# Poverty Level Analysis in Indonesia Using the Stochastic Restricted Maximum Likelihood Approach Method

**Riaman, Sudradjat Supian, Sukono**
Department of Mathematics, FMIPA, Padjadjaran University, Jatinangor 40363, Indonesia
riaman@unpad.ac.id, adjat03@yahoo.com; sukono@unpad.ac.id

**Abdul Talib Bon**
Department of Production and Operations, University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

## Abstract

Analysis of economic data relating to the level of poverty in Indonesia is conducted to determine the factors that affect the poverty index. This needs to be done as a material for the government's consideration in planning national development. Model analysis was performed using a logistic regression model. Logistic regression analysis is a statistical analysis method that describes the relationship between the dependent variable (response) which has two or more categories with one or more independent variables (predictors) on the category scale or interval scale, with the dependent variable coded zero if the province has a poverty percentage below percentage of poverty in Indonesia, and coded if the province has a poverty percentage above or equal to the percentage of poverty in Indonesia. This study aims to analyze the level of poverty in Indonesia using a logistic regression model. Parameter estimation is done using the Stochastic Restricted Maximum Likelihood Estimator (SRMLE), where this method can only be used on data containing multicollinearity. Multicollinearity data often appears in economic data because an increase in the value of one variable will affect the increase in other variables. Four independent variables were taken, namely the unemployment percentage, the underemployment percentage, the poverty depth index and the poverty severity index. Of the four variables, the Poverty Depth Index and Poverty Severity Index are variables that contain multicollinearity. The results of the analysis with a significance level of 0.05 indicate that the percentage of unemployment and depth of the poverty Index are significant variables for the poverty rate in Indonesia.

**Keywords:**
Level of poverty, Logistic Regression Analysis, Stochastic Restricted Maximum Likelihood Estimator, Multicollinearity.

## 1. Introduction

Indonesia has a standardization for poverty where in 2018 the poverty rate in Indonesia was recorded at 9.66% (Badan Pusat Statistik, 2019), in other words 9.66% of Indonesia's population is still below the poverty line. Poverty is one of the problems that is never absent from the attention of the government of a country. Poverty has even become a phenomenal problem in the economic field which has become a measure of the success of state governments from time to time, especially in developing countries. One of the root causes of poverty related to a high population is the existence of job opportunities that cannot accommodate the needs of the workforce, resulting in unemployment which leads to poverty formation.

Unemployment is the number of workers in the economy who are actively looking for work but have not yet obtained it (Kutner, et al, 2004). Apart from unemployment, there are other things that are very influential in the level of poverty in Indonesia, namely the index of the depth and severity of poverty in each region in Indonesia. The Poverty Gap Index (P1), is a measure of the average expenditure gap of each poor person against the poverty line. The higher the index value, the further away the population's average expenditure is from the poverty line. Another factor is the Poverty Severity Index (P2), which provides an overview of the distribution of expenditure among the poor. The higher the index value, the higher the expenditure inequality among the poor.

Logistic regression is an approach to creating predictive models, such as linear regression or whose parameters are estimated using Ordinary Least Squares (OLS) regression. In logistic regression, the dependent variable has a dichotomous scale, namely a nominal data scale with two categories, for example: "yes" and "no", "good" and "bad" or "high" and "low". If the OLS requires conditions or assumptions that the error is normally distributed, on the contrary, in logistic regression this assumption is not needed because in this type of logistic regression it follows a logistic distribution.

In this study, the authors take the unemployment rate, underemployment rate, poverty depth index, and poverty severity index for each province in Indonesia as factors that are thought to be significantly related to the poverty rate occurring in Indonesia. To model the case used binary logistic regression analysis based on the Stochastic Restricted Maximum Likelihood Estimator (SRMLE) approach, as an analysis tool.

### 1.1 Objectives

The purpose of this research is to estimate logistic regression parameters in determining the poverty rate model in Indonesia. The purpose of this study is to obtain a logistic regression model that can be used to determine the level of poverty in Indonesia. Determine what factors affect the level of poverty in Indonesia.

## 2. Literature Review

Logistic regression is a statistical method applied to model categorical independent variables (nominal/ordinal scale) based on one or more predictor modifiers which can be categorical or continuous variables (interval or ratio scale). The method used in this study is binary logistic regression analysis where binary logistic regression analysis is used to analyze the relationship between one response variable and several predictor variables, with the response variable in the form of dichotomous qualitative data, which is 1 to state the existence of a characteristic and 0 to state. the absence of a characteristic (Tampil, 2017).

To estimate the parameters in logistic regression, the Maximum Likelihood Estimator (MLE) method is usually used, but MLE cannot handle the type of data that contains multicollinearity. With the multicollinearity in logistic regression, it will result in high variance values in the Maximum Likelihood Estimator (MLE) estimation (Siray, 2015). Therefore, there must be a special method in determining parameter estimation with data that contains multicollinearity, one method that can overcome multicollinearity is the Stochastic Restricted Maximum Likelihood Estimator (SRMLE).

SRMLE was popularized by Nagarajah and Wijekoon in 2015 where the SRMLE estimator is a development of the Restricted Maximum Likelihood Estimator (RMLE) method studied by Duffy and Sunter in 1989. The Stochastic Restricted Maximum Likelihood Estimator (SRMLE) can be done if there is a linear stochastic limit that is known other than logistic regression model. In addition, the use of the SRMLE method in logistic regression is carried out with data that has multicollinearity (Nagarajah and Wijekoon, 2015). Multicollinearity data often appears in economic data because an increase in the value of one variable will affect the increase in other variables (Sumodiningrat, 1994). One of the economic topics that is often raised is about data on the poverty of an area, in this study the author will use poverty data for each province in Indonesia in 2018.

SRMLE was popularized by Nagarajah and Wijekoon in 2015 where the SRMLE estimator is a development of the Restricted Maximum Likelihood Estimator (RMLE) method studied by Duffy and Sunter in 1989. The Stochastic Restricted Maximum Likelihood Estimator (SRMLE) can be done if there is a linear stochastic limit that is known other than logistic regression model. In addition, the use of the SRMLE method in logistic regression is carried out with data that has multicollinearity (Nagarajah and Wijekoon, 2015). Multicollinearity data often appears in economic data because an increase in the value of one variable will affect the increase in other variables (Sumodiningrat, 1994). One of the economic topics that is often raised is about data on the poverty of an area, in this study the author will use poverty data for each province in Indonesia in 2018.

Regression analysis is a tool used to analyze two or more data in which it is studied how the variables are correlated (Kutner, et al, 2004). The relationship obtained is generally expressed in the form of a mathematical equation which states the functional relationship between the independent variables and one dependent variable. In regression analysis, there are two types of variables, namely Variable Response (dependent variable or dependent variable). This variable's existence is influenced by other variables and is denoted by the symbol $Y$.

Second, the predictor variable (independent variable), which is a variable that is not influenced by other variables and is denoted by $X$.

## 3. Methods

Regression analysis is known in two forms, depending on the number of independent variables ($X$). If the independent variable ($X$) is only one, then the regression analysis is called a simple linear regression analysis. But if

the independent variable ($X$) is associated with more than one dependent variable ($Y$), then regression analysis is called multiple linear regression analysis (multi regression).
The requirements for linear regression include the data must be on an interval or ratio scale, the error is normally distributed and there is no correlation between the predictor variables (multicollinearity) for multiple regression. Here is a linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

$\varepsilon_i$ is an error with an identical distribution to $\varepsilon_i \sim N(0, \sigma^2)$ for $i$ = 1, 2, …, $n$ and $p$ is the number of independent variables, the estimator function model (1) is obtained from the sample data so that it takes the form:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \cdots + \hat{\beta}_p x_{pi}, i = 1,2, \ldots n.$$

A common technique for estimating parameters is to use the Ordinary Least Square (OLS) regression method. This technique is also useful as an indication of the suitability of the model with existing data. The suitability of the model can be determined by comparing the actual value with the estimated $y$ value. Estimated value using the OLS method for the regression coefficient $\hat{\beta}_{OLS}$ is:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\boldsymbol{y}. \tag{2}$$

Logistic regression analysis is a statistical analysis method that describes the relationship between the dependent variable (response) which has two or more categories with one or more independent variables (predictors) on a category scale or an interval scale (Hosmer and Lemeshow, 2000). The logistic regression model is obtained from the standard logistic distribution, with the opportunity density function as follows:

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2} \text{ , for } -\infty < x < \infty. \tag{3}$$

If the response variable is dichotomous, it is called binary logistic regression analysis. It is said to be dichotomous because it contains two values, namely 1 if an event occurs and 0 if the event does not occur.

In binary logistic regression, there are hidden variables that underlie the tendency of the variable $Y = 1$. The hidden variable constructs the response variable $Y$ and is written in

$$y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon_{Y^*}, \tag{4}$$

where $-\infty < y < \infty$. If $y^* \geq 0$ then $y = 1$, if $y^* < 0$ then $y = 0$ (Cakmakyapan dan Goktas, 2013).

In the linear regression model, the mean of the $Y$ response variable is written as follows:

$$E(Y|X = x_i) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \text{ , } -\infty < E(Y|X) < \infty. \tag{5}$$

Some linear regression models the value of $E(Y|X)$, logistic regression model opportunities, namely $P(Y = 1|X)$ where,

$$P(Y = 1|X) = P(Y^* \geq 0) = P(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \geq 0) = P(\boldsymbol{\varepsilon} \leq \mathbf{X}\boldsymbol{\beta})$$

$\pi(x)$ represents $E(Y|X)$, where $\pi(x)$ is the standard cumulative distribution function of the Bernoulli distribution (Nagarajah and Wijekoon, 2015). The logistic regression model is as follows:

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

or

$$\pi(x) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)]} \text{ ,} \tag{6}$$

where $p$ the number of predictor variables. The function of the logistic regression model in equation (6) is not linear, so by using the logit transformation the function becomes linear in its parameters. After the logit transformation, the logistic regression model has an equation:

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p\ , \tag{7}$$

where $\pi(x)$ probability occurs, $1-\pi(x)$ probability does not occur, and $\left[\frac{\pi(x)}{1-\pi(x)}\right]$ is the odds value. Odds are the tendency to experience $y = 1$. Odds ratio is the comparison of the odds for one event to another. Odds ratio describes the change (increase or decrease) in the tendency for each additional unit of the independent variable, if the independent variable is continuous or the difference in trend between categories, if the independent variable is categorical (Hosmer, Lemeshow, and Sturdivant, 2013).

The regression model is assumed not to contain a linear dependency relationship between independent variables. If there is a high correlation between the independent variables, the relationship between the independent variables and the dependent variable will be disrupted so that the multicollinearity problem will arise. If there is multicollinearity, the estimation results of the coefficients will be invalid.

Individually, the independent variable does not affect the dependent variable through the t test, but the coefficient of determination $(R^2)$ can still be relatively high. To detect multicollinearity in multiple linear regression models, the Variance Inflation Factor (VIF) value can be used provided that if the VIF value exceeds the number 10, multicollinearity occurs in the regression model. The formula for finding the VIF value is:

$$VIF_j = \frac{1}{1-{R_j}^2}\ , \qquad j = 1,2,\dots,p \tag{8}$$

where ${R_j}^2$ is the coefficient of determination between $X_j$ and other predictor variables. The value of ${R_j}^2$ is obtained from the formula:

$${R_j}^2 = \frac{\text{JK Regresi j}}{\text{JK Total j}} = \frac{\sum_{i=1}^{n}(\hat{X}_{ij} - \bar{X}_j)^2}{\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2} \tag{9}$$

If the value of VIF > 10, it can be said that there is multicollinearity in the data (Williams, 2015).

The type of data is very influential on the method to be used, for logistic regression we can usually use the MLE estimator as an estimate with data that does not contain multicollinearity and Heteroscedasticity. In this study, the data used are data that contain multicollinearity, the estimation that can be used is SRMLE (Nagarajah and Wijekoon, 2015). In general, the maximum likelihood estimate (MLE) is a method for obtaining an estimate of the unknown parameter by optimizing the likelihood function. To apply the maximum likelihood estimate, first construct the likelihood as a function of the parameter in the specified model, based on the distribution assumption. Stochastic Restricted Maximum Likelihood Estimator (SRMLE) is used when a stochastic linear constraint is available in addition to the logistic regression model. In addition, this SRMLE can also be used to solve multicollinearity problems (Nagarajah and Wijekoon, 2015). The linear limitation can be written as follows:

$$\boldsymbol{h} = \boldsymbol{S\beta}, \tag{10}$$

where

$$\mathbf{S} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}$$

where $\mathbf{S}$ is a full rank matrix $(p \times (p+1))$ with elements consisting of 0.1 and -1 and meets the full rank column and row requirements. In addition, based on equation (19), it can also be seen that the estimated value for the Restricted Maximum Likelihood Estimator (RMLE) is as follows (Duffy dan Santner, 1989):

$$\boldsymbol{\hat{\beta}_{RMLE}} = \boldsymbol{\hat{\beta}_{MLE}} + \boldsymbol{C^{-1}S^T(SC^{-1}S^T)^{-1}\big(h - S\hat{\beta}_{MLE}\big)}.$$

Based on the $\boldsymbol{\hat{\beta}_{RMLE}}$ equation, a new parameter estimator method is obtained, namely the Stochastic Restricted Maximum Likelihood Estimator (SRMLE) by adding the $\mathbf{\Psi}$ identity matrix. The SRMLE parameter estimate is obtained if equation (10) is available in addition to the existing logistic regression model and $\mathbf{\Psi}$ is assumed to be an identity matrix ($p \times p$), so that the model of the parameter estimator can be written in the following equation (Nagarajah and Wijekoon, 2015):

$$\boldsymbol{\hat{\beta}_{SRMLE}} = \boldsymbol{\hat{\beta}_{MLE}} + \mathbf{C^{-1}S^T(\Psi + SC^{-1}S^T)^{-1}}\big(\boldsymbol{h} - \mathbf{S}\boldsymbol{\hat{\beta}_{MLE}}\big). \tag{11}$$

The multicollinearity in the independent variable $X$ causes a small characteristic root (eigenvalue) in the $\mathbf{X^T X}$ matrix (Montmogery and Peck, 1992), so that to estimate equation (20) which assumes multicollinearity in variable X the β value in equation (9) is taken of the eigenvectors with the smallest eigenvalues of the $\mathbf{X^T X}$ matrix. The asymptotic covariance matrix of SRMLE is (Nagarajah and Wijekoon, 2015):

$$\boldsymbol{Var}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{SRMLE}}) = \mathbf{C^{-1}} - \mathbf{C^{-1}S^T}(\boldsymbol{\Psi} + \mathbf{SC^{-1}S^T})^{-1}\mathbf{SC^{-1}} \qquad (12)$$

## 4. Data Collection

The research data used in this study is secondary data regarding the percentage of the poverty index of each province in Indonesia. Much of the data used according to many provinces in Indonesia with 4 factors that influence the percentage of poverty in Indonesia is presented in Table 1.

Table 1. Poverty Data in Indonesia

| i | Nama Provinsi di Indonesia | Persentase Pengangguran | Persentase Setengah Pengangguran | Indeks Kedalaman Kemiskinan | Indeks Keparahan Kemiskinan | Persentase Penduduk Miskin | Y |
|---|---|---|---|---|---|---|---|
| | | X1 | X2 | X3 | X4 | | |
| 1 | Aceh | 6,36 | 11,37 | 2,8 | 0,72 | 15,68 | 1 |
| 2 | Sumatera Utara | 5,56 | 8,38 | 1,46 | 0,33 | 8,94 | 0 |
| 3 | Sumatera Barat | 5,55 | 9,62 | 0,96 | 0,21 | 6,55 | 0 |
| 4 | Riau | 6,20 | 8,68 | 1,05 | 0,24 | 7,21 | 0 |
| 5 | Jambi | 3,86 | 7,61 | 1,26 | 0,31 | 7,85 | 0 |
| 6 | Sumatera Selatan | 4,23 | 8,31 | 2,06 | 0,5 | 12,82 | 1 |
| 7 | Bengkulu | 3,51 | 8,5 | 2,35 | 0,51 | 15,41 | 1 |
| 8 | Lampung | 4,06 | 9,35 | 2,06 | 0,48 | 13,01 | 1 |
| 9 | Kep. Bangka Belitung | 3,65 | 5,96 | 0,73 | 0,15 | 4,77 | 0 |
| 10 | Kep. Riau | 7,12 | 3,87 | 0,59 | 0,11 | 5,83 | 0 |
| 11 | Dki Jakarta | 6,24 | 1,97 | 0,5 | 0,11 | 3,55 | 0 |
| 12 | Jawa Barat | 8,17 | 5,38 | 1,13 | 0,26 | 7,25 | 0 |
| 13 | Jawa Tengah | 4,51 | 5,21 | 1,63 | 0,34 | 11,19 | 1 |
| 14 | Di Yogyakarta | 3,35 | 4,26 | 1,65 | 0,35 | 11,81 | 1 |
| 15 | Jawa Timur | 3,99 | 5,83 | 2,07 | 0,56 | 10,85 | 1 |
| 16 | Banten | 8,52 | 4,23 | 0,91 | 0,25 | 5,25 | 0 |
| 17 | Bali | 1,37 | 2,36 | 0,52 | 0,11 | 3,91 | 0 |
| 18 | Nusa Tenggara Barat | 3,72 | 16,95 | 2,38 | 0,55 | 14,63 | 1 |
| 19 | Nusa Tenggara Timur | 3,01 | 12,4 | 4,55 | 1,44 | 21,03 | 1 |
| 20 | Kalimantan Barat | 4,26 | 9,23 | 1,21 | 0,28 | 7,37 | 0 |
| 21 | Kalimantan Tengah | 4,01 | 7,24 | 0,82 | 0,2 | 5,1 | 0 |
| 22 | Kalimantan Selatan | 4,50 | 5,95 | 0,75 | 0,18 | 4,65 | 0 |
| 23 | Kalimantan Timur | 6,60 | 4,22 | 0,76 | 0,15 | 6,06 | 0 |
| 24 | Kalimantan Utara | 5,22 | 7,7 | 0,91 | 0,17 | 6,86 | 0 |
| 25 | Sulawesi Utara | 6,86 | 8,25 | 1,31 | 0,3 | 7,59 | 0 |
| 26 | Sulawesi Tengah | 3,43 | 10,48 | 2,28 | 0,68 | 13,69 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 27 | Sulawesi Selatan | 5,34 | 7,57 | 1,68 | 0,51 | 8,87 | 0 |
| 28 | Sulawesi Tenggara | 3,26 | 11,16 | 2,09 | 0,55 | 11,32 | 1 |
| 29 | Gorontalo | 4,03 | 9,32 | 3,02 | 0,83 | 15,83 | 1 |
| 30 | Sulawesi Barat | 3,16 | 8,67 | 1,56 | 0,35 | 11,22 | 1 |
| 31 | Maluku | 7,27 | 9,36 | 3,31 | 0,92 | 17,85 | 1 |
| 32 | Maluku Utara | 4,77 | 7,78 | 1,25 | 0,39 | 6,62 | 0 |
| 33 | Papua Barat | 6,30 | 8,56 | 6,5 | 2,38 | 22,66 | 1 |
| 34 | Papua | 3,20 | 7,83 | 5,91 | 1,82 | 27,43 | 1 |
| | **Indonesia** | **5,34** | **6,62** | **1,63** | **0,41** | **9,66** | |

To perform a multicollinearity test on poverty data in Indonesia, VIF is based on equation (8), before that, the coefficient of determination on each variable used must be sought first by using equation (9). In the process of calculating the coefficient of determination and the VIF value, the writer used IBM SPSS 24 software. The results of the calculation of VIF values for each variable can be seen in Table 2.

Table 2 Multicollinearity Test on Poverty Data in Indonesia

| Variable Name | Value of the coefficient of determination | VIF | Description |
|---|---|---|---|
| PE | 0,088 | 1,097 | There is no Multicollinearity |
| SPE | 0,335 | 1,504 | There is no Multicollinearity |
| IDK | 0,979 | 47,900 | There is Multicollinearity |
| IDP | 0,978 | 44,702 | There is Multicollinearity |

A variable is said to have multicollinearity if the VIF value is $\geq 10$. Based on Table 2, there are variables that have a VIF $\geq 10$ value, so that the data contains multicollinearity. Thus, the data can be used to be modeled by estimating the logistic regression model using the SRMLE method.

## 5. Results and Discussion

In determining the logistic regression model from a data set, it is necessary to first find the estimated parameter values of the model. Because the data used is multicollinearity, so to estimate the parameters you can use the SRMLE method using equation (10). To estimate the SRMLE parameter, we must first find the MLE estimator with equation (9).

### 5.1 Numerical Results

The matrix used to find eigenvalues and eigenvectors is the $\mathbf{X^T X}$ matrix where the **X** matrix is a data matrix containing the independent variables. Where the $\mathbf{X^T X}$ matrix is as follows:

$$X^T X = \begin{bmatrix} 34 & 165,19 & 263,56 & 64,02 & 17,24 \\ 165,19 & 891,7699 & 1246,2899 & 300,4873 & 81,2673 \\ 263,56 & 1246,2899 & 2335,7648 & 558,3078 & 152,0347 \\ 64,02 & 300,4873 & 558,3078 & 186,5910 & 55,1868 \\ 17,24 & 81,2673 & 152,0347 & 55,1868 & 16,7862 \end{bmatrix}$$

Based on equation (10), the eigenvalues and eigenvectors can be determined in the $\mathbf{X^T X}$ matrix presented in Table 3.

Table 3 Eigenvalues and Matrix Eigenvectors $\mathbf{X^T X}$

| Eigenvalues | Eigenvectors | | | | |
|---|---|---|---|---|---|
| 0,1572 | -0,1270 | 0,1055 | -0,1037 | 0,3341 | -0,9338 |
| 1,5918 | -0,9845 | 0,1074 | 0,5325 | -0,3138 | 0,1233 |
| 56,1695 | -0,1456 | -0,3166 | 0,2615 | -0,9073 | -0,3276 |
| 177,2844 | -0,6819 | -0,8674 | 0,4675 | 0,1481 | 0,4726 |
| 3229,7087 | 0,9896 | 0,4845 | 0,8426 | 0,2055 | 0,5619 |

Based on Table 3, it can be seen that the smallest eigenvalue is 0.1572, so that the eigenvector of the eigenvalue 0.1572 will be the β value to find the estimated MLE parameter according to equation (8). Before estimating parameters using the MLE method, the matrix **U, z,** and **C** must be determined where to determine the matrix the authors use the following equation:

$$\mathbf{U} = \begin{bmatrix} \pi(x_1)\big(1-\pi(x_1)\big) & 0 & \cdots & 0 \\ 0 & \pi(x_2)\big(1-\pi(x_2)\big) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \pi(x_n)\big(1-\pi(x_n)\big) \end{bmatrix},$$

$$\boldsymbol{z} = \mathrm{logit}(\pi(x_i)) + \frac{y_i - \pi(x_i)}{\pi(x_i)(1-\pi(x_i))},$$

where $\mathbf{C} = \mathbf{X^T U X}$ and $i = 1,2, \ldots n$. Using the Python software, a matrix can be determined **U**, **z** and **C** as follows:

$$\mathbf{U} = \begin{bmatrix} 0{,}2495460031 & 0 & \cdots & 0 \\ 0 & 0{,}2499629896 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0{,}2493697280 \end{bmatrix}$$

$$\mathbf{z} = \begin{bmatrix} 2.003535174 \\ -2.000298529 \\ \vdots \\ 2.004885628 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 8{,}5000 & 41{,}2975 & 65{,}8900 & 16{,}0050 & 4{,}3100 \\ 41{,}2975 & 222{,}9424 & 311{,}5724 & 75{,}1218 & 20{,}3168 \\ 65{,}8900 & 311{,}5724 & 583{,}9412 & 139{,}5769 & 38{,}0086 \\ 16{,}0050 & 75{,}1218 & 139{,}5769 & 46{,}6477 & 13{,}7967 \\ 4{,}3100 & 20{,}3168 & 38{,}0086 & 13{,}7967 & 4{,}1965 \end{bmatrix}$$

Based on the matrix above, the parameter estimator of the MLE method can be determined based on equation (8) given in Table 4.

Table 4 Estimation Results of the MLE Method Based on Equation (8)

| Variable | Estimated Value Regression Coefficient |
|---|---|
| Intercept | -2.2845 |
| PE | -0,2166 |
| SPE | -0.1836 |
| IDK | 4.7623 |
| IDP | -11.0530 |

Based on Table 4, the logistic regression equation with the MLE estimator is obtained, namely:

$$\pi(x) = \frac{\exp(-2{,}2845 - 0{,}2166x_1 - 0{,}1836x_3 + 4{,}7623x_3 - 11{,}053x_4)}{1 + \exp(-2{,}2845 - 0{,}2166x_1 - 0{,}1836x_3 + 4{,}7623x_3 - 11{,}053x_4)}$$

After obtaining the MLE estimate with equation (9), then the SRMLE estimator method is searched based on equation (11). Before looking for parameter estimates using the SRMLE method, the matrix Ψ, S and the vector h must be determined as obtained from equation (10), The estimation results obtained with the SRMLE method are in Table 5.

Table 5 Results of the SRMLE Method Assessment

| Variable | Estimated Value Regression Coefficient |
|---|---|
| Intercept | -0,6527 |
| PE | -0,3514 |
| SPE | 0.9544 |
| IDK | 0,9490 |
| IDP | -0,5201 |

Based on Table 5, the logistic regression equation with the SRMLE estimator is obtained, namely:

$$\pi(x) = \frac{\exp(-0{,}6527 - 0{,}3514x_1 + 0{,}9544x_2 + 0{,}949x_3 - 0{,}5201x_4)}{1 + \exp(-0{,}6527 - 0{,}3514x_1 + 0{,}9544x_2 + 0{,}949x_3 - 0{,}5201x_4)}$$

### 5.2 Proposed Improvements

The proposal for written improvement here is that the data used must be validated based on field surveys, including numerical results that must be analyzed more sharply, and also need additional graphical presentation as material for analysis of the results of data processing carried out. This method will also be compared with other methods, in order to obtain even more accurate results. Finally, I hope the results of this research can be developed to get more satisfying results.

### 5.3 Validation

The next step is to test the significance of the parameters individually using the Wald test to find out which variables have an influence on the regression model. Each variable is examined by comparing the W value obtained using equation (2.27) with the Chi-Square distribution where α = 0.05 and db = 1, obtained $\chi^2_{(0{,}05{,}1)}$= 3.84146. The standard error (SE) is obtained from the main diagonal of the covariance matrix.
Based on the covariance matrix above, the standard error value of each variable can be determined, where the standard error value is shown in Table 6.

Table 6 Value of Standard Error for PE and IDK Variables

| Variable | Standard Error |
|---|---|
| Intercept | 0,7206623433 |
| PE | 0,1393839103 |
| IDK | 0,2269032894 |

Based on Table 6 it can be determined that the Wald value for each variable is to obtain the following values:
Based on the hypothesis:
$H_0$:$\beta_3 = 0$ (variable $X_3$ does not make a significant contribution to π (*x*)).
$H_1$:$\beta_3 \neq 0$ (variable $X_3$ makes a significant contribution to π (x)).

$$W_3 = \left(\frac{\hat{\beta}_3}{\mathrm{SE}(\hat{\beta}_3)}\right)^2 = 13.13783$$

Then $H_0$is rejected because the value of $W_3$= 13.13783 is greater than $\chi^2_{(0{,}05{,}1)}$= 3.84146, meaning that the Poverty Depth Index variable has a significant effect on the binary logistic regression model. Based on the results of statistical testing, the results show that the significant predictor variables for the binary logistic regression model are the Unemployment variable and the Poverty Depth Index. The logistic regression models that are significant in the poverty level are as follows:

$$\hat{\pi}(x) = \frac{\exp(0{,}3635 - 0{,}4083x_1 + 0{,}8224x_3)}{1 + \exp(0{,}3635 - 0{,}4083x_1 + 0{,}8224x_3)}$$

In linear regression, the coefficient $\hat{\beta}_j$ shows the change in the value of the response variable as a result of a one-unit change in the predictor variable. In binary logistic regression, the coefficient is difficult to interpret directly, therefore the odds ratio is used to interpret the coefficient. The odds ratio can be seen in Table 7.

Table 7 Odds Ratio for PE and IDK Variables

| Variable | $\hat{\beta}$ | Exp($\hat{\beta}$) |
|---|---|---|
| Intercept | 0,3635 | 1,438 |
| PE | −0,4083 | 0,665 |
| IDK | 0,8224 | 2,276 |

Significant predictor variables, namely the unemployment variable (PE) and the Poverty Depth Index (IDK) variable, both of these data are continuous data types so that in implementing the data must be in the form of one unit. If the percentage of unemployment in a province increases by one unit, the chance for the province to be closer to poverty is 0.665 times, whereas if the Poverty Depth Index of a province increases by one unit, then the chance for the province to be closer to poverty is 2.276 times

## 6. Conclusion

The poverty rate of each province in Indonesia on average every year can be divided into two, namely the percentage above or equal to the percentage of Indonesia which is coded 1 and the percentage below the percentage of Indonesia which is coded 0 so that it can be modeled using binary logistic regression analysis. The data used contains multicollinearity so that the SRMLE method can be used as a parameter estimator. By performing the likelihood ratio test and Wald test of the SRMLE model parameters with a significant level of 0.05. Based on the logistic regression model with the Stochastic Restricted Maximum Likelihood (SRMLE) parameter estimator, it can be determined the factors that affect the poverty rate in Indonesia each year, namely the unemployment variable and the Poverty Depth Index.

## References

Anton, H. dan Rorres, C. (2004). Aljabar Linear Elementer Versi Aplikasi Edisi Kedelapan Jilid 1. Jakarta: Erlangga.

Cakmakyapan, S. dan Goktas, A. (2013). A Comparison of Binary Logit and Probit Models with A Simulation Study. *Journal of Social and Economics Statistics*, Vol.2, No.1, Summer.

Duffy, D.E. dan Santner T.J. (1989). On the Small Sample Properties of Norm-Restricted Maximum Likelihood Estimators for Logistic Regression Models Vol 18 No 3 Hal. 959-980. *Communications in Statistics-Theory and Method*.

Hosmer DW dan Lemeshow S. (2000). *Applied Logistic Regression. Edition 2*. New York (US).

Hosmer, D.W., Lemeshow, S., dan Sturdivant, (2013). *Applied Logistic Regression (3rd ed).* New York: John Wiley & Sons, Inc.

Kutner, M.H., C.J. Nachtsheim., dan J. Neter. (2004). *Applied Linear Regression Models.* 4th ed. New York: McGraw-Hill Companies, Inc.

Megasari, H, Samsul A dan Idris. (2015). Analisis Perekonomian Dan Kemiskinan di Indonesia. *Jurnal Kajian Ekonomi.* Vol 3 No 6. Universitas Negri Padang.

Montgomery, D.C & Peck. (1992). Introduction to linear Regression Analysis. New York. John Wiley & sons.

Nagarajah, V dan Wijekoon, P. 2015. Stochastic Restricted Maximum Likelihood Estimator in Logistic Regression Model. *Open Journal of Statistics*. Vol 5, 837-851. University of Peradeniya dan University of Jaffna, Sri Lanka.

Siray, G.U, Toker, S. dan Kaciranlar, S. (2015). On the Restricted Liu Estimator in Logistic Regression Model. *Communications in statistics-Simulation and Computation,* 44. Vol 18 No 16 Hal 37-41. Çukurova University, Turkey.

Sukirno, S. (2004). *Makro Ekonomi Teori Pengantar*. Jakarta: Raja Grafindo Persada.

Sumodiningrat. 1994. *Ekonometrika Pengantar*. Yogyakarta: BPFE.

Tampil, Y.A. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *Journal of Conservative Dentistry,* Vol 6 No 2. Universitas Sam Ratulangi Manado.

Williams, R. (2015). *Multicolinearity*. University of Notre Dame, Australia.

Badan Pusat Statistik (BPS). (2019). Kemiskinan dan Ketimpangan. diakses dari http://www.bps.go.id/ , diakses pada tanggal 20 Desember 2019 pada pukul 20.00 WIB.

Cosma S. *Statistics 36-350: Data Mining*, Lecture Notes, 18 November 2009, http://www.stat.cmu.edu/~cshalizi/350/lectures/26/lecture-26.pdf

Srihari, SN. *Machine Learning: Iterative Reweighted Least Squares*. https://cedar.buffalo.edu/~srihari/CSE574/Chap4/4.3.3-IRLS.pdf, State University of New York USA diakses 15 Maret 2020.

**Biography (12 font)**

**Riaman** is a teaching staff in the mathematics study program at the Mathematics Department of the University of Pajajaran since 1997. The mathematics degree was won from the Mathematics Department of the University of Padjadjaran in 1995, while the Masters of Actuarial Science was obtained from the Bandung Institute of Technology in 2000. The field of research currently being undertaken is the field of applied mathematics. More specifically are: fields of financial mathematics, actuarial mathematics, survival models, mathematical statistics, and reliability. Since 2001 until now he has joined as a member of IndoMS, and since 2017 joined IAENG.

**Sukono** is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Currently serves as Head of Master's Program in Mathematics, the field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

**Abdul Talib Bon** is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He’s bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.