JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME 9 (ISSUE 2) DECEMBER 2025: 14-24
https://doi.org/10.21009/JSA.092XX

Linear Mixed Model-Lasso with MLE and REML Estimation on Poverty Data in Java Island

Vera Maya Santi[1]*, Devi Indriana[1], Faroh Ladayya[1], Tonah[2]

[1]*Study Program of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta Jl. Rawamangun Muka, East Jakarta 13320, Indonesia*

[2]*Study Program of Mathematics Education, Faculty of Education and Science, Universitas Swadana Gunung Jati Jl. Perjuang, Cirebon, West 45132, Indonesia*

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

Article History:
Received: 16, November 2025
Revised: 11, December 2025
Accepted: 29, December 2025
Published: 31, December 2025
Available online.

Keywords:
Poverty; Linear Mixed Model; LASSO; MLE; REML

ABSTRACT

Poverty in Indonesia, especially in Java, remains a major challenge despite the island being the economic and political centre of the country. The government has made many efforts but has not been effective in overcoming poverty. The hierarchical structure of poverty data may cause higher-level clusters to be random effect. One approach that can be used to represent the relationship between the poverty rate in each regency/city in Java and the factors that influence it with the province as a random effect is a linear mixed model (LMM). The number of factors that can affect poverty results in multicollinearity. The application of LASSO is used in this study to overcome multicollinearity, select, and generate variables that are significant to poverty in Java. The data used in this study consists of 85 regencies and 34 cities in Java Island involving 20 independent variables. The results show that the factors that influence the poverty rate are average years of schooling, non-food expenditure, number of households with housing assets owned, percentage of households with a dirt floor, and percentage of households with PLN lighting. The LMM-LASSO is a linear model augmented with a LASSO penalty function to address multicollinearity and incorporates random effects into the model. This approach is suitable for modeling the poverty rate, as indicated by its smaller AIC and BIC values compared to the conventional linear mixed model. In addition, based on the ICC value, the province as a random effect contributes significantly to the variability of the data at the district/city observation level in Java Island.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:
V. M. Santi, D. Indriana, F. Ladayya, Tonah, "Linear Mixed Model-Lasso with MLE and REML Estimation on Poverty Data in Java Island", Jurnal Statistika dan Aplikasinya, vol. 9, iss. 2, pp. 14 – 24, December 2025

Copyright © 2025 Author(s)
Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

Every country has a development plan to improve people's welfare by reducing poverty [1]. Poverty in Indonesia, especially in Java, is a major challenge even though this region is the economic and political centre of the country. The factors that cause poverty in Java are very diverse, not only from the economic aspect but also concerning aspects of health, education, and living standards. Although the government has implemented various poverty alleviation programs, the results have not been effective in reducing the high poverty rate [2]. [3] applied panel data regression methods to analyze poverty in the provinces of Java and Bali. Furthermore, [4] applied multiple linear regression analysis to study poverty in East Java Province.

Many previous studies have used linear regression to analyse poverty. A linear regression model that assumes all explanatory variables have fixed effect is called a fixed effects model. However, in certain situations, the explanatory variables can be considered as samples from the population, so it can be modelled as random effect, known as random effects models. A model that contains both fixed effect and random effect is called a linear mixed model (LMM) [5], [6]. The random effect in LMM is used to accommodate the possibility of correlation between observed objects with each other in the same group (cluster) [5], [6]. For example, a random effect is hierarchical data of several individuals within a group of provinces, which can have a random effect if the individuals are assumed to be correlated [7].

The large number of explanatory variables used in LMM can cause multicollinearity [8]. To overcome this problem, variable selection using penalties, such as LASSO (Least Absolute Shrinkage and Selection Operator), can be applied. LASSO is a method that minimizes the sum of squared errors by considering the shrinkage of regression coefficients towards zero, which can reduce insignificant coefficients [9], [10]. LASSO has proven to be effective in improving the stability and efficiency of estimation in linear regression models and linear mixed models [8], [11].

Previous research by Santi et al. (2021) showed that GLMM-LASSO performed better than GLMM in assessing factors affecting Indonesian students' science literacy scores on PISA 2015 data based on AIC, BIC, and standard error [11]. Dwinata et al. (2021) also found that GLMM-LASSO was effective in variable selection and analysis of Covid-19 patients in the ICU [8].

2. METHODS

Material and Data

The data used in this study are secondary data obtained from publication books and websites of Badan Pusat Statistik (BPS) of Banten, DKI Jakarta, West Java, Central Java, D.I Yogyakarta, and East Java in 2022. The research unit used is 119 regencies/cities in Java Island consisting of 85 regencies and 34 cities.

Table 1 Research Variables

Notation	Variables	Scale
Y	Percentage of poverty	Interval
X_1	Life Expectancy at Birth (year)	Ratio
X_2	Number of Health Centers	Ratio
X_3	Number of Family Planning Facilities	Ratio
X_4	Number of Health Insurance Recipients	Ratio
X_5	Average Years of Schooling (year)	Ratio
X_6	Primary School Net Enrollment Rate (%)	Interval
X_7	Junior High School Net Enrollment Rate (%)	Interval
X_8	Senior High School Net Enrollment Rate (%)	Interval
X_9	Number of School	Ratio
X_{10}	District/City Minimum Wage (IDR)	Interval
X_{11}	Realization of Food Aid Budget (IDR)	Interval
X_{12}	Percentage of PDRB	Interval
X_{13}	Food Expenditure (IDR)	Ratio

Notation	Variables	Scale
X_{14}	Non-Food Expenditure (IDR)	Interval
X_{15}	Number of Households with Housing Assets Owned	Ratio
X_{16}	Percentage of Households with Proper Sanitation	Interval
X_{17}	Percentage of Households with Access to Adequate Drinking Water	Interval
X_{18}	Percentage of Gas Fuel Usage for Cooking	Interval
X_{19}	Percentage of Households with a Dirt Floor	Interval
X_{20}	Percentage of Households with PLN Lighting	Interval

Research Method

The analytical approach used in this study is a linear mixed model by penelized LASSO (LMM-LASSO). The development of this linear model involves fixed effects and random effects to measure the relationship between response variables and explanatory variables [4], [5]. LASSO is used to perform variable selection by shrinking the coefficients of certain predictors to zero in overcoming multicollinearity problems. The procedural steps of the analysis carried out in this study are as follows:

1. Collecting hierarchical data, regency/city data from each province in Java Island.
2. Explore the data by presenting descriptive statistics on the response variable to determine the general characteristics of the data.
3. Standardize the observation data on the explanatory variables using z-score transformation based on the following formula:

$$Z^{score} = (x - \mu)/\sigma \quad (1)$$

with x = explanatory variable; μ = mean of the explanatory variable; σ = standard deviation of the explanatory variable

4. Checking multicollinearity between explanatory variables with the Variance Inflation Factor (VIF) value. The VIF value can be calculated with the following equation:

$$\text{VIF}_p = \frac{1}{1 - \text{R}_p^2}, \text{j} = 1,2, \dots, p \quad (2)$$

5. Aplication of Linear Mixed Model-LASSO.

Linear mixed model-LASSO (LMM-LASSO) is a statistical analysis method used to overcome multicollinearity problems and reduce model complexity in hierarchical data analysis [6]. The model is as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta_l} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{\varepsilon} \quad (3)$$

where $\boldsymbol{y}(n \times 1)$ is the vector of poverty rate percentage, $\mathbf{X}(n \times (p+1))$ is the fixed effect design matrix, $\mathbf{Z}\ (n \times q)$ is random efect design matrix, $\boldsymbol{\beta_l}((p\ +\ 1)\ \times\ 1)$ is the vector of fixed effect parameters, $\boldsymbol{u}(q \times 1)$ is the vector of random effect parameters and $\boldsymbol{\varepsilon}(n \times 1)$ is the vector of errors.

5.1 Determining the optimum λ using 10-fold cross validation.

Cross-validation is one of the simplest methods most often used to estimate the prediction error of a model in the LASSO method. Cross-validation is usually used to determine the optimal value of λ. The cross-validation prediction error estimate is written as [12]:

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}^{-k(i)}(x_i)) \quad (4)$$

where L is a loss function that measures the difference between the actual value y_i and predicted value of $\hat{f}$ model.

5.2 Estimation Parameters

1. Estimation of Fixed Effect Parameters ($\boldsymbol{\beta}_l$)

For the purpose of variable selection, the parameter estimation method $\boldsymbol{\beta}_l$ using maximum likelihood is defined by the penelized LASSO log-likelihood function denotated by $\text{l}_\text{p}(\boldsymbol{\beta_l}; \boldsymbol{\theta})$ as follows:

$$l_p(\boldsymbol{\beta_l};\boldsymbol{\theta}) = l(\boldsymbol{\beta},\boldsymbol{\theta}) + \lambda \sum_{i=1}^{j} \|\boldsymbol{\beta}_j\| \quad (5)$$

$l(\boldsymbol{\beta};\boldsymbol{\theta})$ is *log-likelihood* function from linear mixed model and $\lambda \sum_{i=1}^{j}(\|\boldsymbol{\beta}_j\|)$ is the function of penelized LASSO fungsi penalti LASSO with tuning parameter λ. Then, the parameter $\boldsymbol{\beta}_l$ of the LMM-LASSO can be estimated by $\widehat{\boldsymbol{\beta}}^{lasso}$ [13]:

$$\widehat{\boldsymbol{\beta}}^{lasso} = argmin\{l_p(\boldsymbol{\beta},\boldsymbol{\theta})\}$$

$$\widehat{\boldsymbol{\beta}}^{lasso} = argmin\left\{\boldsymbol{l}(\boldsymbol{\beta},\boldsymbol{\theta}) + \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_j|\right\} \quad (6)$$

2. Estimation of Variance ($\boldsymbol{\theta}$)

Restricted Maximum Likelihood Estimation (REML) is an alternative method to estimate the parameter $\boldsymbol{\theta}$, where the data is transformed to remove the influence of fixed effects [5]. The complexity of estimating the parameter $\boldsymbol{\theta}$ using the REML can be solved using a numerical iteration algorithm, Newton Raphson [14], [15]. To estimate the variance $\boldsymbol{\sigma}$, where $\mathrm{V} = \sum_{i=0}^{q} Z_i Z_i \boldsymbol{\sigma}_i$ can be estimated by the following equation [5]:

$$\boldsymbol{\sigma}^{(t+1)} = \boldsymbol{\sigma}^{t} - \alpha[\boldsymbol{l_\sigma}''(\boldsymbol{\theta})]^{-1}\boldsymbol{l_\sigma}'(\boldsymbol{\theta}) \quad (7)$$

where

$\boldsymbol{l_\sigma}'(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\sigma_i}} = \{c\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{y}'\mathbf{P}\mathbf{Z_i}\mathbf{Z_i'}\mathbf{P}\boldsymbol{y} - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{tr}(\mathbf{V^{-1}}\boldsymbol{Z_i}\boldsymbol{Z_i'})\}$ and

$l_\sigma''(\theta) = \dfrac{\partial^2 l}{\partial \boldsymbol{\sigma_i}\partial \boldsymbol{\sigma_j}'} = \{\dfrac{\mathbf{1}}{\mathbf{2}}\boldsymbol{y}'\mathbf{P}\mathbf{Z_i}\mathbf{Z_i'}\mathbf{P}\mathbf{Z_j}\mathbf{Z_j'}\mathbf{P}\boldsymbol{y} + \dfrac{\mathbf{1}}{\mathbf{2}}\boldsymbol{y}'\mathbf{P}\mathbf{Z_j}\mathbf{Z_j}'\mathbf{P}\mathbf{Z_i}\mathbf{Z_i}'\mathbf{P}\boldsymbol{y} - \dfrac{\mathbf{1}}{\mathbf{2}}\mathbf{tr}(\mathbf{V^{-1}}\boldsymbol{Z_j}\boldsymbol{Z_j'}\mathbf{V^{-1}}\boldsymbol{Z_i}\boldsymbol{Z_i'})\}_{i,j=0}^{q}$

5.3 Random Effect Prediction

Linear mixed model assumes that the random effect $\boldsymbol{u}$ follows a normal distribution with mean $\mathbf{0}$ and variance-covariance $\mathbf{D}$. Based on the observed data, the conditional expected value of the random effect is as follows [16]:

$$\widehat{\boldsymbol{u}}_{\boldsymbol{i}}(\boldsymbol{\beta},\boldsymbol{\theta}) = \widehat{\boldsymbol{u}}_{\boldsymbol{i}} = \boldsymbol{E}[\boldsymbol{u}_i|\boldsymbol{y}_i] = \widehat{\boldsymbol{D}}\boldsymbol{Z_i'}\widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{-\mathbf{1}}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) \quad (8)$$

This conditional expectation is known as the Best Linear Unbiased Predictor (BLUP) of the random effect. BLUP is unbiased and has the smallest variance compared to all other unbiased estimators. The predictor $\widehat{\boldsymbol{u}}_{\boldsymbol{i}}$ is also often reffered to *Empirical BLUP* (EBLUP) because it is calculated based on the estimation of the parameters $\boldsymbol{\beta}$ dan $\boldsymbol{\theta}$.

6. Significance Test of Parameters

Significance test of parameters is conducted after obtaining the regression model estimation. The purpose of this test is to identify independent variables that have a significant influence on the model.

6.1 Likelihood Ratio Test (LRT).

Simulatneous test of the fixed effect parameters can be conducted with *Likelihood Ratio Test* (LRT). LRT is a test based on the value of the likelihood function for two models that define the hypothesis being tested. The LRT statistic can be seen in the following equation [16]:

$$-2\ln(\frac{L_0}{L_1}) = -2\ln(L_0) - (-2\ln(L_1)) \sim \boldsymbol{\chi}^2_{df} \quad (9)$$

6.2 T-test

Partial test of fixed effect parameters can be conducted with t-test. The t-statistics can be seen in the following equation [16]:

$$t = \frac{\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}}{se(\widehat{\boldsymbol{\beta}}_j)} \quad (10)$$

7. Diagnostics Test

7.1 Normality Test

The normality test is used to evaluate whether the residual values of the model follow a normal distribution. The assumption of residual normality is checked by looking at the Normal Q-Q Plot and the Kolmogorov-Smirnov test.

7.2 Homoscedasticity Test

Homogeneous variation of the residuals is also known as the homoscedasticity assumption. If this assumption is violated, the residuals are heteroscedastic, which can lead to biased model estimates. Heteroscedasticity tests are examined with scatter plots and the Breusch Pagan test.

8. Goodness of Fit

Goodness of fit test is conducted to determine the performance of the estimated model. The criteria that are often used to see the feasibility of the model are through the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Root Mean Square Error (RMSE) values. In addition, to see the diversity in the LMM explained by random effects can be seen from the Intraclass Correlation Coefficients (ICC).

3. RESULTS

Desctiptive Statistics

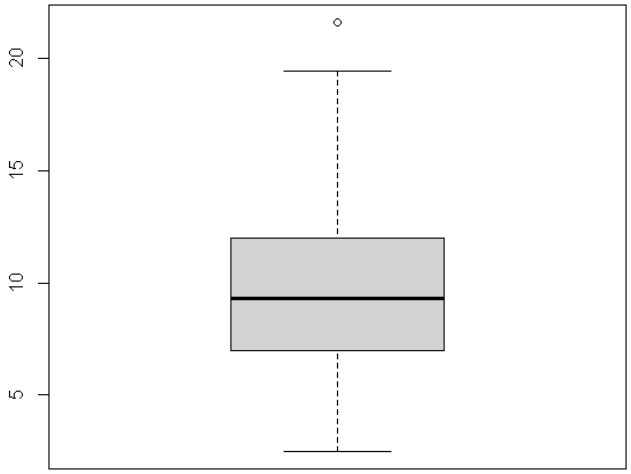


Figure 1. Boxplot of Poverty Rate Percentage

Based on the boxplot in Figure 1, it can be seen that the data has a symmetric distribution. Theresponse variable in this study, the poverty percentage, is assumed to be nomally distributed. This can also be checked by looking at the Normal Q-Q Plot (quantile-quantile plot) in Figure 2. It can be seen that the points spread along a straight line, indicating that the response variable is normally distributed.

The average poverty percentage in Java Island is 9.59%. The standard deviation of 3.81% shows the variation in the poverty rate between districts/municipalities in Java Island and a range of 19.11%. This shows that there is a significant difference in the percentage of poor people in the various districts or cities analyzed. The lowest poverty rate is found in Tangerang Selatan in Banten at 2.50%, while the highest at 21.61% belongs to Sampang in East Java.

Multicollinearity Test

Multicollinearity test can be done by referring to the Variance Inflation Factor (VIF) value. If the VIF value exceeds 5, it indicates multicollinearity among the explanatory variables. In this study, the variables indicated to have a large correlation between variables are average years of schooling, number of schools, realization of food aid budget, food expenditure, non-food expenditure, and number of households with owned house assets.

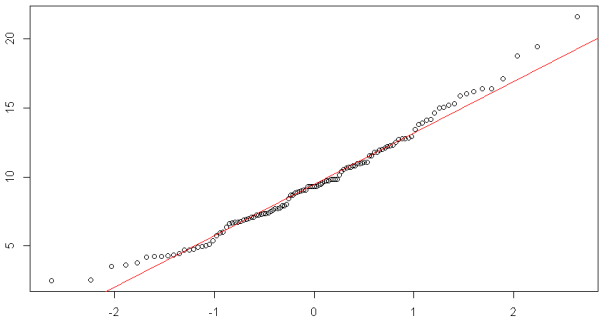
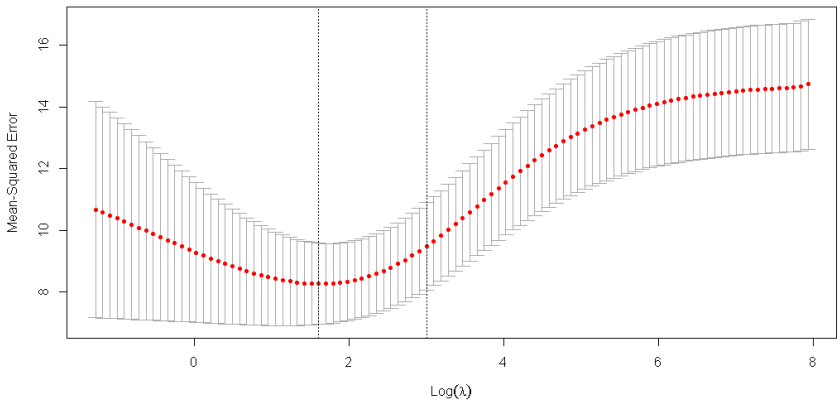
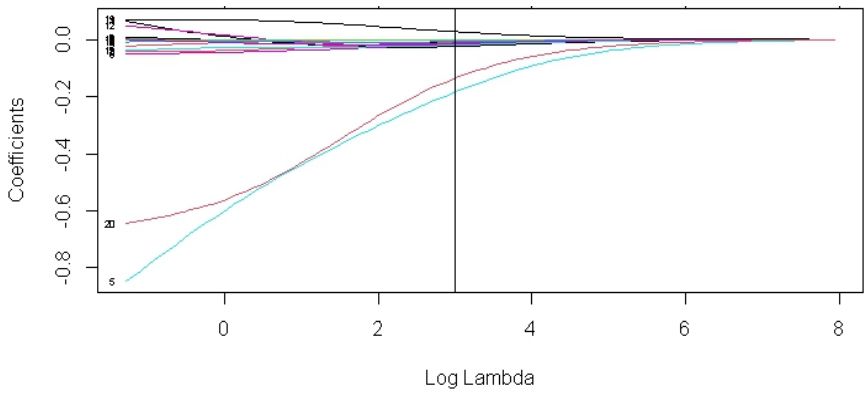


Figure 2. Normal Q-Q Plot of Poverty Rate Percentage

Optimum Lambda Value and Specification of Linear Mixed Model-LASSO



a)



b)

Figure 3. a) Prediction Error Curve for λ values. b) Coefficient selection for LMM-LASSO

The cross-validation results using k-fold with $k = 10$ in Figure 3a show that the optimum lambda value is marked by the right vertical line. In this model, the optimum lambda of 20.199 was selected as the tuning parameter for LMM-LASSO modeling. Figure 3b shows the change in $\boldsymbol{\beta}$ coefficient values as the value of λ increases. As λ increases, LASSO starts to shrink the coefficients of less significant variables to zero, so that only variables that have a large influence on the model remain. The vertical line on the plot shows the optimum λ value. The larger the value of λ, the more coefficients shrink to zero. The estimation results are presented in Table 2 below.

Table 2. Result of The Estiamtion Parameters

Variables	Estimated Coefficient	SE	p-*value*
Intecept	10.051	0.981	0.000*
Life Expectancy at Birth (year)	-		
Number of Health Centers	-		
Number of Family Planning Facilities	-		
Number of Health Insurance Recipients	-		
Average Years of Schooling (year)	-1.913	0.251	0.000*
Primary School Net Enrollment Rate (%)	-		
Junior High School Net Enrollment Rate (%)	-		
Senior High School Net Enrollment Rate (%)	-0.239	0.162	0.140
Number of School	-		
District/City Minimum Wage (IDR)	-0.101	0.170	0.550
Realization of Food Aid Budget (IDR)	-		
Percentage of PDRB	-		
Food Expenditure (IDR)	0.292	0.285	0.307
Non-Food Expenditure (IDR)	-0.973	0.304	0.001*
Number of Households with Housing Assets Owned	-0.576	0.124	0.000*
Percentage of Households with Proper Sanitation	-0.236	0.129	0.067
Percentage of Households with Access to Adequate Drinking Water	-0.175	0.114	0.125
Percentage of Gas Fuel Usage for Cooking	-		
Percentage of Households with a Dirt Floor	0.445	0.119	0.000*
Percentage of Households with PLN Lighting	-0.347	0.129	0.007*

*significant with $\alpha \leq 5\%$

The simultaneous test with the LRT-test resulted in a p-value of 0.000 ($p < 0.05$). This shows that the independent variables in the model, including random effects, jointly contribute significantly to the poverty percentage in Java. From Table 2, the partially significant variables are average years of schooling (X_5), non-food expenditure (X_{13}), number of households with housing assets owned (X_{15}), percentage of households with a dirt floor (X_{19}), and percentage of households with PLN lighting (X_{20}).

The result of random effect prediction $\boldsymbol{u}$ written as $\hat{\boldsymbol{u}}$ are presented in Table 3.

Table 3. Random Effect Prediction

Province	Province Code	$\hat{u}$
DKI Jakarta	31	1.152
Jawa Barat	32	-0.832
Jawa Tengah	33	-0.546
D.I. Yogyakarta	34	3.829
Jawa Timur	35	-0.339
Banten	36	-3.262

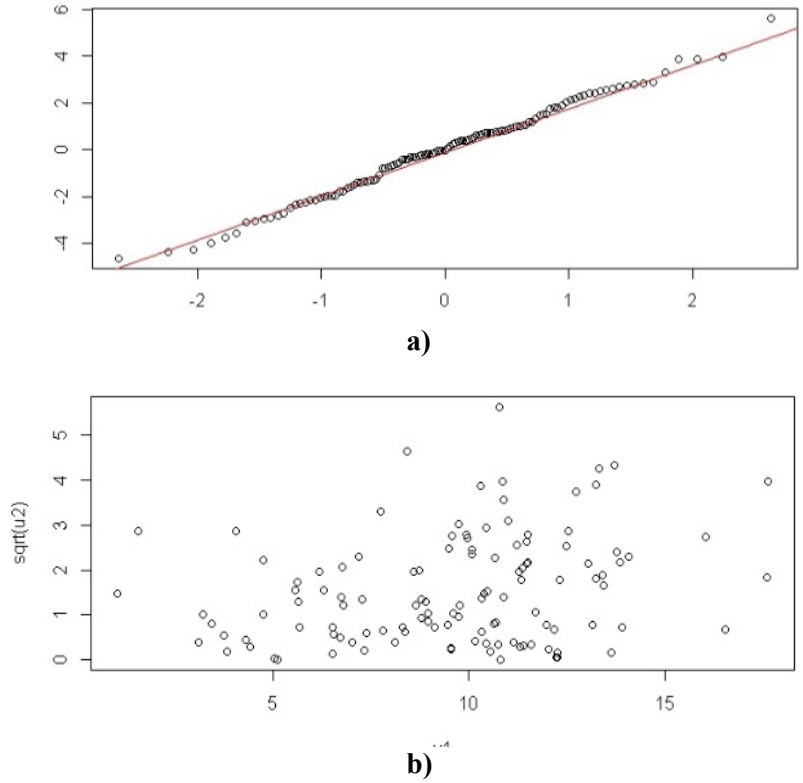
The result of random effect prediction ($\hat{\boldsymbol{u}}$) on Table 3 shows that there is a variation in the deviation of the poverty rate among provinces. The province of D.I. Yogyakarta has the largest positive deviation ($\hat{\boldsymbol{u}} = 3.829$), followed by DKI Jakarta ($\hat{\boldsymbol{u}} = 1.152$), indicating that the random effect causes a higher

poverty rate than the Java Island average. In contrast, Banten recorded the largest negative deviation ($\widehat{\boldsymbol{u}} = -3.262$). indicating that the random effect leads to a much lower poverty rate, followed by West Java ($\widehat{\boldsymbol{u}} = -0.832$), Central Java ($\widehat{\boldsymbol{u}} = -0.546$), and East Java ($\widehat{\boldsymbol{u}} = -0.339$).

Based on the results of parameter significance in Table 2 and random effect prediction in Table 3, after transformation back to the original explanatory variables, the model formed by linear mixed model-LASSO is as follows:

$$\hat{y}_i = 65.821 - 1.196X_5 - 0.023X_8 - 1.07 \times 10^{-7}X_{10} + 1.84 \times 10^{-6}X_{13} - 2.727 \times 10^{-6}X_{14} - 3.157 \times 10^{-8}X_{15} + 0.063X_{19} - 0.439X_{20} + 1.152Z_1 - 0.831Z_2 - 0.546Z_3 + 3.827Z_4 - 0.339Z_5 - 3.262Z_6$$

Diagnostics Test



a)

b)

Figure 4. a) Normal quantile-quantile plot of residuals. b) Residuals scatter plot

Normality tests were conducted using Q-Q plot and Kolmogorov-Smirnov test. The Q-Q plot shows that the model's residuals follow the normal distribution trend line. The Kolmogorov-Smirnov test yields a p-value of 0.775 ($p > 0.05$) indicating the normality assumption is fulfilled. For homoscedasticity, the scatter plot between the difference and the predicted value shows a random scatter plot with no particular pattern indicating the non-existence of heteroscedasticity symptoms. In addition, the Breusch-Pagan test yields a p-value of 0.220 ($p > 0.05$). Thus, based on diagram visualization and formal tests, the model fulfilled the assumptions of normality and homoscedasticity.

Goodnest of Fit

Table 4. Goodness of Fit Results and ICC

Critera	Results	
	LMM-LASSO	LMM
AIC	550.433	572.239
BIC	594.496	636.159
RMSE	2.086	1.988
Variance Random Effect	5.654	2.663

Critera	Results	
	LMM-LASSO	LMM
Variance Error	4.387	3.990
ICC	0.563	0.403

Table 4 shows that the linear mixed model-LASSO (LMM-LASSO) performs better than the linear mixed model (LMM) based on the AIC and BIC values. In addition, the ICC (Intraclass Correlation Coefficient) value is calculated based on the random effect variance (σ_u^2) of 5.684 and error variance (σ_e^2) of 4.387. The result shows an ICC value of 0.563 or 56.3%. The ICC value that is greater than 0.5 $(\sigma_u^2 > \sigma_e^2)$ indicates that the random effects, which is province level, in the model make a considerable contribution to the diversity of data at the district/city observation level in Java.

4. DISCUSSIONS

Average years of schooling has a significant negative effect on the poverty rate in Java Island. where every 1-year increase in average years of schooling reduces poverty by 1.913%. This suggests that better education can reduce poverty, in line with research by [17] and [18]. In addition, non-food expenditure also has a significant negative effect on poverty, with every 1 rupiah increase reducing poverty by 0.973%. These expenditures reflect a better quality of life, which aligns with [19] and [20]. Home ownership also plays an important role, where every increase of one household with owned housing assets reduces poverty by 0.576%, as found in [21] and [22]. Other variables, such as the percentage of households with dirt floors and the percentage of households with PLN lighting, have a significant influence on poverty. An increase in the percentage of households with dirt floors increases poverty by 0.445% in accordance with the research of [22] and [23]. Meanwhile, every 1% increase in households with PLN lighting reduces poverty by 0.347%. supporting the results of the study by [22] and [24].

Overall, factors such as education, non-food expenditure, home ownership, physical condition of the house, and access to electricity have a significant effect on poverty in Java. It reflects the close relationship between quality of life and poverty reduction. These variables show results consistent with previous studies that reveal that improvements in access to education, basic facilities, and infrastructure are instrumental in reducing poverty in a region.

Moreover, the result of random effect prediction shows that there are variations in poverty rates among provinces in Java Island after considering fixed factors. DI Yogyakarta and DKI Jakarta have higher poverty rates than the average, while Banten, West Java, Central Java, and East Java show lower poverty rates. This finding reflects that in addition to fixed factors, significant differences in policies, infrastructure, and local socio-economic conditions also affect poverty rates among provinces in Java.

5. CONCLUSION

Linear mixed model-LASSO was found to be suitable for modeling poverty data involving random effects from provinces. This is based on the considerable diversity of the random effects indicating the contribution of each province to the variation in poverty rates, as well as the fulfillment of model diagnostics, such as normality of the residuals and homoscedasticity. Significance test shows that the explanatory variables simultaneously have a significant influence on the poverty rate. Partially, the average years of schoolin, non-food expenditure, the number of households with owned housing assets, the percentage of households with dirt floor type, and the percentage of households with PLN lighting, are proven to have a significant effect on the poverty rate. Further research remains open by employing several alternative methods to model data affected by multicollinearity, such as ridge regression, SCAD, elastic net, MCP, and partial least squares. If penalty functions are to be used, the authors suggest

increasing the number of explanatory variables beyond the number of observations (high-dimensional data), as this may further improve model performance in modeling poverty rates.

6. ACKNOWLEDGMENTS

The author expresses gratitude to Universitas Negeri Jakarta for providing the necessary funding and assistance to complete this research.

7. REFERENCES

[1] M. P. Todaro and S. C. Smith. *Economic development*. Pearson. 2020.
[2] S. Sumarto and I. De Silva. "Beyond the Headcount: Examining The Dynamics and Patterns of Multidimensional Poverty In Indonesia." 2014. [Online]. Available: www.tnp2k.go.id
[3] Wulandari, H., & Aisyah, S. (2021). Analisis Determinan Kemiskinan Di Provinsi Jawa Dan Bali. Journal Of Economics Research And Policy Studies, 1(2), 106–116. Https://Doi.Org/10.53088/Jerps.V1i2.136
[4] Priseptian, L., Priana Primandhana, W., & Ekonomi Dan Bisnis Universitas Pembangunan Nasional Veteran Jawa Timur, F. (2022). Analisis Faktor-Faktor Yang Mempengaruhi Kemiskinan. FORUM EKONOMI, 24(1), 45–53. Http://Journal.Feb.Unmul.Ac.Id/Index.Php/FORUMEKONOMI
[5] C. E. McCulloch and S. R. Searle. "Generalized. Linear. and Mixed Models." 2001.
[6] V. M. Santi. S. Azzahra. and D. Siregar. "Analisis Skor Literasi Membaca Siswa Indonesia Menggunakan Linier Mixed Models." *MUST: Journal of Mathematics Education. Science and Technology*. vol. 7. no. 2. p. 116. Dec. 2022. doi: 10.30651/must.v7i2.14420.
[7] V. M. Santi. I. Hasari. and D. Handayani. "Linear Mixed Models to Analyze Indonesia's PISA Reading Literacy Score." in *AIP Conference Proceedings*. American Institute of Physics Inc.. Jan. 2024. doi: 10.1063/5.0183760.
[8] A. Dwinata. K. A. Notodiputro. and B. Sartono. "A Combination of Generalized Linear Mixed Model and LASSO Methods for Estimating Number of Patients Covid 19 in the Intensive Care Units." *CAUCHY: Jurnal Matematika Murni dan Aplikasi*. vol. 7. no. 1. pp. 13–21. Nov. 2021. doi: 10.18860/ca.v7i1.11575.
[9] R. Tibshirani. "Regression Shrinkage and Selection via the Lasso." 1996.
[10] V. M. Santi. K. Anwar Notodiputro. and B. Sartono. "A Study of Several Variable Selection Methods in Modelling the Mathematics Scores of Indonesian Students in Programme for International Student Assessment (PISA) based on Convex Penalized Likelihood." 2019.
[11] 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." in *Journal of Physics: Conference Series*. IOP Publishing Ltd. Apr. 2021. doi: 10.1088/1742-6596/1869/1/012140.
[12] J. Friedman. T. Hastie. and R. Tibshirani. "The Elements of Statistical Learning Preface to the Second Edition." 2008.
[13] Q. Wen. "On Lasso Estimation of Linear Mixed Model for High Dimensional Longitudinal Data." 2021.
[14] J. Jiang. "Linear and Generalized Linear Mixed Models and Their Applications." 2007.
[15] V. M. Santi. K. A. Notodiputro. I. Indahwati. and B. Sartono. "Restricted Maximum Likelihood Estimation for Multivariate Linear Mixed Model in Analyzing Pisa Data for Indonesian Students." *BAREKENG: Jurnal Ilmu Matematika dan Terapan*. vol. 16. no. 2. pp. 607–614. Jun. 2022. doi: 10.30598/barekengvol16iss2pp607-614.
[16] W. West and Gałecki. "Linear Mixed Models: A Practical Guide Using Statistical Software. Second Edition." 2015.
[17] N. P. Syabrina *et al.*. "Pengaruh pertumbuhan ekonomi. rata-rata lama sekolah dan tingkat pengangguran terhadap tingkat kemiskinan di Provinsi Jambi." 2021.
[18] A. Rafiqi. "Pengaruh Rata-Rata Lama Sekolah. Pengeluaran Riil Perkapita. Pertumbuhan Ekonomi dan Pengangguran Terhadap Tingkat Kemiskinan di Provinsi D.I Yogyakarta." 2020.
[19] 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." in *IOP*

Conference Series: Materials Science and Engineering. Institute of Physics Publishing. Jul. 2019. doi: 10.1088/1757-899X/546/5/052044.

[20] R. Hasanah. Syaparuddin. and Rosmeli. "Pengaruh angka harapan hidup. rata-rata lama sekolah dan pengeluaran perkapita terhadap tingkat kemiskinan pada Kabupaten /Kota di Provinsi Jambi." 2021.

[21] M. F. Fadilah and M. U. Basuki. "Analisis Faktor-Faktor yang Mempengaruhi Kerentanan Kemiskinan Relatif di Kota Jakarta Barat Tahun 2018." *Diponegoro Journal of Economics*. vol. 9. no. 2. p. 168. 2020. [Online]. Available: https://ejournal2.undip.ac.id/index.php/dje

[22] Risnawati. F. Zain Olilingo. B. Rantow Payu. and I. Abdul. "Determinan Kemiskinan di Kabupaten Parigi Moutong." *Jurnal Studi Ekonomi dan Pembangunan*. vol. 1(2). 2023. [Online]. Available: https://ejurnal.ung.ac.id/index.php/jsep

[23] A. D. Puteri and H. B. Notobroto. "Indikator Karakteristik Fisik Rumah Dominan dalam Penentuan Status Kemiskinan untuk Program Rehab Rumah tidak Layak Huni di Kabupaten Sidoarjo." 2016.

[24] F. Andrianus and K. Alfatih. "Pengaruh Infrastruktur terhadap Kemiskinan: Analisis Data Panel 34 Provinsi di Indonesia." *Jurnal Informatika Ekonomi Bisnis*. pp. 54–60. Mar. 2023. doi: 10.37034/infeb.v5i1.206.