BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2557–2570
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss4pp2557-2570

MODELING POVERTY IN WEST JAVA PROVINCE USING NEGATIVE BINOMIAL REGRESSION WITH PENALIZED SMOOTHLY CLIPPED ABSOLUTE DEVIATION

Vera Maya Santi[1*], **Aulia Baihaqi**[2], **Dania Siregar**[3]

[1,2,3]*Study Program of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta*
Jln. Rawamangun Muka, Jakarta, 13220, Indonesia

*Corresponding author's e-mail: ** **vmsanti@unj.ac.id**

Article History:
Received: 23rd January 2025
Revised: 22nd March 2025
Accepted: 27th May 2025
Available online: 1st September 2025

Keywords:
Multicollinearity;
Negative Binomial Regression;
Poverty;
SCAD;
Variables Selection.

ABSTRACT

The number of poor people is an example of discrete or count data. One commonly used regression model for count responses is the Negative Binomial regression. Regression modeling with many predictor variables results in the problem of multicollinearity. This condition causes the parameter estimator to become unstable. One method to overcome this problem is to use the penalty function to optimize the selection of predictor variables. This study aims to analyze the factors influencing the number of poor people in West Java Province using Negative Binomial regression with the Smoothly Clipped Absolute Deviation (SCAD) penalty function. The research data was sourced from the Central Bureau of Statistics in 2022, covering 27 districts/cities in West Java Province with 21 predictor variables. The method applied selects variables and estimates parameters simultaneously in the Negative Binomial regression model. Based on the AIC value, it was found that the Negative Binomial penalized SCAD model (AIC = 628.12) had better performance than the Negative Binomial regression model (AIC = 634.34). The Negative Binomial penalized SCAD regression model yielded five significant predictor variables with R^2 value of 92.8%. This model is simpler than the Negative Binomial regression model with six predictor variables. The regional minimum wage, number of cooperatives, percentage of the population who have health insurance, the pure college enrollment rate, and non-food expenditure are important variables as factors affecting the number of poor people in West Java Province.
.

How to cite this article:

V. M. Santi, A. Baihaqi, and D. Siregar, "MODELING POVERTY IN WEST JAVA PROVINCE USING NEGATIVE BINOMIAL REGRESSION WITH PENALIZED SMOOTHLY CLIPPED ABSOLUTE DEVIATION," *BAREKENG: J. Math. & App.*, vol. 19, iss. 4, pp. 2557-2570, December, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Poverty is a condition of a person's inability to meet basic daily needs, which include food and non-food needs measured in terms of expenditure [1]. The number of poor people in Indonesia fluctuates from year to year. The poor population in 2020 was around 26.42 million. This number then increased to 27.54 million people in 2021, or by approximately 1.12 million people from 2020, and in 2022 it decreased to around 26.26 million people. The largest poor population in Indonesia is on the island of Java, one of which is in West Java Province. According to Susenas data published by BPS (Central Bureau of Statistics) in 2022, the number of poor people in West Java Province has reached 4,070.98 thousand people and occupies the second position in the province with the largest poor population in Indonesia [2], [3].

Various factors can cause poverty, such as uneven economic development, low level of education, inadequate employment, and so on [4]. Previous research modelling poverty using multiple regression analysis found that levels of education, health, and asset ownership have a significant effect on poverty in North Sulawesi [5]. Another study by Munira and Juliansyah found that population density and open unemployment rates affect poverty in Aceh Province [6].

In this study, the researchers modeled data on the number of poor people in West Java Province, which is a discrete or count type. Negative Binomial regression analysis is commonly used statistical method for count data and as an alternative to Poisson regression in dealing with the problem of overdispersion in data [7]-[9]. When performing regression modeling with many predictor variables involved, there is often a problem of multicollinearity. The existence of this multicollinearity causes the parameter estimator to be unstable [10]. Therefore, the selection of independent variables is important in the analysis.

One of the variable selection methods that can be applied to Negative Binomial regression is the penalty method. Some common penalty methods used in regression models are Least Absolute Shrinkage and Selection Operator (LASSO), Elastic net, Smoothly Clipped Absolute Deviation (SCAD), and Minimum Concave Penalty (MCP) [11]-[16]. This study applies the SCAD penalty function to the Negative Binomial regression model for the case of poor population data. Previous studies by Wang et al. and Olanrewaju and Ojo have applied the SCAD penalty function to Negative Binomial regression [12], [15]. Studies conducted by Wang et al., Santi et al., and Olanrewaju and Ojo have shown that the SCAD penalty function outperforms several other regularization methods, including Lasso, Group Lasso, and MCP [12], [13], [15].

This study involves many predictor variables that were considered to have the potential to affect poverty in West Java Province. This study aims to analyze factors influencing poverty in West Java by addressing equidispersion assumption violations through Negative Binomial regression. Furthermore, the addition of the penalty function, SCAD penalty, is needed in the Negative Binomial regression model to overcome multicollinearity and select relevant predictor variables. One method for selecting optimal regularization parameters for penalty functions is Cross-Validation [17]. Model evaluation criteria such as the Akaike Information Criterion (AIC) and R-square (R^2) are used to see how the model fits into the data [18]-[20]. The findings of this study are expected to identify influential variables as a factor causing poverty in West Java Province.

2. RESEARCH METHODS

2.1 Data

This research employs secondary data from the West Java Provincial BPS (Central Bureau of Statistics) publication in 2022 (https://jabar.bps.go.id/id/publication/2023/02/28/57231a828abbfdd50a21fe31/provinsi-jawa-barat-dalam-angka-2023.html) [2] and West Java Governor Decree Number 561 of 2021 related to the regional minimum wage [21]. The observation unit in this study consists of 27 districts/cities. The response variable (Y), predictor variable (X), and offset are presented in Table 1:

Table 1. Research Variable

Symbol	Information
Y	Number of Poor People (people)
X_1	Population Density (people/km^2)
X_2	Economic Rate (%)
X_3	Workforce Placement

Symbol	Information
X_4	Regional Minimum Wage (Rupiah)
X_5	Number of Cooperatives (Unit)
X_6	Social Food Assistance Budget (Rupiah)
X_7	Open Unemployment Rate (%)
X_8	Expected Years of Schooling (Year)
X_9	Mean Years of Schooling (Year)
X_{10}	Number of Schools (Unit)
X_{11}	Life Expectancy (Year)
X_{12}	Number of Health Facilities (Unit)
X_{13}	Percentage of Health Insurance Recipients (%)
X_{14}	Percentage of People Who have Health Complaints (%)
X_{15}	Number of Undernourished Toddlers (people)
X_{16}	Number of Natural Disasters
X_{17}	Percentage of Home Ownership (%)
X_{18}	Pure High School Participation Rate (%)
X_{19}	Pure College Enrollment Rate (%)
X_{20}	Food Expenditure (Rupiah)
X_{21}	Non-food expenditure (Rupiah)
Offset	Number of Population (people)

2.2 Negative Binomial Regression

This regression model is a special case of GLM (Generalized Linear Model), which assumes the response variable follows a negative binomial distribution. The negative binomial distribution is the distribution of the Poisson-Gamma mixture [20]. Suppose Y_i follows Negative Binomial with a rate parameter μ_i and dispersion parameter $\theta(>0)$, density function of $Y_i, i = 1,2,\dots,n$ is given in **Equation (1) [19]-[22]**:

$$f(y_i;\mu_i,\theta) = \frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\theta}\right)}\left(1-\frac{1}{1+\theta\mu_i}\right)^{y_i}\left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}};\ \theta>0;\ \ y_i=0,1,2,\dots \tag{1}$$

Mean and variance value for Y_i are given by $E[Y_i]=\mu_i$ and $V[Y_i]=\mu_i+\theta\mu_i^2$ respectively. If $\theta = 0$, then $V[Y_i]=\mu_i$, so that the Negative Binomial distribution will be Poisson(μ_i). If $\theta > 0$, then the variance value of the Negative Binomial distribution will be greater than the mean. As a result, the Negative Binomial distribution permits for overdispersion in the data. If there are variable predictors $X_1, X_2, \dots, X_j$ and exposure t_i, expected value of Y_i can be expressed as shown in **Equation (2) [20]**:

$$E\left(\frac{y_i}{t_i}\right) = \frac{\mu_i}{t_i}$$

From **Equation (2)**, the Negative Binomial regression model for the expected value of Y_i per unit t_i of variable predictors with a natural logarithmic (ln) link function can be expressed as shown in **Equation (3) [8], [20]**:

$$\ln\left(\frac{\mu_i}{t_i}\right) = \boldsymbol{x}_i^T\boldsymbol{\beta}$$

$$\ln(\mu_i) = \ln(t_i) + \left(\beta_0+\beta_1x_{i1}+\dots+\beta_px_{ip}\right) \tag{3}$$

where $\ln(t_i)$ is a constant and is known as an offset.

2.3 Negative Binomial Penalized SCAD

The addition of SCAD (Smoothly Clipped Absolute Deviation) penalty to the Negative Binomial regression model was carried out to deal with the multicollinearity problem that occurs when analyzing count data. The use of the penalty method in this regression model makes it possible to estimate parameters and select important variables simultaneously. For the purpose of selecting variables, in the parameter estimation method using Maximum Likelihood is defined a penalty log-likelihood function notated with $\mathcal{L}_p(\beta;y,\theta)$ as shown in **Equation (4) [12]**.

$$\mathcal{L}_p(\beta; y, \theta) = \mathcal{L}(\beta; y, \theta) - n \sum_{j=1}^{p} p_{\lambda,\gamma}(|\beta_j|) \tag{4}$$

where $\mathcal{L}_p(\beta; y, \theta)$ is the log-likelihood function of the Negative Binomial distribution and $p_{\lambda,\gamma}(|\beta_j|)$ is the penalty function used, i.e., SCAD with tuning parameter λ. The SCAD penalty function is defined by [17] as shown in Equation (5):

$$p_{\lambda,\gamma}(|\beta_j|) = \begin{cases} \lambda|\beta_j|, & \text{if } |\beta_j| \le \lambda, \\ \dfrac{\gamma\lambda|\beta_j| - 0,5({\beta_j}^2 + \lambda^2)}{\gamma - 1}, & \text{if } \lambda < |\beta_j| \le \gamma\lambda \\ \dfrac{\lambda^2(\gamma^2 - 1)}{2(\gamma - 1)}, & \text{if } |\beta_j| > \gamma\lambda \end{cases}$$

where $\gamma > 0$ is a fixed parameter. The commonly used value of this parameter is $\gamma = 3.7$ [11].

2.4 Cross-Validation

Parameter selection of λ in SCAD can use the cross-validation method [17]. One type of cross-validation that is commonly used is k-fold cross-validation. The principle used in the k-fold cross-validation process is to divide the data into as many as k parts. The recommended k-value is five or ten [23]. Cross-validation applied to Negative Binomial penalized SCAD applies the principle in estimating its parameters based on log-likelihood values. The optimal λ parameter is selected based on the maximum log-likelihood value [12].

2.5 Model Evaluation

Model evaluation is performed to determine model performance and determine the best model among possible models that are matched to the data. This study used model evaluation criteria, namely AIC and R^2. The formulas for calculating the two criteria are presented in Equation (6) and Equation (7) [20], [22].

$$AIC = 2k - 2\ln(\mathcal{L}_k) \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{7}$$

2.6 Research Procedure

The data were processed using R software. The following are the stages of data analysis in this study:

1. Explore data to determine the characteristics of the data.
2. Goodness-of-fit of the Negative Binomial distribution on the response variable using the Chi-Square test.
3. Standardize the predictor variables using z-score transformation.
4. Check multicollinearity among predictor variables using the Variance Inflation Factor (VIF) value.
5. Specify and estimate parameters of the Negative Binomial regression model.
6. Determine the optimum λ value for the SCAD penalty regularization parameter using 10-fold cross-validation.
7. Specify and estimate parameters of the Negative Binomial Regression penalized SCAD model.
8. Evaluate the model using AIC and R^2.
9. Model interpretation.

3. RESULTS AND DISCUSSION

3.1 Data Exploration

West Java Province consists of 27 districts/cities, with details of 18 districts and 9 cities. According to information published by BPS in 2022, West Java Province ranks second with the highest number of poor people in Indonesia. Figure 1 shows a map of the distribution of the number of poor people by district/city in West Java Province for 2022.

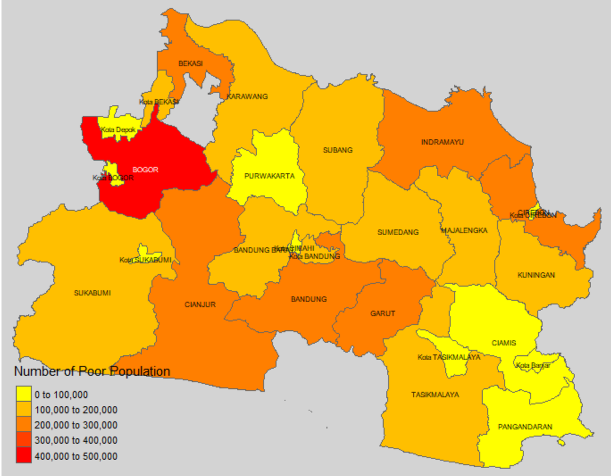

Figure 1. Distribution Map of Poor People per District/City in West Java Province

Based on **Figure 1**, it can be seen that the greater poor people in a district/city, the color on the map in the district/city will show more dark yellow to red. On the other hand, if the number of poor people in a district/city is getting less, then the color on the map in the district/city will show a lighter yellow color. Districts/cities with a relatively high number of poor people compared to other regions in West Java Province include Bogor Regency, Bandung Regency, Bekasi Regency, Cianjur Regency, and Garut Regency. These districts are also districts with a higher population than other regions. The higher the population in an area, the more likely it is that the number of poor people in that area will also be relatively higher.

The distribution of the number of poor people in West Java Province can be seen through the histogram graph along with the density plot shown in **Figure 2** below.

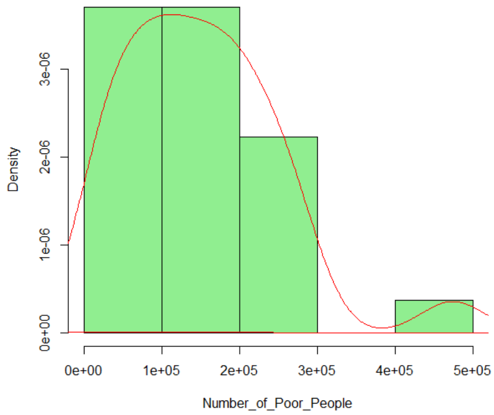

Figure 2. Histogram and Density Plot of the Number of Poor People

Based on **Figure 2**, it can be seen that the shape of the distribution of the response variable Y (number of poor people) is not in the form of a symmetrical bell curve (normally distributed), but a curve that is skewed to the right. The count data is identified as having a Negative Binomial distribution **[24]**. So the data of the response variable Y, namely the number of poor people in West Java Province, was identified as a Negative Binomial distribution.

Descriptive statistics for the response variable Y (number of poor population) and 21 predictor variables are shown in **Table 2** below.

Table 2. Descriptive Statistics of Response Variable and Predictor Variables

Variable	Mean	Standard Deviation	Range	Minimum	Maximum
Y	150,777.78	102,413.71	462	127	474.7
X_1	3,824.56	4,547.78	14,393	383	14,776
X_2	5.11	0.71	3.75	2.88	6.63
X_3	6,192.89	5,956.63	24,105	33	24,138
X_4	3,072,179.26	1,009,196.85	2,964,822	1,852,099	4,816,921
X_5	1,015.44	577.57	2,260	183	2,443
X_6	338.92	202.75	796.74	39.29	836.03

Variable	Mean	Standard Deviation	Range	Minimum	Maximum
X_7	7.80	2.33	9.22	1.56	10.78
X_8	12.84	0.78	2.50	11.78	14.28
X_9	8.78	1.45	4.64	6.83	11.47
X_{10}	744.37	483.37	1,600	99	1,699
X_{11}	72.72	1.40	5.53	69.95	75.48
X_{12}	301.26	179.81	561	41	602
X_{13}	66.76	16.20	55.67	41.21	96.88
X_{14}	33.67	11.49	44.55	15.14	59.69
X_{15}	3,669.26	2,914.22	12,064	473	12,537
X_{16}	48.96	59.69	268	4	272
X_{17}	82.34	10.01	36.55	58.98	95.53
X_{18}	61.48	8.46	28	48.44	76.44
X_{19}	18.55	8.16	29.03	7.90	36.93
X_{20}	695,376.63	128,480.95	545,846	508,099	1,053,945
X_{21}	700,648.78	326,143.8	1,223,868	335,866	1,559,734

Table 2 shows that West Java Province in 2022 has an average number of poor people of 150,777 people, with a standard deviation and range of 102,413 and 462,000 people. The very large standard deviation and range value show that the number of poor people between districts/cities in West Java Province varies widely. The highest number of poor people is in Bogor Regency with 474,700 people, while the lowest number of poor people is in Banjar City with 12,700 people.

Based on Table 2, the range value of each predictor variable is very diverse, so it is necessary to standardize these variables. The standardization process is carried out using the z-score transformation method such that the distribution of transformed data will have a mean equal to zero and a standard deviation value of one.

3.2 Goodness of Fit Test of Response Variable Distribution

Chi-square test is performed to determine whether the response variable data follows the Negative Binomial distribution or not, with the following hypothesis:

H_0 : Response variable follows the Negative Binomial distribution
H_1 : Response variable does not follow the Negative Binomial distribution
The results of the Chi-Square test are displayed in Table 3 as follows:

Table 3. Chi-Square Test of Response Variable

Chi-Square	Degree of freedom (df)	p-value
3.90	$k-(c+1)=7-3=4$	0.42

According to Table 3 above, the values of $\chi^2 = 3.90 < \chi^2_{0.05;4}(9.48)$ and $p-value > \alpha$, so there is no reason to reject H_0. It means that the response variable data, namely the number of poor people in West Java Province, comes from a Negative Binomial distribution.

3.3 Multicollinearity Test

To find out whether there is multicollinearity can refer to the Variance Inflation Factor (VIF) value. If the VIF value > 5, then there is an indication of multicollinearity between predictor variables [25]. The VIF values of each predictor variable are presented in the table below.

Table 4. VIF Value of Predictor Variable

Predictor Variable	VIF	Indication of Multicollinearity
X_1	74.21	Yes
X_2	4.14	No
X_3	11.06	Yes
X_4	35.66	Yes
X_5	7.45	Yes
X_6	58.87	Yes
X_7	10.56	Yes
X_8	20.94	Yes
X_9	73.21	Yes

Predictor Variable	VIF	Indication of Multicollinearity
X_{10}	33.00	Yes
X_{11}	7.13	Yes
X_{12}	18.68	Yes
X_{13}	13.54	Yes
X_{14}	8.62	Yes
X_{15}	18.10	Yes
X_{16}	9.76	Yes
X_{17}	17.31	Yes
X_{18}	13.65	Yes
X_{19}	18.52	Yes
X_{20}	98.38	Yes
X_{21}	178.69	Yes

According to **Table 4**, from the 21 predictor variables, there is only one variable that does not indicate multicollinearity, namely variable X_2, while the remaining variables indicate multicollinearity. It means that there is a serious multicollinearity problem in regression models. For this reason, it is necessary to handle this multicollinearity problem. One of which is by using the Smoothly Clipped Absolute Deviation penalty method for shrinking the coefficients in the Negative Binomial regression model.

3.4 Fitting the Negative Binomial Regression Model

The parameter estimation of the Negative Binomial regression model with the Maximum Likelihood Estimation (MLE) method and Fisher Scoring numerical iterations is shown in the following **Table 5**.

Table 5. Parameter Estimation of the Negative Binomial Regression Model

Parameter	Etimates	Std. Error	*p-value*
β_0	-2.514	0.022	$< 2 \times 10^{-16***}$
β_1	0.584	0.192	0.002^{**}
β_2	-0.044	0.045	0.328
β_3	0.020	0.074	0.785
β_4	0.248	0.134	0.064
β_5	-0.197	0.061	0.001^{**}
β_6	0.348	0.172	0.043^{*}
β_7	-0.032	0.073	0.657
β_8	0.073	0.102	0.476
β_9	-0.152	0.191	0.427
β_{10}	-0.317	0.128	0.014^{*}
β_{11}	0.134	0.059	0.025^{*}
β_{12}	0.027	0.096	0.775
β_{13}	0.026	0.082	0.748
β_{14}	0.003	0.066	0.952
β_{15}	0.120	0.095	0.209
β_{16}	-0.036	0.069	0.607
β_{17}	0.296	0.093	0.001^{**}
β_{18}	0.059	0.083	0.476
β_{19}	-0.137	0.096	0.154
β_{20}	-0.335	0.222	0.131
β_{21}	-0.342	0.299	0.252
AIC			634.34
Residual Deviance			27.085
Dispersion Parameter			5.417

Information : *) significant at the level of significance 5%

Some variables that have a significant effect with p-value < alpha (0.05) in the Negative Binomial regression model, including Population Density (X_1), Number of Cooperatives (X_5), Social Food Assistance Budget (X_6), Number of Schools (X_{10}), Life expectancy (X_{11}), and Percentage of Home Ownership (X_{17}).

3.5 Selection of optimum λ value

The value of λ is obtained by using 10-fold cross-validation. The result is presented as follows.

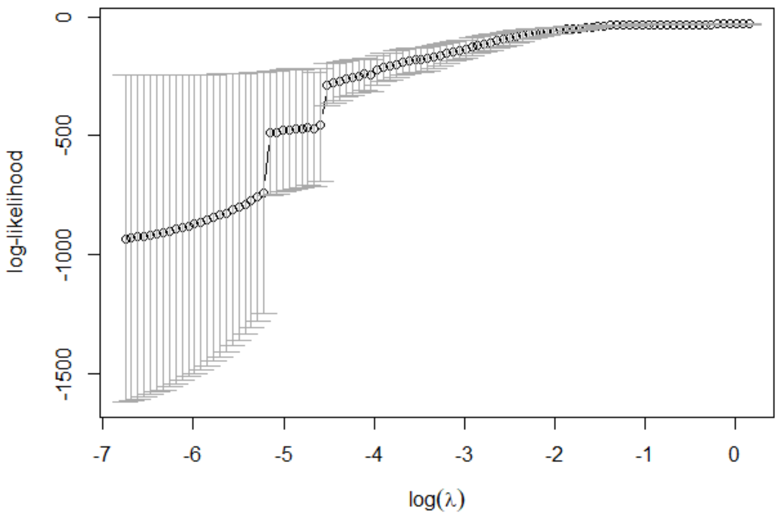

Figure 3. Log-Likelihood Value Prediction Curve for Lambda Value (λ)

In **Figure 3**, the symbol (o) shows the average log-likelihood value for a given λ value and symbol ([) indicates the log-likelihood value interval for a λ value. The optimum λ value is reached when the average log-likelihood value begins to stabilize towards a certain value, and is the maximum value. It is marked in the figure with the symbol (o), which are in a straight line with progressively smaller interval lengths and are not much different from each other.

Based on **Figure 3**, there is a recommended value of λ, i.e., the optimum lambda of 1.166867. Then the parameter estimation of the Negative Binomial penalized SCAD regression was carried out using the Maximum Likelihood Estimation method with IRLS (Iteratively Reweighted Least Square) numerical iteration algorithm and Coordinate Descent.

3.6 Fitting Negative Binomial Penalized SCAD Model

The following table is the result of parameter estimation of the Negative Binomial penalized SCAD model.

Table 6. Parameter Estimation of Negative Binomial Penalized SCAD Model

Parameter	Estimates
β_0	**-2.500**
β_1	0.000
β_2	0.000
β_3	0.000
β_4	**-0.002***
β_5	**-0.019***
β_6	0.000
β_7	0.000
β_8	0.000
β_9	0.000
β_{10}	0.000
β_{11}	0.000
β_{12}	0.000
β_{13}	**-0.024***
β_{14}	0.000
β_{15}	0.000
β_{16}	0.000
β_{17}	0.000
β_{18}	0.000
β_{19}	**-0.072***
β_{20}	0.000
β_{21}	**-0.141***
AIC	628.121
Residual Deviance	27.155
Dispersion Parameter	5.431

Information : *) significant at the level of significance 5%

Based on the results in Table 6, it can be seen that the parameter estimator value consists of zero and nonzero values. This is due to the addition of the SCAD penalty function into the Negative Binomial regression model, so that for variables that do not affect the model, the parameter estimators will be depreciated to zero, while for variables that affect the model, the parameter estimators are not zero or close to zero. This model produces dispersion parameter values that are not much different from the values generated in the Negative Binomial regression model (5.417). It can be said that both models have equally good performance in handling overdispersion in the data.

3.7 Model Evaluation

After obtaining the results of parameter estimation for the Negative Binomial regression model and the Negative Binomial penalized SCAD regression model contained in Table 5 and Table 6, the parameter estimation for the two models can be compared. It is briefly presented in the table below.

Table 7. Parameter Estimation for the Selection of Variables

Parameter	Estimate Parameter Coefficient	
	Negative Binomial	Negative Binomial penalized SCAD
Intercept	-2.514	-2.500
β_1	0.584	-
β_4	-	-0.002
β_5	-0.197	-0.019
β_6	0.348	-
β_{10}	-0.317	-
β_{11}	0.134	-
β_{13}	-	-0.024
β_{17}	0.296	-
β_{19}	-	-0.072
β_{21}	-	-0.141

Based on Table 7, the Negative Binomial model produces six variables that are statistically significant in the model, while the Negative Binomial penalized SCAD model produces five variables that affect the model. There is only one variable selected in the Negative Binomial model, which is also selected in the Negative Binomial penalized SCAD model, namely the Number of Cooperatives (X_5). The Negative Binomial model produces a greater number of variables than the Negative Binomial penalized SCAD model. It means that the Negative Binomial model is more complex than the Negative Binomial penalized SCAD model. Furthermore, to choose the best model can be identified from model performance based on several criteria, such as AIC and R^2 value. The results are shown in the table below.

Table 8. Comparison of AIC and R^2

Criterion	Negative Binomial Model	Negative Binomial Penalized SCAD
AIC	634.34	628.12
R^2	0.980	0.928

Based on Table 8, the Negative Binomial penalized SCAD model has relatively smaller AIC values compared to the Negative Binomial model. This shows that Negative Binomial penalized SCAD regression has a better performance than Negative Binomial regression in modeling the number of poor people in West Java Province with a multicollinearity problem. Then, based on the R-square value of the Negative Binomial penalized SCAD model of 0.928, which means that 92.8% of the diversity in response variables can be explained by predictor variables, and the remaining 6.2% is explained by other variables outside the model.

Based on the previous explanation, it was concluded that the best model chosen was the Negative Binomial penalized SCAD regression to model the number of poor people in West Java Province. The equation of the Negative Binomial penalized SCAD regression can be expressed in variable X with known $Z_j = \frac{X_j - \mu_j}{\sigma_j}$; $j = 1,2,\ldots,21$. The result is obtained in Equation (8),

$$\ln(\mu_i) = \ln(t_i) + \left(-2.500 - 0.002\left(\frac{X_4 - 3{,}072{,}179.26}{1{,}009{,}196.85}\right) - 0.019\left(\frac{X_5 - 1{,}015.44}{577.57}\right) - 0.024\left(\frac{X_{13} - 66.76}{16.20}\right) - 0.072\left(\frac{X_{19} - 18.55}{8.16}\right) - 0.141\left(\frac{X_{21} - 700{,}648.78}{326{,}143.8}\right)\right)$$

$$\ln(\mu_i) = \ln(t_i) + \begin{pmatrix} -3.10 - 1.98 \times 10^{-9} X_4 - 3.28 \times 10^{-5} X_5 - 1.48 \times 10^{-3} X_{13} \\ -8.82 \times 10^{-3} X_{19} - 4.32 \times 10^{-7} X_{21} \end{pmatrix} \tag{8}$$

where $\ln(t_i)$ is offset, which is the number of population in each district/city in West Java Province. Based on the model, the interpretation can be explained as follows:

1. Regional Minimum Wage (X_4)

The amount of the district/city minimum wage negatively affects the number of poor people in West Java Province. It means that every one rupiah increase in the minimum wage of districts/cities will result in a reduction in the average number of poor people in each district/city in West Java Province by $\exp(-1.98 \times 10^{-9}) = 0.99$ times, assuming the other predictor variables are constant. The findings of this study are in line with research conducted by Utami and Masjkuri, which concludes that the minimum wage has a negative influence on the number of poor people in each district/city in East Java Province [26]. Then, another study by Priseptian and Primandhana also concluded that the minimum wage has a negative influence on poverty in East Java [27]. In addition, according to Sumarsono, the determination of the minimum wage limit is expected to guarantee that the workforce meets the needs of daily life and to encourage the welfare of the workforce [28].

2. Number of Cooperatives (X_5)

The existence of cooperative units has a negative influence on the number of poor people in West Java Province. Every addition of one cooperative unit in a district/city will cause a decrease in the average number of poor people in each district/city in West Java Province by $\exp(-3.28 \times 10^{-5}) = 0.99$ times, assuming other predictor variables are constant. This research is in line with a study by Rusdianti, which declares that the existence of cooperatives through lending business capital to the community can help overcome poverty problems and improve community welfare [29]. Sharp et al. (1996) in Kuncoro identified one of the causes of poverty from the economic side as differences in access to capital [30]. The existence of a cooperative unit as a business entity providing capital can help develop the economic potential of the community in an effort to improve community welfare.

3. Percentage of Health Insurance Recipients (X_{13})

Ownership of health insurance negatively affects the number of poor people in West Java Province. This condition means that every 1% increase in health insurance recipients in each district/city in West Java Province will cause an average decrease in the number of poor people in each district/city in West Java Province by $\exp(-1.48 \times 10^{-3}) = 0.99$ times, assuming the other predictors remain the same. This study is in line with research by Dartanto, which explains that the provision of national health insurance has been able to protect and save up to more than one million people from poverty in 2015-2016 [31]. Johnston stated that providing health insurance to the poor can help reduce health care costs, so that the treatment process becomes easier to do [32]. Health insurance plays a role in reducing the burden of health expenditure in a community. such as for the cost of health services and medicines.

4. Pure College Enrollment Rate (X_{19})

According to the results of the study, a person's participation in studying at the university level negatively affects the number of poor people in West Java Province. Every 1% increase in the pure enrollment rate of universities in each district/city in West Java Province will result in a reduction in the average number of poor people in each district/city in West Java Province by $\exp(-8.82 \times 10^{-3}) = 0.99$ times assuming the other predictor variables are constant. This result is in line with a study by Nurhasanah et al., which concluded that the education participation rate negatively affects the poverty rate in Jambi Province [33]. Furthermore, this research is supported by Dejanvry and Sadoulet in [34] which states that education can reduce inequality and poverty. The higher a person's level of education, the better the level of knowledge and skills possessed. This allows the person to get an adequate job and avoid poverty.

5. Non-Food Expenditure (X_{21})

The results of this study show that the large amount of family expenditure for non-food needs has a negative influence on the number of poor people in West Java Province. Every one rupiah increase in expenditure on non-food needs in each district/city in West Java Province will cause a decrease in the average number of poor people in each district/city in West Java Province by $\exp(-4.32 \times 10^{-7}) = 0.99$ times assuming the other predictor variables are constant. The results are in line with research by Massaid, which states that an increase in non-food expenditure significantly affects poverty reduction in Indonesia [35]. Then, a similar study was also conducted by Hasanah, which explains that per capita expenditure negatively affects the poverty rate in districts/cities in Jambi Province [36]. The higher per capita expenditure, especially expenditure on non-food, means a better economic condition of the community in meeting their needs, resulting in a decrease in the number of poor people in the community.

4. CONCLUSION

Analyzing count data as a response variable influenced by many predictor variables is a challenge because it requires consideration of assumptions in regression modeling. In this study, we use a Negative Binomial penalized SCAD regression analysis approach to deal with the problem of unmet assumptions, namely equidispersion and multicollinearity. This modeling is applied to data on the number of poor people in West Java Province with several variables that have a high potential to influence it.

The Negative Binomial penalized SCAD regression model yields five predictor modifiers with values AIC = 628.12. This model performs better and forms a simpler model than the Negative Binomial regression model with six predictor variables and values AIC = 634.34. Based on the R-square value of the Negative Binomial penalized SCAD model, 92.8% of the diversity in response variables can be explained by predictor variables. Some of the predictor variables selected in the model include the regional minimum wage, the number of cooperatives, the percentage of the population who have health insurance, the pure college enrollment rate, and non-food expenditure. The application of Negative Binomial regression in this study has not fully resolved the overdispersion issue in the data. Therefore, alternative approaches such as Generalized Poisson, Quasi-Poisson, Zero-Inflated Poisson, and others are recommended for further consideration.

AUTHOR CONTRIBUTIONS

Vera Maya Santi: Conceptualization, Formal Analysis, Methodology, Writing - Original Draft, Writing - Review and Editing, Project Administration, Supervision, Validation, Software. Aulia Baihaqi: Conceptualization, Data Curation, Formal Analysis, Validation, Software, Writing - Original Draft. Dania Siregar: Formal Analysis, Validation, Writing - Review and Editing. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This research was funded by a competitive research grant under the national collaboration scheme, provided by the Institute for Research and Community Service (LPPM), Universitas Negeri Jakarta.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the editorial and reviewer team of the *Barekeng* Journal for their valuable support in enhancing the quality of this manuscript. We also extend our thanks to the West Java Provincial BPS (Central Bureau of Statistics) for providing the data that supported this research. Our appreciation goes to the Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, for providing research infrastructure, and to the Institute for Research and Community Service (LPPM UNJ) for supporting and funding this study through its competitive research grant program.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest related to this article.

REFERENCES

[1] M. A. M. Wibisono and T. Sirait, “MULTIDIMENSIONAL POVERTY MODELING IN CENTRAL JAVA, DI YOGYAKARTA, AND EAST JAVA PROVINCES,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, pp. 1939–1954, Jul. 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp1939-1954.

[2] Badan Pusat Statistik, “PROVINSI JAWA BARAT DALAM ANGKA,” Jawa Barat, 2023.

[3] [BPS] Badan Pusat Statistik, “DATA DAN INFORMASI KEMISKINAN KABUPATEN/KOTA,” Jakarta, 2022.

[4] M. Kasim, *KARAKTERISTIK KEMISKINAN DI INDONESIA DAN STRATEGI PENANGGULANGANNYA: STUDI KASUS DI PADANG PARIAMAN*. Jakarta: Indomedia, 2006.

[5] E. H. Jacobus, P. Kindangen, and E. N. Walewangko, “ANALYSIS OF FACTORS AFFECTING HOUSEHOLD POVERTY IN NORTH SULAWESI (ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEMISKINAN RUMAH TANGGA DI SULAWESI UTARA),” *Jurnal Pembangunan Ekonomi Dan Keuangan Daerah*, vol. 19, no. 3, 2018.doi: https://doi.org/10.35794/jpekd.19900.19.7.2018

[6] Munira and H. Juliansyah, “ANALYSIS OF THE INFLUENCE OF POPULATION DENSITY, PER CAPITA EXPENDITURE, AND OPEN UNEMPLOYMENT RATE ON THE POOR POPULATION IN ACEH PROVINCE FROM 2000 TO 2019 (ANALISIS PENGARUH KEPADATAN PENDUDUK, PENGELUARAN PERKAPITA, DAN TINGKAT PENGANGGURAN TERBUKA TERHADAP PENDUDUK MISKIN DI PROVINSI ACEH TAHUN 2000-2019),” *Jurnal Ekonomi Regional Unimal*, vol. 5, no. 1, 2022.doi; https://doi.org/10.29103/jeru.v5i1.7916

[7] D. Handayani, A. F. Artari, W. Safitri, W. Rahayu, and V. M. Santi, “COUNT REGRESSION MODELS FOR ANALYZING CRIME RATES IN THE EAST JAVA PROVINCE,” *J Phys Conf Ser*, vol. 2123, no. 1, p. 012028, Nov. 2021, doi: https://doi.org/10.1088/1742-6596/2123/1/012028.

[8] A. Agresti, *AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS*, Third. Florida: Wiley Series in Probability and Statistics, 2019.

[9] Y. Tiara, M. N. Aidi, E. 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, Apr. 2023, doi: https://doi.org/10.30598/barekengvol17iss1pp0417-0426.

[10] Z. Y. Algamal, R. E. Shamany, and N. N. Alobaidi, “A NEW RIDGE ESTIMATOR FOR THE NEGATIVE BINOMIAL REGRESSION MODEL,” *Thailand Statistician*, vol. 19, no. 1, pp. 115–124, 2021.

[11] H. Wang, R. Li, and C.-L. Tsai, “TUNING PARAMETER SELECTORS FOR THE SMOOTHLY CLIPPED ABSOLUTE DEVIATION METHOD,” *Biometrika*, vol. 94, no. 3, pp. 553–568, Aug. 2007, doi: https://doi.org/10.1093/biomet/asm053.

[12] Z. Wang, S. Ma, M. Zappitelli, C. Parikh, C.-Y. Wang, and P. Devarajan, “PENALIZED COUNT DATA REGRESSION WITH APPLICATION TO HOSPITAL STAY AFTER PEDIATRIC CARDIAC SURGERY,” *Stat Methods Med Res*, vol. 25, no. 6, pp. 2685–2703, Dec. 2016, doi: https://doi.org/10.1177/0962280214530608.

[13] V. M. Santi, K. A. Notodiputro, and B. Sartono, “VARIABLE SELECTION METHODS APPLIED TO THE MATHEMATICS SCORES OF INDONESIAN STUDENTS BASED ON CONVEX PENALIZED LIKELIHOOD,” *J Phys Conf Ser*, vol. 1402, no. 7, p. 077096, Dec. 2019, doi: ttps://doi.org/10.1088/1742-6596/1402/7/077096.

[14] V. M. Santi, K. A. Notodiputro, B. Sartono, and W. Rahyu, “GENERALIZED LINEAR MIXED MODELS BY PENALIZED LASSO IN MODELLING THE SCORES OF INDONESIAN STUDENTS,” *J Phys Conf Ser*, vol. 1869, no. 1, p. 012140, Apr. 2021, doi: https://doi.org/10.1088/1742-6596/1869/1/012140.

[15] R. O. Olanrewaju and J. F. Ojo, “NON-CONVEX PENALIZED ESTIMATION OF COUNT DATA RESPONS VIA GENERALIZED LINEAR MODEL,” *Asian Journal of Fuzzy and Applied Mathematics*, vol. 8, no. 3, 2020.doi: https://doi.org/10.24203/ajfam.v8i3.6443

[16] M. Arayeshgari, L. Tapak, G. Roshanaei, J. Poorolajal, and A. Ghaleiha, “APPLICATION OF GROUP SMOOTHLY CLIPPED ABSOLUTE DEVIATION METHOD IN IDENTIFYING CORRELATES OF PSYCHIATRIC DISTRESS AMONG COLLEGE STUDENTS,” *BMC Psychiatry*, vol. 20, no. 1, p. 198, 2020, doi: https://doi.org/10.1186/s12888-020-02591-3.

[17] J. Fan and R. Li, “VARIABLE SELECTION VIA NONCONCAVE PENALIZED LIKELIHOOD AND ITS ORACLE PROPERTIES,” *J Am Stat Assoc*, vol. 96, no. 456, 2001.doi: https://doi.org/10.1198/016214501753382273

[18] A. J. Dobson and A. G. Barnett, *AN INTRODUCTION GENERALIZED LINEAR MODELS*, 4th ed. US: Taylor & Francis Group, 2018.

[19] A. Agresti, *CATEGORICAL DATA ANALYSIS*, 2nd ed. Hoboken: John Wiley & Sons, 2002.

[20] J. M. Hilbe, *NEGATIVE BINOMIAL REGRESSION*, 2nd ed. New York: Cambridge University Press, 2011.

[21] Pemerintah Jawa Barat, *KEPUTUSAN GUBERNUR JAWA BARAT*. Jawa Barat, 2021.

[22] A. C. Cameron and P. K. Trivedi, *REGRESSION ANALYSIS OF COUNT DATA*. New York: Cambridge University Press, 1998.

[23] A. J. Izenman, *MODERN MULTIVARIATE STATISTICAL TECHNIQUES: REGRESSION, CLASSIFICATION, AND MANIFOLD LEARNING*. New York: Springer, 2008.

[24] A. K. Yadav and S. K. Shah, “NEGATIVE BINOMIAL DISTRIBUTION TO EXPLAIN THE DOMESTIC FIRE INCIDENCE IN NEPAL,” *Nepalese Journal of Statistics*, pp. 51–66, Dec. 2021, doi: https://doi.org/10.3126/njs.v5i1.41229.

[25] J. F. Hair, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, *PARTIAL LEAST SQUARES STRUCTURAL EQUATION MODELING (PLS-SEM) USING R*. Cham: Springer International Publishing, 2021. doi: https://doi.org/10.1007/978-3-030-80519-7.

[26] H. W. Utami and S. U. Masjkuri, “THE INFLUENCE OF ECONOMIC GROWTH, MINIMUM WAGE, OPEN UNEMPLOYMENT RATE, AND EDUCATION ON THE NUMBER OF POOR POPULATION (PENGARUH PERTUMBUHAN EKONOMI, UPAH MINIMUM, TINGKAT PENGANGGURAN TERBUKA DAN PENDIDIKAN TERHADAP JUMLAH PENDUDUK MISKIN),” *Jurnal Ekonomi dan Bisnis Airlangga*, vol. 28, no. 2, pp. 105–116, 2018.doi: https://doi.org/10.20473/jeba.V28I22018.105-116

[27] L. Priseptian and W. P. Primandhana, “ANALYSIS OF FACTORS AFFECTING POVERTY IN EAST JAVA (ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEMISKINAN DI JAWA TIMUR),” *Forum Ekonomi UPN* , vol. 24, no. 1, pp. 45–53, 2022.doi: https://doi.org/10.30872/jfor.v24i1.10362

[28] S. Sumarsono, *EKONOMI MANAJEMEN SUMBER DAYA MANUSIA DAN KETENAGAKERJAAN*. Jember: Penerbit Graha Ilmu, 2003.

[29] E. Rusdianti, "POVERTY ALLEVIATION STRATEGIES THROUGH THE ESTABLISHMENT AND DEVELOPMENT OF COOPERATIVES (STRATEGI PENGENTASAN KEMISKINAN MELALUI PROGRAM PENDIRIAN DAN PENGEMBANGAN KOPERASI)," *Jurnal Dinamika Sosial Budaya*, vol. 21, no. 2, 2019.doi: https://doi.org/10.26623/jdsb.v21i2.1765

[30] M. Kuncoro, *DASAR-DASAR EKONOMIKA PEMBANGUNAN*. Yogyakarta: UPP STIM YKPN, 2010.

[31] T. Dartanto *et al.*, "DAMPAK PROGRAM JKN-KIS TERHADAP KEMISKINAN," Jakarta, 2017.

[32] E. M. Johnston, S. McMorrow, T. W. Thomas, and G. M. Kenny, "ACA MEDICAID EXPANSION AND INSURANCE COVERAGE AMONG NEW MOTHERS LIVING IN POVERTY," *Pediatrics*, vol. 145, no. 5, 2020.doi: https://doi.org/10.1542/peds.2019-3178

[33] Nurhasanah, M. Syafri, and J. K. Edi, "ANALYSIS OF THE INFLUENCE OF EDUCATION LEVEL AND ECONOMIC GROWTH ON POVERTY RATE IN JAMBI PROVINCE (ANALISIS PENGARUH TINGKAT PENDIDIKAN DAN PERTUMBUHAN EKONOMI TERHADAP TINGKAT KEMISKINAN DI PROVINSI JAMBI)," *Jurnal Ekonomi Sumberdaya dan Lingkungan*, vol. 8, no. 3, 2019.doi: https://doi.org/10.22437/jels.v8i3.11993

[34] K. P. Doshi, *INEQUALITY AND ECONOMIC GROWTH*. California: University of San Diego, 2000.

[35] A. Massaid, M. Hanif, D. Febrianti, and N. Chamidah, "MODELLING OF POVERTY PERCENTAGE OF NON-FOOD PER CAPITA EXPENDITURES IN INDONESIA USING LEAST SQUARE SPLINE ESTIMATOR," *IOP Conf Ser Mater Sci Eng*, vol. 546, no. 5, p. 052044, Jun. 2019, doi: https://doi.org/10.1088/1757-899X/546/5/052044.

[36] R. Hasanah, Syaparuddin, and Rosmeli, "THE INFLUENCE OF LIFE EXPECTANCY, AVERAGE YEARS OF SCHOOLING, AND PER CAPITA EXPENDITURE ON POVERTY RATES IN REGENCIES/CITIES OF JAMBI PROVINCE (PENGARUH ANGKA HARAPAN HIDUP, RATA-RATA LAMA SEKOLAH, DAN PENGELUARAN PERKAPITA TERHADAP TINGKAT KEMISKINAN PADA KABUPATEN/KOTA DI PROVINSI JAMBI)," *Jurnal Perspektif Ekonomi dan Pembangunan Daerah*, vol. 10, no. 3, 2021.doi: https://doi.org/10.22437/pdpd.v10i3.16253