ridge Regression (JNBR) for modeling stunting cases under simultaneous overdispersion and multicollinearity remain limited. In addition, limited attention has been given to the selection of optimal ridge parameters within the JNBR framework for public health count data. Therefore, this study aims to model the number of stunting cases among children under five years of age in North Sumatra Province using the Jackknife Negative Binomial Ridge Regression estimator to accommodate overdispersed count data and address multicollinearity simultaneously. In addition, this study focuses on determining the optimal ridge parameter within the JNBR framework based on the Mean Squared Error (MSE) criterion in order to obtain more stable parameter estimates.

2. RESEARCH METHOD

This study used secondary cross-sectional data for the year 2023 obtained from the North Sumatra Provincial Health Office. The response variable represents the number of stunting cases at the district/city level in North Sumatra Province. To account for differences in population size across districts/cities, a population offset term defined as log (Balita) was incorporated into the regression model. This specification allows the model to estimate the incidence rate of stunting rather than the absolute number of cases. To address overdispersion and multicollinearity among predictors, the Jackknife Negative Binomial Ridge (JNBR) regression approach was employed. Ridge regularization was applied to reduce the instability of parameter estimates caused by high intercorrelations among explanatory variables, while jackknife resampling was used to improve estimation stability and reduce bias. The variables used in the analysis are presented in Table 1.

Table 1: Description of Variables

Variable	Description
Y	Number of stunting cases in each district/city
X_1	Number of infants receiving exclusive breastfeeding
X_2	Number of infants with low birth weight
X_3	Number of pregnant women receiving iron tablets
X_4	Number of pregnant women with chronic energy deficiency
X_5	Number of children receiving early initiation of breastfeeding
X_6	Number of infants receiving vitamin A

The analytical procedures in this study are outlined as follows:

a. Data on the number of stunting cases were collected from the Health Office of North Sumatra Province.
b. Descriptive statistical analysis of the dataset.
c. Estimation of the Poisson regression model.
In Poisson regression, the dependent variable representing the number of occurrences of an event is assumed to follow a Poisson distribution conditional on the independent variables $X_{1i}, X_{2i}, \ldots, X_{ki}$. Since the mean of the Poisson distribution must be positive, a log link function is used to relate the expected value of the response variable $(E(Y_i) = \lambda_i)$ to a linear combination of the explanatory variables, which may be continuous, ordinal, nominal, or a combination thereof.

A Poisson regression model with a log link can be expressed in the following form [17]:

$$Y_i \sim \text{Poisson}(\mu_i)$$

And an offset term can be expressed as follows

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \ln(t_i) \tag{1}$$

where

$$\mathbf{x}_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix}^T, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_p \end{bmatrix}^T$$

for $i = 1, 2, \ldots, n$, where Y_i denotes the number of events at the i-th observation, $\mu_i = \mathbb{E}(Y_i)$ is the expected number of events, and t_i is the exposure variable.

d. Overdispersion testing
Overdispersion in Poisson regression arises when the variance exceeds the mean. Consequently, applying Poisson regression to overdispersed data may lead to invalid inferences, as standard errors are typically underestimated [7]. To assess equidispersion, the following formula is employed:

$$\phi = \frac{\text{deviance}}{\text{degrees of freedom}} \tag{2}$$

A value of the deviance divided by the degrees of freedom greater than one indicates overdispersion, whereas a value less than or equal to one suggests equidispersion [18].

e. Negative Binomial regression modeling
In the Poisson distribution, the mean and variance are equal. However, in some cases the observed data exhibit a variance greater than the mean, a condition commonly referred to as overdispersion. Therefore, the Negative Binomial distribution plays an important role in parametric statistical analysis to handle overdispersed data.

Negative Binomial regression is formulated using a log link function. When incorporating an offset term, the model is expressed as follows:

$$\ln(\mu_i) = \ln(E_i) + \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \tag{3}$$

where E_i denotes the exposure variable representing the number of under-five children in each district/city. The probability mass function of the Negative Binomial distribution is given by [19]:

$$f(y,\mu,\alpha) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(\alpha^{-1})\, y!} \left(\frac{1}{1+\alpha\mu}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y} \tag{4}$$

(a) Partial significance testing using the Wald test
The Wald test is used to determine whether each individual independent variable has a significant effect on the dependent variable. The hypotheses are formulated as follows:

$$H_0 : \beta_j = 0 \quad \text{(The independent variable has no significant effect on the dependent variable)}$$
$$H_1 : \beta_j \neq 0 \quad \text{(The independent variable has a significant effect on the dependent variable)}$$

The test statistic is given by:

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \tag{5}$$

where:
$\hat{\beta}_j$: estimated value of the parameter,
$SE(\hat{\beta}_j)$: standard error of $\hat{\beta}_j$.
The decision rule is:
Reject H_0 if $W_j > \chi^2_{(\alpha,1)}$ or *p-value* $< \alpha$.

(b) Simultaneous significance testing using the likelihood ratio test
This test is used to determine whether an independent variable, such as a categorical factor, should be included in the model. In addition, the Likelihood Ratio (LR) test assists in selecting the model that best fits the observed data [20]. The hypotheses are formulated as follows:

$$H_0 : \quad \beta_1 = \beta_2 = \cdots = \beta_p = 0 \quad \text{(no explanatory variables have effect)}$$
$$H_1 : \quad \exists \beta_j \neq 0,\ j = 1,\ldots,p \quad \text{(at least one variable has an effect)}$$

The test statistic is defined as:

$$LR = -2\left(L_{\text{null}} - L_{\text{full}}\right) \tag{6}$$

where:
L_{null} : the log-likelihood evaluated at the estimated parameters from the intercept-only model,
L_{full} : the log-likelihood evaluated at the estimated parameters from the full model including all independent variables.
The decision criterion is:

$$\text{Reject } H_0 \text{ if } LR > \chi^2_{(\alpha,\,df)} \text{ or if p-value} < \alpha.$$

f. Multicollinearity testing
To detect the presence of multicollinearity between independent variables, indicators such as the Pearson correlation coefficient and the Variance Inflation Factor (VIF) can be used. An explanatory variable is considered independent if its VIF value is below 10 [17]. In this study, the method used to calculate the VIF values is based on the correlation matrix, as expressed in the following formula.

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}\, s_{kk}}} = \frac{\sum_{j=1}^{n}(x_{ji}-\bar{x}_i)(x_{jk}-\bar{x}_k)}{\sqrt{\sum_{j=1}^{n}(x_{ji}-\bar{x}_i)^2 \cdot \sum_{j=1}^{n}(x_{jk}-\bar{x}_k)^2}} \tag{7}$$

Subsequently, the Variance Inflation Factor (VIF) values are obtained from the main diagonal elements of the inverse correlation matrix $\mathbf{R}^{-1}$.

g. Modeling stunting cases among children under five using the Jackknife Negative Binomial Ridge Regression (JNBR) estimator to address multicollinearity, which involves [21]:

(a) Performing centering and scaling of the independent variables using the following equation.

$$X_{ij}^{*} = \frac{x_{ij} - \bar{x}_j}{S_j\sqrt{n-1}} \tag{8}$$

where:

$$s_j = \sqrt{\frac{\sum(x_{ij} - \bar{x}_j)^2}{n-1}}$$

(b) Performing a negative binomial regression analysis using maximum likelihood estimation based on the centered and scaled variables.

(c) Constructing the matrix $X^T\hat{W}X$ based on the transformed data, where $\hat{W}$ is a diagonal matrix given by $\frac{\mu}{1+\alpha\mu}$.

(d) Forming the matrix $\mathbf{G}$, which consists of the eigenvectors of $X^T\hat{W}X$

(e) Constructing $Z = XG$.

(f) The estimator γ_{ML} is computing used $\hat{\gamma}_{ML} = G^{-1}\hat{\beta}_{ML}$

(g) Determining the ridge parameter k for the Negative Binomial model

a) Schafer et al. 1984 [22]

$$k_1 = \frac{1}{\hat{\alpha}^2_{\max}} \tag{9}$$

b) Alkhasimi et al. 2006 [23]

$$k_2 = \text{median}(s_j) \tag{10}$$

c) Kibria et al. 2003 [9]

$$k_3 = \text{median}(q_j) \tag{11}$$

$$k_4 = \max(q_j) \tag{12}$$

with:

$$\begin{aligned}
\hat{a}^2_{\max} &= \text{nilai maksimum dari } (G\beta^*_{ML})^2,\\
S_j &= \frac{t_j\sigma^2}{(n-p)\sigma^2 + t_j\hat{a}_j^2},\\
t_j &= \text{nilai eigen dari } X^TX,\\
\sigma^2 &= \frac{(\gamma-\mu)^2}{n-p-1},\\
q_j &= \frac{\lambda_{\max}}{(n-p)\sigma^2 + \lambda_{\max}\hat{a}_j^2},\\
\lambda_{\max} &= \text{nilai eigen maksimum dari } G.
\end{aligned}$$

For robustness, the selected k value was additionally evaluated based on the minimum MSE criterion across competing ridge specifications.

(h) Distributing the value of k to obtain the Jackknife Negative Binomial Ridge Regression estimator. The jackknife negative binomial ridge regression estimator can be written as follows:

$$\hat{\gamma}_{\text{JNBR}} = \left(\mathbf{I} - k^2\mathbf{B}^{-2}\right)\hat{\gamma}_{\text{ML}} \tag{13}$$

And the variance for the jackknife negative binomial ridge regression is:

$$\text{Var}(\hat{\gamma}_{\text{JNBR}}) = \left(\mathbf{I} + k^2\mathbf{B}^{-2}\right)\mathbf{\Lambda}^{-1}_{\text{NBRR}}\left(\mathbf{I} + k^2\mathbf{B}^{-2}\right) \tag{14}$$

The estimator of the jackknife negative binomial ridge regression is biased. The bias for the jackknife negative binomial ridge regression is as follows:

$$\text{Bias}(\hat{\gamma}_{\text{JNBR}}) = -k^2\mathbf{B}^{-2}\hat{\gamma}_{\text{ML}} \tag{15}$$

The MSE of the jackknife negative binomial ridge regression is:

$$\text{MSE}(\hat{\gamma}_{\text{JNBR}}) = \left(\mathbf{I} + k\mathbf{B}^{-1}\right)\mathbf{\Lambda}^{-1}_{\text{NBRR}}\left(\mathbf{I} + k\mathbf{B}^{-1}\right) + k^4\mathbf{B}^{-2}\hat{\gamma}_{\text{ML}}\hat{\gamma}^T_{\text{ML}}\mathbf{B}^{-2} \tag{16}$$

(i) The selection of the optimal parameter k is carried out by comparing the Mean Squared Error (MSE) values across different k values. The k value that results in the lowest MSE is considered optimal, as it indicates that the model with this parameter has the lowest prediction error.

(j) Transforming the model back to the original scale to obtain the final Negative Binomial regression equation.

3. RESULT AND ANALYSIS

In this study, descriptive data analysis is used to determine the mean, minimum, maximum and standard deviation values. The following table presents the descriptive statistics of the variables used in this study.

Table 2: Descriptive Statistics of Variables

Var	N	**Min**	**Max**	**Mean**	**Variance**
Y	33	73	2933	764.9393	463987.5587
X_1	33	100	29756	4709.7273	36163311.0795
X_2	33	1	285	66.6667	4395.7917
X_3	33	13	17245	2236.5758	12398632.5644
X_4	33	7	599	108.0606	14243.0587
X_5	33	312	23912	3901.0909	20050543.6477
X_6	33	331	16382	2671.6700	11034823.6042

Descriptive analysis indicates substantial variability across districts/cities, reflecting heterogeneity in maternal and child health conditions. No missing observations were identified in the dataset.

3.1. Poisson Regression Model

In this study, Poisson regression is employed as the baseline model for analyzing count data. This model is part of the Generalized Linear Model (GLM) framework, which is widely used for modeling count-type dependent variables. The use of this model aims to provide an initial representation of the relationship between independent and dependent variables, as well as to serve as a reference for assessing model adequacy prior to considering more complex modeling approaches. Results of the analysis of the poisson regression model using R:

Table 3: Poisson Regression Model Estimation

Variable	**Parameter**	**Estimation**
Constant	β_0	-2.696
X_1	β_1	-0.0002
X_2	β_2	0.0053
X_3	β_3	-0.0002
X_4	β_4	0.0022
X_5	β_5	0.00002
X_6	β_6	0.0001

The resulting Poisson Regression model is as follows:

$$\ln\left(\frac{\mu_i}{\text{balita}_i}\right) = -2.696 - 0.0002X_1 + 0.0053X_2 - 0.0002X_3 + 0.0022X_4 + 0.00002X_5 + 0.0001X_6$$

3.2. Overdispersion Test

Overdispersion is examined by comparing the ratio of the deviance to the degrees of freedom. The dispersion value is obtained as

$$\phi = \frac{9659}{26} = 371.5132.$$

In the Poisson regression framework, the equidispersion assumption requires this ratio to be approximately equal to 1. However, the obtained dispersion value is substantially greater than 1, indicating the presence of over dispersion in the data. This violation suggests that the standard Poisson regression model is not appropriate. Therefore, the Negative Binomial regression model is employed, as it incorporates an additional dispersion parameter that allows the variance to exceed the mean, making it more suitable for modeling over dispersed count data.

3.3. Testing the Estimation of the Negative Binomial Regression

Parameter estimation in the Negative Binomial regression model was conducted using the Maximum Likelihood Estimation (MLE) method. The estimated parameters obtained using R software are presented in Table 4.

Thus, the resulting Negative Binomial regression model is expressed as

$$\ln\left(\frac{\mu_i}{\text{balita}_i}\right) = -2.788 - 0.0001805X_1 + 0.006085X_2 - 0.0001903X_3 + 0.002674X_4 + 0.00003729X_5 + 0.00004954X_6$$

Table 4: Negative Binomial Regression Model Estimation

Variable	**Parameter**	**Estimation**	**Std. Error**
Constant	β_0	-2.788	0.2062
X_1	β_1	-0.0001805	0.00007370
X_2	β_2	0.006085	0.003333
X_3	β_3	-0.0001903	0.00008799
X_4	β_4	0.002674	0.001555
X_5	β_5	0.00003729	0.0001260
X_6	β_6	0.00004954	0.0001119
a		1.878	

Based on the estimated Negative Binomial regression model with an offset term log(Balita), the response is interpreted as the incidence rate of stunting per under-five population. Therefore, the regression coefficients are interpreted in terms of the incidence rate ratio (IRR $= \exp(\beta)$), which represents the multiplicative change in the stunting incidence rate for a one-unit increase in each explanatory variable, holding other variables constant.

An increase in exclusive breastfeeding (X_1) is associated with a decrease in the incidence rate of stunting by a factor of 0.9998436, corresponding to an approximate 0.018% reduction in the incidence rate of stunting. In contrast, low birth weight (X_2) is associated with an increase in the incidence rate by a factor of 1.0061035, or approximately a 0.610355% increase. Similarly, pregnant women receiving iron tablets (X_3) is associated with a slight reduction in the incidence rate of stunting by a factor of 0.9998097 (approximately 0.019028% decrease), pregnant women with chronic energy deficiency (X_4) is associated with an increase in the incidence rate by a factor of 1.0026775 (approximately 0.26775% increase). Early initiation of breastfeeding (X_5) and vitamin A supplementation (X_6) show very small positive associations with the incidence rate of stunting, with IRR values of 1.0000372 and 1.0000495, corresponding to increases of approximately 0.00372% and 0.00495%, respectively, holding all other variables constant.

3.4. Model Comparison

To evaluate the performance of the model, a comparison between the Poisson and Negative Binomial models was conducted using the Akaike Information Criterion (AIC). A lower AIC value indicates a better model fit.

Table 5: Comparison of statistical models

Model	**AIC**
Poisson	9939.1
Negative Binomial	505.04

The result indicates that the Negative Binomial model yields a lower AIC value compared to the Poisson model, suggesting that it provides a better fit to the data.

3.5. Partial Significance Testing of Model Parameters

Partial significance testing of model parameters is conducted using the Wald test to evaluate the statistical significance of each explanatory variable in explaining the response variable. The following are the results of the hypothesis test.

Table 6: Wald Test Values

Var	$\hat{\beta}_j$	**S.E**	W	***p-value***
X_1	-0.0001805	0.0000737	-2.450	0.0143
X_2	0.006085	0.003333	1.826	0.0679
X_3	-0.0001903	0.00008799	-2.163	0.0306
X_4	0.002674	0.001555	1.719	0.0855
X_5	0.00003729	0.000126	0.296	0.7672
X_6	0.00004954	0.0001119	0.443	0.6580

Using a significance level of $\alpha = 0.1$, statistical significance is determined based on the p-value criterion. The results show that X_1, X_2, X_3, and X_4 have p-values less than 0.1, indicating that these variables have a statistically significant partial effect on stunting in toddlers. Meanwhile, X_5, and X_6 are not statistically significant. These findings are consistent with previous studies [24], [25], [26], which have reported that exclusive breastfeeding (X_1) and pregnant women receiving iron tablets (X_3) significantly reduce the risk of stunting, while low birth weight (X_2) and pregnant women with chronic energy deficiency (X_4) represent a key determinant of child growth outcomes due to its adverse effect on maternal nutritional adequacy and fetal development during pregnancy.

Although early breastfeeding initiation (X_5), and vitamin A supplementation (X_6) are theoretically associated with stunting, these variables were not statistically significant in the current model. This may reflect aggregation bias at the district level, multicollinearity suppression effects, or limited statistical power due to the relatively small sample size.

3.6. Simultaneous Significance Testing of Model Parameters

The Likelihood Ratio (LR) test is commonly employed to assess and compare the fit of nested models and, in some cases, non-nested models [20].

The test statistic used:

$$LR = -2\left(L_{null} - L_{full}\right)$$

Reject H_0 if $LR > \chi^2_{(0.1;5)}$ or p-value < 0.1. The test result shows $LR = 13.754 > \chi^2_{(0.1;6)} = 10.645$, which means that there is at least one variable that has a significant effect on the model.

3.7. Multicollinearity Test

In this study, the method used to calculate the VIF values is based on the correlation matrix. Then, to obtain the VIF values, the inverse of the correlation matrix R is computed, resulting as follows:

$$\mathbf{R}^{-1} = \begin{bmatrix} 11.9142 & -1.5026 & 5.7446 & 0.0374 & -10.9291 & -2.9576 \\ -1.5026 & 2.9246 & -1.3416 & 0.3414 & 0.0750 & -0.0737 \\ 5.7446 & -1.3416 & 5.7026 & -1.2501 & -5.2018 & -2.5449 \\ 0.0374 & 0.3414 & -1.2501 & 2.0651 & -1.7632 & 1.3846 \\ -10.9291 & 0.0750 & -5.2018 & -1.7632 & 18.9804 & -3.7803 \\ -2.9576 & -0.0737 & -2.5449 & 1.3846 & -3.7803 & 8.2598 \end{bmatrix}$$

The VIF values, obtained from the main diagonal of the inverse correlation matrix $\mathbf{R}^{-1}$, indicate that two variables exceed the threshold of 10, suggesting the presence of multicollinearity. Specifically, the variables X_1 and X_5 exhibit high VIF values, indicating a strong linear relationship between these predictors. To address this issue, the Jackknife Negative Binomial Ridge Regression (JNBR) method is applied, combining ridge regression and the jackknife technique to reduce multicollinearity effects and improve the stability of parameter estimates.

3.8. Parameter Estimation of the Transformed Data Model

Although the Negative Binomial regression has demonstrated an adequate fit to the data, the multicollinearity test indicates the presence of high correlations among the independent variables. To address this issue, the Jackknife Negative Binomial Ridge Regression (JNBR) approach was employed.

Prior to addressing multicollinearity, centering and scaling were performed to standardize the variables and enhance the stability of the analysis, as defined in Equation (8).

Table 7: Data Obtained from Centering and Scaling

X_1	X_2	X_3	X_4	X_5	X_6
-0.1208	-0.1244	-0.0934	-0.0785	-0.1237	-0.1185
0.1259	0.3208	-0.0611	0.0369	0.0679	0.0632
⋮	⋮	⋮	⋮	⋮	⋮
-0.0907	-0.0844	0.0149	0.0813	-0.0673	-0.0708
-0.1129	0.0355	-0.0695	0.1006	-0.0936	-0.0898

The transformed data is attached. Next, parameter estimation for the Negative Binomial regression model was performed on the transformed data using the Maximum Likelihood estimator. The following are the estimated parameter values after transformation.

Next, a matrix is constructed based on the transformed data, given by $\mathbf{X^T\hat{W}X}$, where $\mathbf{W}$ is a diagonal matrix whose elements are defined as $\left(\frac{\mu}{1+a\mu}\right)$. Based on this formulation, the transformation matrix $\mathbf{Z} = \mathbf{XG}$ is then obtained.

Table 8: Negative Binomial Regression Model Estimation on Transformed Data

Parameter	Parameter Estimator
β_0	-3.0915
β_1	-6.2054
β_2	2.2823
β_3	-3.7904
β_4	1.8052
β_5	0.9447
β_6	0.9309
a	1.878

The parameter estimation is subsequently performed using the estimator $\hat{\beta}^*_{ML}$ as the coefficient estimator for $\mathbf{Z}$, which is further interpreted as the estimator $\hat{\gamma}_{ML}$ for the parameters associated with $\mathbf{X}$. The resulting parameter estimates are presented as follows.

Table 9: Parameter Estimates of $\hat{\gamma}_{ML}$

Parameter ($\hat{\gamma}_{ML}$)	Parameter Estimator
γ_1	1.7097
γ_2	-1.5859
γ_3	-0.7875
γ_4	-2.9210
γ_5	-4.3340
γ_6	-5.4523

3.9. Estimation of Ridge Parameter k

Based on the Jackknife Negative Binomial Ridge Regression method used in this study, the bias constant k is determined and subsequently used to construct $\mathbf{K} = k\mathbf{I}$. Equations (9), (10), (11), and (12) are employed to compute the value of k. The estimated values of the parameter k are presented in Table 10.

Table 10: Parameter Estimates of k_i

Parameter (k_i)	Parameter Estimator
k_1	0.0435
k_2	0.0124
k_3	2.6555×10^{-7}
k_4	2.6555×10^{-7}

3.10. Jackknife Negative Binomial Ridge Regression Estimator

Table 11 presents the parameter estimators, bias, variance, and MSE values obtained from the Jackknife Negative Binomial Ridge Regression method. These results are based on the estimation of k using $\hat{\gamma}_{ML}$ and are computed using Equations (13), (14), (15), and (16).

Based on Table 11, the smallest MSE value is 48.8816, corresponding to the ridge parameter $k_1 = 0.0435$. The dispersion parameter was estimated using maximum likelihood estimation through the glm.nb procedur in R. Thus, the resulting regression equation with dispersion parameter $a = 1878$ is obtained as follows.

$$\ln\left(\frac{\mu_i}{\text{balita}_i}\right) = 1.7091X^*_{i1} - 1.5757X^*_{i2} - 0.7649X^*_{i3} - 2.7616X^*_{i4} - 3.1917X^*_{i5} - 2.7167X^*_{i6}.$$

The regression model is expressed in a centering and scaling form. Therefore, the coefficients cannot yet be substantively interpreted, and a back-transformation into the original form without centering and scaling is required.

3.11. Comparison of Estimator Performance

A comparison of estimator performance was conducted to valuate the ability of the Jackknife Negative Binomial Ridge Regression (JNBR) method to address multicollinearity compared with the standard Negative Binomial Maximum Likelihood Estimation (NB-MLE) estimator. The evaluation was based on the Mean Squared Error

Table 11: Parameter Estimator of Jackknife Negative Binomial Ridge Regression

Parameter (γ_{JNBR})	k_1	k_2	k_3	k_4
γ_1	1.7091	1.7096	1.7096	1.7096
γ_2	-1.5757	-1.5850	-1.5859	-1.5859
γ_3	-0.7649	-0.7851	-0.7875	-0.7875
γ_4	-2.7616	-2.9023	-2.9210	-2.9210
γ_5	-3.1917	-4.1021	-4.3340	-4.3340
γ_6	-2.7167	-4.5392	-5.4522	-5.4522
Bias	8.8138	0.8877	1.4613×10^{-18}	1.4613×10^{-18}
Variance	40.0678	74.3495	94.1042	94.1042
MSE	48.8816	75.2372	94.1042	94.1042

(MSE) and Root Mean Squared Error (RMSE) of the parameter estimators. The comparison results are presented in the following table 12.

Table 12: Comparison of Estimator Performance

Estimator	**MSE**	**RMSE**
NB-MLE	94.1042	9.7007
JNBR (k_1)	48.8816	6.9915

The results indicate that the Jackknife Negative Binomial Ridge Regression (JNBR) estimator produced smaller MSE and RMSE values than the standard Negative Binomial Maximum Likelihood Estimator (NB-MLE). The MSE value decreased from 94.1042 for NB-MLE to 48.8816 for JNBR, while the RMSE value decreased from 9.7007 to 6.9915. This reduction suggests that the JNBR estimator was able to reduce the overall estimation error and improve the stability of the regression coefficients in the presence of multicollinearity.

The improvement occurs because the ridge component shrinks unstable coefficient estimates, thereby reducing estimator variance, while the jackknife procedure helps reduce the bias introduced by ridge regularization. Consequently, the JNBR estimator achieves a better bias–variance trade-off than the conventional NB-MLE estimator. These findings indicate that JNBR provides more efficient and stable parameter estimation for Negative Binomial regression when strong correlations exist among predictor variables.

3.12. New Negative Binomial Regression

Since the final model is expressed in standardized form, the regression coefficients were transformed back to the original scale using the following transformation:

$$\boldsymbol{\beta} = \mathbf{D}^{-1}\boldsymbol{\beta}^* \tag{17}$$

where:

$$\mathbf{D} = \operatorname{diag}\left(S_1\sqrt{n-1}, S_2\sqrt{n-1}, \ldots, S_p\sqrt{n-1}\right) \tag{18}$$

Thus, the coefficient transformation for each predictor variable is given by:

$$\beta_j = \left(\frac{1}{S_j\sqrt{n-1}}\right)\beta_j^*, \qquad j = 1, 2, \ldots, p \tag{19}$$

The matrix $\mathbf{D}$ has dimensions $p \times p$, while the parameter vector $\boldsymbol{\beta}^*$ has dimensions $p \times 1$, ensuring dimensional consistency in the back-transformation process.

Accordingly, the Negative Binomial regression model utilizing the Jackknife Negative Binomial Ridge Regression estimator can be expressed as follows:

$$\begin{aligned}\ln(\mu_i) = &-2.4452 - 1.2335 \times 10^{-4}X_1 + 5.1004 \times 10^{-3}X_2 - 1.2702 \times 10^{-4}X_3 \\ &+ 2.4175 \times 10^{-3}X_4 - 2.6640 \times 10^{-5}X_5 + 9.3187 \times 10^{-6}X_6.\end{aligned}$$

After incorporating the offset term log(number of under-five children), the estimated model represents the incidence rate of stunting per under-five population across districts/cities. Therefore, the regression parameters are interpreted in terms of the incidence rate ratio (IRR $= \exp(\beta)$), which reflects the relative change in the stunting incidence rate after controlling for the population at risk and other covariates included in the model. For improved interpretability, the effects are reported for a 10-unit increase in each explanatory variable, computed as $\exp(10\beta)$.

The estimated $\text{IRR}_1$0values indicate variation in the direction of association with the incidence rate of stunting across districts/cities in North Sumatra Province. In general, some variables yield IRR values below one, indicating a negative association with the stunting incidence rate, while others produce IRR values above one, indicating a positive association. However, the presence of extremely large IRR values for several coefficients suggests high sensitivity in the estimation process, likely driven by multicollinearity and differences in the scale of the explanatory variables. This condition implies that the quantitative interpretation based on the magnitude of IRR is not fully stable. Therefore, the results should be interpreted primarily in terms of the direction of associations and overall patterns among stunting determinants, rather than the absolute magnitude of the estimated effects.

Accordingly, policy interpretations should be made cautiously, as the model captures associative relationships rather than causal effects. Nonetheless, the results suggest that maternal and child health-related factors remain relevant in explaining variation in stunting incidence across districts/cities, particularly low birth weight and maternal chronic energy deficiency.

3.13. Study Limitations

This study has several limitations. The relatively small sample size ($n = 33$) may lead to unstable parameter estimation in the presence of multicollinearity, although the JNBR approach was applied to mitigate this issue. In addition, the use of cross-sectional data from 2023 limits the analysis to variation across districts/cities and does not allow for causal interpretation or assessment of temporal dynamics. The model incorporates a log (Balita) offset to account for differences in population size; therefore, results are interpreted in terms of incidence rates, although unobserved heterogeneity across regions may still influence the estimates. Furthermore, the limited sample size implies that the asymptotic properties underlying Maximum Likelihood Estimation and Wald-based inference should be interpreted with caution. Nevertheless, the JNBR method provides more stable estimation under conditions of overdispersion and multicollinearity.

4. CONCLUSION

The results indicate that the estimated effects of the explanatory variables on the number of stunting cases are generally small when interpreted through the exponentiated coefficients. Although the direction of associations varies across predictors, the overall magnitude of effects remains limited. From a methodological perspective, the comparison of estimator performance shows that the JNBR (k1) approach outperforms the NB-MLE estimator. Specifically, JNBR yields a lower MSE (48.8816) and RMSE (6.9915) compared to NB-MLE, which produces an MSE of 94.1042 and an RMSE of 9.7007. These results indicate that the JNBR framework provides more accurate and stable parameter estimates under conditions of overdispersion and multicollinearity. Future research is recommended to extend the analysis by incorporating longitudinal data structures, additional socioeconomic covariates, and exposure-adjusted modeling frameworks to enhance both statistical robustness and epidemiological interpretability.

REFERENCES

[1] T. A. E. Permatasari, Y. Chadirin, Ernirita, A. N. Syafitri, and D. A. Fadhilah, "The accuracy of a novel stunting risk detection application based on nutrition and sanitation indicators in children aged under five years," *BMC Nutrition*, vol. 11, no. 1, 2025, https://doi.org/10.1186/s40795-025-01074-6.

[2] S. J. W. Astuti, Suindyah, Dwiningwarni, and S. Atmojo, "Modeling environmental interactions and collaborative interventions for childhood stunting: A case from indonesia," *Dialogues in Health*, vol. 6, p. 100206, 2025, https://doi.org/10.1016/j.dialog.2025.100206.

[3] M. Ekholuenetale, A. Barrow, C. E. Ekholuenetale, and G. Tudeme, "Impact of stunting on early childhood cognitive development in benin: Evidence from demographic and health survey," *Egyptian Pediatric Association Gazette*, vol. 68, no. 1, 2020, https://doi.org/10.1186/s43054-020-00043-x.

[4] R. Regassa, T. Belachew, M. Duguma, and D. Tamiru, "Factors associated with stunting in under-five children with environmental enteropathy in slum areas of jimma town, ethiopia," *Frontiers in Nutrition*, vol. 11, p. 1335961, 2024, https://doi.org/10.3389/fnut.2024.1335961.

[5] Yuwanti, F. M. Mulyaningrum, and M. M. Susanti, "Faktor-faktor yang mempengaruhi stunting pada balita di kabupaten grobogan," *Jurnal Keperawatan dan Kesehatan Masyarakat Cendekia Utama*, vol. 10, no. 1, 2021, https://doi.org/10.31596/jcu.v10i1.704.

[6] H. Alrweili, "Kibria–lukman hybrid estimator for handling multicollinearity in poisson regression model: Method and application," *International Journal of Mathematics and Mathematical Sciences*, vol. 2024, no. 1, pp. 1–31, 2024, https://doi.org/10.1155/2024/1053397.

[7] Y. Tiara, M. N. Aidi, Erfiani, and R. Rachmawati, "Overdispersion handling in poisson regression model by applying negative binomial regression," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 1, pp. 0417–0426, 2023, https://doi.org/10.30598/barekengvol17iss1pp0417-0426.

[8] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970, https://doi.org/10.1080/00401706.1970.10488634.

[9] B. M. G. Kibria, "Performance of some new ridge regression estimators," *Communications in Statistics–Simulation and Computation*, vol. 32, no. 2, pp. 419–435, 2003, https://doi.org/10.1081/SAC-120017499.

[10] K. Liu, "A new class of biased estimate in linear regression," *Communications in Statistics–Theory and Methods*, vol. 22, no. 2, pp. 393–402, 1993, https://doi.org/10.1080/03610929308831027.

[11] M. H. Quenouille, "Notes on bias in estimation," *Biometrika*, vol. 43, no. 3/4, pp. 353–360, 1956, https://doi.org/10.2307/2332914.

[12] J. W. Tukey, "Bias and confidence in not quite large samples," *Annals of Mathematical Statistics*, vol. 29, no. 2, p. 614, 1958, https://doi.org/10.1214/aoms/1177706647.

[13] J. K. Tay, B. Narasimhan, and T. Hastie, "Elastic net regularization paths for all generalized linear models," *Journal of Statistical Software*, vol. 106, no. 1, pp. 1–31, 2023, https://doi.org/10.18637/jss.v106.i01.

[14] A. Aziz, H. Nayem, and B. M. G. Kibria, "A comparative evaluation of ols, ridge and lasso in the presence of multicollinearity and outliers," Sep. 2025, https://doi.org/10.5281/zenodo.17230601.

[15] L. Liu, J. Gao, G. Beasley, and S.-H. Jung, "Lasso and elastic net tend to over-select features," *Mathematics*, vol. 11, no. 17, p. 3738, 2023, https://doi.org/10.3390/math11173738.

[16] T. Koç and H. Koç, "A new effective jackknifing estimator in the negative binomial regression model," *Symmetry*, vol. 15, no. 12, p. 2107, 2023, https://doi.org/10.3390/sym15122107.

[17] M. A. Haris and P. R. Arum, "Negative binomial regression and generalized poisson regression models on the number of traffic accidents in central java," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 2, pp. 471–482, 2022, https://doi.org/10.30598/barekengvol16iss2pp471-482.

[18] A. Tamimi and I. Husein, "Application of negative binomial regression in analyzing factors that influence stunting in toddlers," *Desimal: Jurnal Matematika*, vol. 7, no. 1, pp. 103–112, 2024, https://doi.org/10.24042/djm.v7i1.22070.

[19] M. M. Majore, D. T. Salaki, and J. D. Prang, "Penerapan regresi binomial negatif dalam mengatasi overdispersi regresi poisson pada kasus jumlah kematian ibu," *d'CartesiaN: Jurnal Matematika dan Aplikasi*, vol. 9, no. 2, pp. 133–139, 2020, https://doi.org/10.35799/dc.9.2.2020.29150.

[20] F. Suryadi, S. Jonathan, K. Jonatan, and M. Ohyver, “Handling overdispersion in poisson regression using negative binomial regression for poverty case in west java,” *Procedia Computer Science*, vol. 216, pp. 517–523, 2023, https://doi.org/10.1016/j.procs.2022.12.164.

[21] K. A. Palinoan, A. K. Jaya, and A. Islamiyati, “Pemodelan regresi binomial negatif menggunakan estimator jackknife negative binomial ridge regression pada data angka kematian bayi provinsi sulawesi selatan,” *Basis: Jurnal Ilmiah Matematika*, vol. 3, no. 2, pp. 1–8, 2024, https://doi.org/10.30872/basis.v3i2.1140.

[22] R. L. Schaeffer, L. D. Roi, and R. A. Wolfe, “A ridge logistic estimator,” *Communications in Statistics–Theory and Methods*, vol. 13, no. 11, pp. 2005–2020, 1984, https://doi.org/10.1080/03610928408828664.

[23] M. A. Alkhamisi, G. Khalaf, and G. Shukur, “Some modifications for choosing ridge parameter,” *Communications in Statistics–Theory and Methods*, vol. 35, no. 11, pp. 2005–2020, 2006, https://doi.org/10.1080/03610920600762905.

[24] H. Hadi, Fatimatasari, W. Irwanti, C. Kusuma, R. D. Alfiana, M. I. N. Asshiddiqi, S. Nugroho, E. C. Lewis, and J. Gittelsohn, “Exclusive breastfeeding protects young children from stunting in a low-income population: A study from eastern indonesia,” *Nutrients*, vol. 13, no. 12, p. 4264, 2021, https://doi.org/10.3390/nu13124264.

[25] Y. B. Nisar, M. J. Dibley, and V. M. Aguayo, “Iron-folic acid supplementation during pregnancy reduces the risk of stunting in children less than 2 years of age: A retrospective cohort study from nepal,” *Nutrients*, vol. 8, no. 2, p. 67, 2016, https://doi.org/10.3390/nu8020067.

[26] E. K. Wati, R. Murwani, M. I. Kartasurya, and Sulistiyani, “Determinants of chronic energy deficiency incidence in pregnant women: A cross-sectional study in banyumas,” *Narra J*, vol. 4, no. 1, pp. 1–8, 2024, https://doi.org/10.52225/narra.v4i1.742.