MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: October 18, 2024
https://doi.org/10.31893/multiscience.2025187

The effectiveness of Liu-Estimator in predicting poverty levels in Indonesia: Comparative study and application of simulation

Netti Herawati[a] ✉ | Eva Silviana[a] | Khoirin Nisa[a] | Subian Saidi[a]

[a]Department of Mathematics, University of Lampung, Indonesia.

Abstract Several logistic regression studies have frequently encountered problems where the data contains multicollinearity. This condition negatively impacts parameter estimation. The Least Absolute Shrinkage and Selection Operator (LASSO) and the Liu Estimator are methods that can be employed to eliminate multicollinearity among independent variables. The objective of this research is to evaluate the effectiveness of the Liu Estimator in removing multicollinearity compared to the Maximum Likelihood Estimator (MLE) and LASSO, using simulation data for sample sizes of n = 25, 50, and 75 in a binary logistic model with 6 independent variables and a multicollinearity level of 0.99 between independent variables. The method with the smallest MSE and AIC values is used for comparison. The results of this research indicate that the Liu Estimator method is more effective than MLE and LASSO in addressing multicollinearity issues, as it produces lower MSE and AIC values for all sample sizes studied. The application of the Liu Estimator to poverty levels in Indonesia reveals that population density, the Human Development Index (HDI), average years of schooling, per capita expenditure, literacy rate, and life expectancy significantly influence the poverty rate in Indonesia.

Keywords: MLE, LASSO, liu estimator, multicollinearity, logistic regression, proverty level

1. Introduction

Regression analysis is a statistical method that is widely used to determine the relationships between one or more independent variables and dependent variables. Generally, the data used in regression analysis are continuous data with a normal distribution. However, in several studies, the dependent variable studied uses data categories, which state success and failure events or have values of 0 and 1. In regression analysis, the model used to analyze the relationship between the independent variable and the dependent variable uses dichotomous data, namely, the logistic regression model. The assumption of logistic regression analysis includes that there is no multicollinearity between independent variables; it does not require a linear relationship between independent variables and dependent variables; it does not require heteroscedasticity assumptions; independent variables do not have the same diversity between groups of variables; dependent variables are dichotomous; at least a sample of 50 sample data points for a predictor variable is necessary; and the independent variable does not require the assumption of multivariate normality. In several logistic regression studies, problems are often found where the data used contain multicollinearity. There are several methods used to overcome this multicollinearity problem, including least absolute shrinkage and selection (LASSO) and the Liu estimator. The LASSO method shrinks the coefficient (parameter β) to be exactly 0 or close to 0. This method has been proven to be able to remove multicollinearity in binary logistic regression (Herawati et al, 2020). On the other hand, Liu's method is an alternative in the form of a biased loss estimator that uses an estimated value of d, where the shrinkage parameter d can take a value between 0 and 1. This method is also able to address multicollinearity in binary logistic regression. The advantages and capabilities of Liu's method in overcoming multicollinearity have been studied as well as the application of this method to several real-life problems by several researchers (Mansson et al., 2012; Mansson et al., 2015; Jahufer, 2013; Saputri, et al., 2024).

In this study, the performance of the LASSO and LIU estimators was proven to eliminate multicollinearity using simulated data containing multicollinearity with several different samples. The best method has the smallest MSE and AIC values among the methods evaluated. Furthermore, this best method is applied to real data on poverty levels in Indonesia.

Poverty is a problem that continues to haunt all developing countries, including Indonesia. However, the poverty level in Indonesia in 2024 is expected to decrease spatially in both urban and rural areas. For example, in urban areas, the poverty level rate appears to have decreased to 7.09 percent from the previous level of 7.29 percent in March 2023. Moreover, in rural areas, the percentage of poor people decreased from 12.22% to 11.79% in March 2023. A decline in the poverty level has also occurred throughout Indonesia, with the greatest decline occurring in Bali and Nusa Tenggara. Examining the solid factors of domestic economic activity and various government social assistance programs, especially in response to the increase in food

 Received: August 8, 2024 | Accepted: July 27, 2024

inflation in early 2024, led to a reduction in the poverty level rate in March 2024 (Central Bureau of Statistics, 2024). This fact does not necessarily mean that the poverty rate in Indonesia has been effectively resolved and that Indonesia is free from poverty problems. Therefore, the best method obtained from simulation data will be applied to data on poverty levels in Indonesia to determine which factors still influence poverty levels in Indonesia.

2. Materials and methods

2.1. Binary logistic regression analysis

Logistic regression analysis is a regression analysis in which the dependent variable has binary or dichotomous properties with one or more independent variables (Hosmer & Lemeshow, 2000). A dichotomous or binary variable is a variable that has only two categories, namely, 0 and 1. The dependent variable is symbolized by y. Because it has two categories, for example, the category that states success events y = 1 and the category that states failure events y = 0.

According to Agresti (2002), the variable Y is a variable that follows the Bernoulli distribution. The probability function for Y with parameter (x) is as follows:

$$f(y) = \pi(x)^y \,(1-\pi(x))^{1-y} \;, y = 0,1 \quad (1)$$

The probability of the variable Y for a given value of x is denoted as π(x). According to Hosmer & Lemeshow (2000), the logistic regression model involving p as a predictor variable is as follows:

$$\pi(x) = \frac{\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+ \beta_p x_p)} \quad (2)$$

Maximum likelihood estimation (MLE) is a method used to estimate parameters in logistic regression. In maximizing the likelihood function, the MLE method provides an estimator parameter β and requires the data to follow a certain distribution. When y_i spreads binomially, from equation (2), the likelihood function is obtained as follows:

$$L(\beta|y) = \prod_{i=1}^{n} f(y|\beta)$$
$$= \prod_{i=1}^{N} \pi_i^{y_i}\,(1-\pi_i)^{1-y_i}$$
$$= (\pi_i)^{\sum y_i}(1-\pi_i)^{\sum 1-y_i} \quad (3)$$

Therefore, the likelihood function is obtained as follows:

$$(\beta|y) = \prod_{i=1}^{N} \frac{n_i!}{y_i!(n_i-y_i)!}\pi_i^{y_i}\,(1-\pi_i)^{n_i-y_i} \quad (4)$$

Then, the maximum possibility β can be obtained by setting each p + 1 in equation (4) to the same value as zero. Equation (4) can be written as follows:

$$\widehat{\beta_{MLE}} = \left(X^T\widehat{W}X\right)^{-1}X^T\hat{z} \quad (5)$$

where $\widehat{W} = diag[\widehat{\pi_\iota}(1-\widehat{\pi_\iota})]$ and $\hat{z}$ is the vector where the i[th] element has a value $\hat{z}_i = \log(\hat{\pi}_i) + \frac{y_i-\hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$.

2.2. Least absolute shrinkage and selection operator (LASSO)

The least absolute shrinkage and selection operator (LASSO) method can be used to overcome multicollinearity problems. LASSO was introduced by Tibshirani (1996), where LASSO works to shrink the correlated coefficients (parameter β) to be exactly zero or close to zero. According to Hastie, Tibshirani & Wainwright (2015), parameter estimation in LASSO is as follows:

$$\pi_i = \frac{1}{1+\exp\,(-\beta^T x_i')} \quad (6)$$

Lagrangian constraints ($\mathbf{L^1}$-norm) can be incorporated in estimating log-likelihood parameters in logistic regression. The combined log-likelihood equation for vector β is as follows:

$$l(\beta) = L(\beta \mid y_i, \ldots, y_n)$$
$$= \sum_{i=1}^{n}\left[y_i ln\left(\frac{1}{1+\exp\,(-\beta^T x_i')}\right) + (1-y_i)\left(\frac{\exp\,(\beta^T x_i')}{1+\exp\,(-\beta^T x_i')}\right)\right]$$
$$= \sum_{i=1}^{n}[(1-y_i)\beta^T x_i' + \ln\,(1+\exp(-\beta^T x_i'))\,] \quad (7)$$

The combined equation between the log-likelihood and Lagrangian constraints produces the following equation:

$$l(\beta) \quad = - \sum_{i=1}^{n}\left[(1-y_i)\beta^T x_i' \ + \ln\left(1+(-\beta^T x_i')\right)\right] - \lambda \sum_{k}^{p}\left|\beta_j\right| \tag{8}$$

Therefore, we obtain logistic regression parameter estimates with LASSO

$$\beta_\lambda^{LASSO} = argmax\,\{l(\beta) - \lambda \sum_{j=1}^{p}\left|\beta_j\right| \tag{9}$$

$\boldsymbol{\lambda}$ is the bias value in the LASSO method, where the λ value > 0. The λ value is obtained via several methods, including the cross-validation method (Tibshirani, 1996).

2.3. Liu's Method

Liu's method is an alternative method of logistic regression in the form of a bias loss estimator and generalized direct estimator proposed for linear regression models by Liu (1993) to overcome multicollinearity problems. Compared with other methods, the Liu estimator has advantages, such as having a scalar mean square error (SMSE) value that is smaller than the ridge estimate. Therefore, according to Liu (1993), the use of an alternative estimation method where the resulting parameter can be useful as a linear function of the shrinkage parameter d is recommended.

The shrinkage parameter d can take values between zero and one, and when d is less than one, we have $\|\hat{\beta}_d\| \leq \|\hat{\beta}_{MLE}\|$. Liu's method was further developed by Mansson et al. (2012) in the logistic regression model as follows:

$$\widehat{\beta_d} = \left(X'\widehat{W}X + I\right)^{-1}\left(X'\widehat{W}X + dI\right)\hat{\beta}_{MLE} \tag{10}$$

The estimates for the d value proposed by Hoerl & Kennard (1970) are as follows:

$$d_1 = max\left[0,\ \frac{\hat{a}_{j\,max}^2 - 1}{\frac{1}{\lambda_{j\,max}} + \widehat{a_{max}^2}}\right] \tag{11}$$

Furthermore, the proposed estimator, which is based on the concepts outlined in Kibria (2003), is as follows:

$$:d_2 = max\left[0,\ median\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\lambda_j} + \widehat{a_j^2}}\right] \tag{12}$$

$$d_3 = max\left[0,\ \frac{1}{p}\sum_{j}^{J}\left(\frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \widehat{a_j^2}}\right)\right] \tag{13}$$

Finally, the following estimator was proposed in which other quantiles in addition to the median were used and successfully applied by Khalaf & Shukur (2005).

$$d_4 = max\left[0, max\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{a}_j^2}\right] \tag{14}$$

$$d_5 = \ max\left[0, min\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} - \hat{a}_j^2}\right] \tag{15}$$

$$d_6 = \ max\left[0, median\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} - \hat{a}_j^2}\right] \tag{16}$$

2.4. Data and analysis

In this study, the data used were simulation data with n = 25, 50, and 75 and independent variables used as many as 6 variables (p = 6), with a correlation level between variables of 0.99 and 100 repetitions. Data simulation for data X via Monte Carlo simulation is based on McDonald & Galarneau (1975) with the following equation:

$$X_{ij} = (1-\rho^2)^{1/2} Z_{ij} + \rho Z_{i(p+1)} \tag{17}$$

where Z represents the data generated in the standard normal or normally distributed form and where N(0,1) and ρ are specified. The performance of the MLE, LASSO, and Liu methods used in this study was illustrated via simulation studies to show how this method can improve the estimation of logistic model parameters containing multicollinearity via R. The dependent variable Y is generated by the probability of logistic regression.

$$P(y_i = 1)Y = \ \pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)} \tag{18}$$

where β_0= 0 and $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$. The multicollinearity of the independent variables is calculated via the VIF. If the VIF value is > 10, it can be concluded that there is significant multicollinearity between the independent variables. The best method for estimating parameters is evaluated by the MSE and AIC. Next, the best method obtained will be applied to real data, namely, data on poverty levels in Indonesia, which consists of 6 independent variables and contains multicollinearity.

3. Results

To start the analysis, the first step involved simulating the data as described in the research methods section. After the data were generated, correlation values between independent variables were examined for each sample size used. The correlation values for n=25, 50 and 75 are presented in Table 1 below.

Table 1 Correlation values between variables for *n* = 25, 50 and 75.

n = 25	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	0.9855553	0.9813981	0.9892030	0.9863767	0.9820822
X_2	0.9855553	1	0.9847641	0.9876933	0.9917750	0.9873204
X_3	0.9813981	0.9847641	1	0.9841272	0.9822016	0.9738643
X_4	0.9892030	0.9876933	0.9841272	1	0.9859681	0.9864211
X_5	0.9863767	0.9917750	0.9822016	0.9859681	1	0.9908091
X_6	0.9820822	0.9873204	0.9738643	0.9864211	0.9908091	1
n = 50	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	0.9830985	0.9855550	0.9785363	0.9817809	0.9818400
X_2	0.9830985	1	0.9819366	0.9799110	0.9799350	0.9807683
X_3	0.9855550	0.9819366	1	0.9857661	0.9831067	0.9835635
X_4	0.9785363	0.9799119	0.9857661	1	0.9765319	0.9754445
X_5	0.9817809	0.9799350	0.9831067	0.9765319	1	0.9814711
X_6	0.9818400	0.9807683	0.9835635	0.9754445	0.9814711	1
n = 75	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	0.9873363	0.9832745	0.9878074	0.9859102	0.9841967
X_2	0.9873362	1	0.9873720	0.9882705	0.9890132	0.9852008
X_3	0.9832745	0.9873720	1	0.9818524	0.9866369	0.9821433
X_4	0.9878074	0.9882705	0.9818524	1	0.9868148	0.9821531
X_5	0.9859102	0.9890132	0.9866369	0.9868148	1	0.9853063
X_6	0.9841967	0.9852008	0.9821433	0.9821531	0.9853063	1

Table 1 shows that the independent variables have high correlation values above 0.5. This shows that there is a high possibility that there is a strong relationship between variables, which allows for multicollinearity between variables. For this reason, an evaluation was carried out on the VIF values between independent variables for n = 25, 50, and 75. The results of the analysis can be seen in Table 2.

Table 2 VIF values.

n	Independent Variable					
	X_1	X_2	X_3	X_4	X_5	X_6
25	60.1037	86.6308	44.8282	81.0369	104.1233	72.4655
50	50.6109	44.7655	67.2899	41.7368	42.2690	43.5353
75	61.2666	79.4079	52.0325	61.8288	69.9952	47.1938

Table 2 shows that VIF values > 10 were obtained for all the independent variables in all the sample data studied (n=25, 50, 75). This means that multicollinearity occurs due to the existence of independent variables in each sample size used. On the basis of the results of the correlation test and the VIF values above, it can be concluded that there is multicollinearity between the independent variables of n = 25, 50, and 75. This is in accordance with what is expected in these research data. After appropriate data are obtained, the next step is to estimate the parameters via the MLE, LASSO and Liu methods and compare the results on the basis of the standard error (SE) value of each estimator for n=25, 50, and 75. The results of the analysis can be seen in Table 3 below.

From Table 3, it can be concluded that the MLE, LASSO, and Liu estimators at n = 25, 50, and 75 have varying $\hat{\beta}_p$ values. Compared with that of the MLE method, the estimated value of $\hat{\beta}_p$ in the LASSO and Liu methods is closer to the actual parameter, namely, $\hat{\beta}_p$ =1. However, if we look in detail at the SE value for the Liu method, the SE value for Liu is smaller than the SE values for the MLE and LASSO methods. This shows that the estimation of the parameters $\hat{\beta}_1 - \hat{\beta}_6$ via the Liu method is much better than that via the MLE and LASSO methods. The smaller the SE value of a parameter is, the better it is at estimating the parameter, especially in data that contain multicollinearity, and the parameter estimator is good for use in the model used. In addition, to ensure the results, the MSE and AIC values of the MLE, LASSO and Liu methods are evaluated to determine the best method. The results of the analysis of the MSE and ACI values are shown in Table 4.

Table 3 Values of $\hat{\beta}_p$ and SE in the MLE, LASSO, and Liu methods for *n* = 25, 50 and 75.

n = 25	$\hat{\beta}_p$			SE		
	MLE	LASSO	LIU	MLE	LASSO	LIU
$\hat{\beta}_1$	1.15e+14	1.1168	-1.24e – 05	19.389348	0.9573	0.0999
$\hat{\beta}_2$	-1.07e+12	0.6428	-3.16e – 07	8.15527051	4.4381	0.0999
$\hat{\beta}_3$	3.28e+ 13	2.0811	-2.32e – 06	19.4251196	0.0157	0.0999
$\hat{\beta}_4$	-1.40e+14	0.6656	-1.24e – 05	5.78962215	2.884	0.0996
$\hat{\beta}_5$	-6.59e+12	0.8240	-2.28e – 08	37.2348995	0.1000	0.0999
$\hat{\beta}_6$	1.002e+14	2.8538	-1.23e - 08	15.3285868	1.4645	0.1000
n = 50	$\hat{\beta}_p$			SE		
	MLE	LASSO	LIU	MLE	LASSO	LIU
$\hat{\beta}_1$	125.5341	0.4807	0.0017	16.5464	45.0760	0.09999014
$\hat{\beta}_2$	64.3883	0.7259	0.0001	4.7514	3.4023	0.0999736
$\hat{\beta}_3$	276.4802	1.1891	0.0004	45.0222	156.6259	0.09988864
$\hat{\beta}_4$	2.9544	1.6238	0.0018	5.6277	11.7517	0.09999941
$\hat{\beta}_5$	-145.624	2.1738	5.12e – 05	15.0407	0.0001	0.09999941
$\hat{\beta}_6$	82.0230	2.1177	1.50e – 05	3.6686	176.3854	0.1
n = 75	$\hat{\beta}_p$			SE		
	MLE	LASSO	LIU	MLE	LASSO	LIU
$\hat{\beta}_1$	3.2821	0.9562	0.0015	0.71536876	17.9985	0.09999991
$\hat{\beta}_2$	18.9339	1.1492	0.0003	3.86477248	0.1000	0.09999903
$\hat{\beta}_3$	4.3312	1.3643	0.0005	12.0116452	16.6810	0.09999583
$\hat{\beta}_4$	16.9225	1.7549	0.0016	5.92102649	1.7256	0.1000
$\hat{\beta}_5$	6.7331	0.9265	5.05e – 06	3.12660553	3.1920	0.1000
$\hat{\beta}_6$	12.6864	1.0088	3.72e – 06	5.06832898	4.8707	0.1000

Table 4 MSE values for MLE, LASSO, and LIU.

Sample Sizes	MSE			AIC		
	MLE	LASSO	LIU	MLE	LASSO	LIU
n =25	267.36	1.3455	0.0599	14.05	-48.2964	-59.4711
n =50	349.7437	1.3903	0.0599	14.03	-105.0664	-109.4559
n =75	343.0046	1.2709	0.0598	14.00	-172.44	-177.8635

In Table 4, it can be seen that at n = 25, 50, and 75, the smallest MSE value was obtained via the Liu method compared with the MSEs of the MLE and LASSO methods. The average MSE value for Liu's method is 0.0599 for the three data samples. Therefore, the best method that can be used to overcome multicollinearity is the Liu method. This is because the smaller the MSE value of a model is, the better and more accurate the modeling value obtained will be. Apart from that, the AIC value of the Liu method was also proven to be smaller than those of MLE and LASSO. The larger the sample size is, the smaller the AIC value obtained. Therefore, the best method that can be used to overcome multicollinearity on the basis of the MSE and AIC is the Liu method.

4. Discussion

4.1. Application of Liu's method to real data

The real data used in this research are secondary data obtained from the Central Statistics Agency, Indonesia. The dependent variable (Y) is a nominal variable that has only two categories or levels, namely, 0/poor and 1/not poor. Moreover, the independent variable (X) consists of 6 variables, namely, population density (X_1), the human development index (X_2), the average length of schooling (X_3), per capita expenditure (X_4), the literacy rate (X_5) and life expectancy (X_6). Before the data were analyzed via Liu's method, multicollinearity was first checked. Table 5 shows the correlations between the independent variables.

Table 5 shows that there is a high correlation between X_4 and X_1, X_2 and X_4, and X_2 and X_3. This finding indicates that the intervariables have a strong relationship and allows the variables to have multicollinearity. To ensure this, the VIF values are checked for each independent variable: X_1 = 3.10, X_2 = 33.44, X_3 = 12.69, X_4 = 12.03, X_5 = 2.48 and X_6 = 4.87. The variables X_2, X_3, X_4 and X_5 have a VIF value of > 10, which indicates that the variable contains multicollinearity. Next, the β_p and SE parameters were estimated on the poverty level percentage data in Indonesia via the best method resulting from the simulation above, namely, the Liu method. The parameter estimation results from data on the percentage of poverty level in Indonesia are presented in Table 6.

Table 5 Correlation between Independent Variables and Poverty Level Data in Indonesia.

Correlation	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	0.5700	0.5492	0.7099	0.1661	0.2657
X_2	0.5700	1	0.8215	0.8535	0.3921	0.6806
X_3	0.5492	0.8215	1	0.6357	0.6445	0.3812
X_4	0.7099	0.8535	0.6357	1	0.2641	0.4114
X_5	0.1661	0.3921	0.6445	0.2641	1	0.2035
X_6	0.2657	0.6806	0.3812	0.4114	0.2035	1

Table 6 $\hat{\beta}_p$ and W_i for poverty level data in Indonesia.

Variable	Liu Method		$X^2_{1,0.05}$
	$\hat{\beta}_p$	W_i	
$\hat{\beta}_1$	1.8014	505.2514	3.841
$\hat{\beta}_2$	-0.3443	1519.193	
$\hat{\beta}_3$	-0.3444	6471.665	
$\hat{\beta}_4$	1.8014	13390	
$\hat{\beta}_5$	-0.3443	2828.419	
$\hat{\beta}_6$	-0.3443	678.9153	

Table 6 gives the estimated values of $\hat{\beta}_1$--$\hat{\beta}_6$ for the Liu method. On the basis of these values, the logistic regression model for poverty level percentage in Indonesia based on Liu's method is as follows:

$$\hat{Y} = 1.8014 - 0.3443X_2 - 0.3443X_3 + 1.8014X_4 - 0.3443X_5 - 0.3443X_6$$

Next, parameter testing was carried out on the model to determine which independent variables influence the percentage of poverty level in Indonesia via the Wald test with $H_0: \hat{\beta}_i = 0$; $H_1: \hat{\beta}_i \neq 0$ and the critical value of the Wald test= 3.841. Reject H_0 if $W_i > X^2_{1,0.05}$. From the results of calculating the Wald value for each independent variable presented in Table 1 above, it can be concluded that all variables (X_1, X_2, X_3, X_4, X_5, and X_6) have $W_i > X^2_{1,0.05} = 3.841$. All the variables, namely, population density, the human development index, the average length of schooling, per capita expenditure, the literacy rate, and life expectancy, significantly affect the poverty level of people in Indonesia.

The variables that influence poverty levels in Indonesia from the test results above will be explored more deeply by examining their relationships one by one. The first variable that influences the level of poverty in Indonesia is the population density. The population density of a country can increase the number of productive ages or workers. As the workforce continues to grow, unemployment will increase, and the number of people who exploit nature because they are unable to meet their daily needs will increase, which will automatically affect the level of poverty in the country (Dita & Legowo, 2022).

The second variable is the human development index, which has a significant influence on poverty levels in Indonesia. This finding is in line with previous research showing that the human development index can increase the percentage of poverty level in Indonesia (Mukhtar et al, 2019). The human development index is a measure of quality of life. Quality of life is measured on the basis of three basic dimensions of human development achievements, namely, a long and healthy life, knowledge, and a decent life. Health dimensions are measured via life expectancy at birth. Moreover, the combination of the literacy rate and average years of schooling is a measure of the knowledge dimension and is an indicator of people's ability to purchase a number of basic needs, which is calculated from the average amount of expenditure per capita as an income approach from development achievements toward a decent life.

Furthermore, a variable that influences poverty levels in Indonesia is the average length of schooling. The results of this research show that the average length of schooling influences poverty levels in Indonesia. This finding is in accordance with previous research, which also shows that there is a very strong relationship between the average length of schooling and the percentage of poor people, namely, the higher the average number of years of schooling is, the lower the percentage of poor people (Asro & Ahmad, 2018). This also occurs because the average length of schooling is the factor most considered in increasing the human development index (Listiani, et al., 2022).

Next is the per capita expenditure variable for all households in Indonesia for personal consumption, which also has an effect on the poverty level in Indonesia. This is obvious since per capita expenditure is closely related to income level. If the income level is low, then it cannot meet the expenses needed for living needs. The results of this study are the same as the results of previous research, which showed that average per capita expenditure had a significant influence on poverty level levels in 2014 (Setiawan & Adjim, 2017). Finally, the results of the model also show that literacy rates and life expectancy rates are factors that can reduce poverty rates in Indonesia, where literacy rates are the proportion of the population of certain age groups who have the ability to read and write in Latin letters, Arabic letters, and other letters. The higher the literacy rate is, the more the poverty rate decreases significantly. The quality of good human resources positively affects a country's economic progress. Moreover, life expectancy is an important parameter for assessing the health of individuals in an area. With an increase in life expectancy, people have a greater life expectancy, which has the potential to increase economic opportunities to achieve higher incomes (Johri & Jain, 1984; Messias, 2003).

5. Conclusions

On the basis of the results of research using simulation data with n=25, 50, and 75, the Liu method is the best method for solving multicollinearity problems because the MSE value of the Liu method is smaller than the MSE values of the MLE and LASSO methods. In addition, Liu's method for poverty level data in Indonesia revealed that population density, the human development index, the average length of schooling, per capita expenditure, the literacy rate, and life expectancy significantly influence poverty levels in Indonesia. These results suggest that Indonesia needs improved education to obtain human resources that are able to compete both now and in the future. For future research, this method can be applied in different contexts and data.

Ethical considerations

Not applicable.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References

Agresti, A. (2002). *An Introduction to Categorical Data Analysis*. 2nd Edtion. John Wiley & Sons, Inc.

Asro, A., & Ahmad, A. (2018). Analysis of factors that affect poverty level in Indonesia. *ESPACIOS. 39* (45), 14. http://www.1.revistaespacios.com/a18v39n45/a18v39n45p14.pdf

Central Bureau of Statistics. (2024). *Poverty Profile in Indonesia 2023*. https://www.bps.go.id/

Dita, C. Y. E., & Legowo, M. (2022). Analysis of population density that affects poverty and environmental degradation. *Proceedings of the 2022 National Seminar on Social Sciences (SNIIS)*. Department of Social Sciences FISH, State University of Surabaya. https://proceeding.unesa.ac.id/index.php/sniis/article/view/34.

Hastie, T., Thibsirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity the LASSO and Generalizations*. John Wiley & Sons, Inc.

Herawati, N., Nisa, K., & Nusyirwan. (2020). Selecting the Method to Overcome Partial and Full Multicollinearity in Binary Logistic Model. *International Journal of Statistics and Applications*, *10*(3), 55-59. doi:10.5923/j.statistics.20201003.01

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67. https://doi.org/10.2307/1267351

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistics Regression*. JohnWiley & Sons, Inc.

Jahufer, A. (2013)..Detecting Global Influential Observations in Liu Regression Model. *Open Journal of Statistics,* .3(1), 5-11. https://doi.org/ 10.4236/ojs.2013.31002

Johri, C.K., & Jain, B.K. (1984). Poverty level, Literacy and Life Expectancy: A Taxonomic Analysis. *Indian Journal of Industrial Relations*, *20*(2), 211-221. https://www.jstor.org/stable/i27768806

Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics- Theory and Method*, *34*(5), 1177-1182. https://doi.org/10.1081/STA-200056836

Kibria, B. M. G. (2003). Performance of Some New Ridge Regression Estimators. *Communications in Statistics Simulation and Computation*, *32*(2), 419-435. https://doi.org/10.1081/SAC-120017499

Listiani, N., Zulfikar, R., Jaelani, A., & Susanto, D. (2022) The Years of Schooling Contribution as Factors That Most Considered in Increasing Human Development Index. *International Journal of Educational, Research & Social Sciences*, *3*(4), 1599-1606. https://doi.org/10.51601/ijersc.v3i4.431

Liu, K. (1993). A New Class of Blased Estimate in Linear Regression. *Communications ins Statistics- Theory and Method*, *22*(2), 393-402. https://doi.org/10.1080/03610929308831027

Mansson, K., Kibria, B. M. G, & Shukur, G. (2015). A restricted Liu estimator for binary regression models and its application to an applied demand system. *Journal of Applied Statistics, 43*(6),1-9. https://doi.org/10.1080/02664763.2015.1092110

Mansson, K., Kibria, B. M. G, & Shukur, G. (2015). Performance of Some Weighted Liu Estimators for Logit Regression Model: An Application to Swedish Accident Data. *Communication in Statistics- Theory and Methods*. 44(2). 363–375. https://doi.org/10.1080/03610926.2012.745562

Mansson, K., Kibria, B. M. G., & Shukur, G. (2012). On Liu estimators for the logit regression model. *Econ. Modell*, *29*, 1483–1488. https://doi.org/10.1016/j.econmod.2011.11.015

McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*, *70*(350), 407-416. https://doi.org/10.1080/01621459.1975.10479882

Messias, E. (2003). Income inequality, illiteracy rate, and life expectancy in Brazil. *Am J Public Health*. *93*(8):1294-1296. https://doi.org/10.2105/AJPH.93.8.1294

Mukhtar, S., Saptono, A., & Arifin, A.S. (2019). The Analysis of the Effects of Human Development Index and Opened Unemployment Levels to the Poverty level In Indonesia. *Jurnal Ecoplan*, *2*(2), 77-89.

Saputri, S.A., Herawati, N., Sutrisno, A. , Nusyirwan, & Nisa, K. (2024). Comparison of MLE, Lasso, and Liu Estimator Methods to Overcome Multicollinearity In Multinomial Logistic Regression: Simulation Study. *Global Scientific and Academic Research Journal of Multidisciplinary Studies*. 3(5), 08-12.

Setiawan, A. V., & Adjim, F. (2017). Determinant of Poverty level in Indonesia. *Economics Development Analysis Journal*, *6*(1), 22-29. https://doi.org/10.15294/edaj.v6i1.22197

Tibshirani, R. (1996). Regression Shrinkage and Selection via LASSO. *Journal of the Royal Statistical Society*, *58*(1), 267-288. : https://www.jstor.org/stable/2346178