priviet lab.
RESEARCH & PUBLISHING

Priviet Social Sciences Journal

Volume 6
Issue 4 April, 2026

Article 20

07-04-2026

SAR probit regression modeling using the fisher scoring approach: A case study of poverty levels on the island of Sumatra

Panji Jiblathar, Germansah, Fadhlul Mubarak, Ellys Agustina, Nabilla Rida Tri Nisa

To cite this article: Jiblathar, P., Germansah, G., Mubarak, F., Agustina, E., & Nisa, N. R. T. (2026). SAR probit regression modeling using the fisher scoring approach: A case study of poverty levels on the island of Sumatra. *Priviet Social Sciences Journal, 6*(4), 241-256. https://doi.org/10.55942/pssj.v6i4.1727

To link to this article: https://doi.org/10.55942/pssj.v6i4.1727

Private Social Sciences Journal
E-ISSN 2798-6314
P-ISSN 2798-866X
https://doi.org/10.55942/pssj.v6i4.1727

Volume 6
Issue 4
April Edition 2026

SAR probit regression modeling using the fisher scoring approach: A case study of poverty levels on the island of Sumatra

Panji Jiblathar[1*], Germansah[1], Fadhlul Mubarak[1], Ellys Agustina[2], & Nabilla Rida Tri Nisa[3]

[1]UIN Sulthan Thaha Saifuddin Jambi, Jl. Jambi Ma. Bulian KM.16 Simpang Sungai Duren Kab. Muaro Jambi, Jambi 36361, Indonesia
[2]Dinas Badan Pengelola Keuangan dan Pendapatan Daerah Kabupaten Kerinci, Jl. Ke Bukit Tengah, Mukai Tinggi, Kecamatan Siulak, Kabupaten Kerinci, Provinsi Jambi 37162, Indonesia
[3]Politeknik Negeri Media, Jl. Srengseng Sawah No.17, RT.8 RW.3, Srengseng Sawah, Kec. Jagakarsa, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12630, Indonesia
*e-mail: panjijiblathar@uinjambi.ac.id

Received 02 February 2026
Revised 10 March 2026
Accepted 03 April 2026

ABSTRACT

The probit model is commonly used to study categorical response data. However, failing to account for spatial autocorrelation factors between regions can lead to inconsistent and biased parameter estimation results. This study focuses on examining the parameter estimation of the Spatial Autoregressive (SAR) Probit model through the Maximum Likelihood Estimation (MLE) method using the Fisher Scoring numerical iteration scheme. The model was implemented on poverty data across 131 regencies/cities in the Sumatra region for 2022. Empirical findings indicate that the GRDP growth rate, open unemployment rate, per capita expenditure, and expected years of schooling significantly affect the poverty rate. The results of this model development provide a prediction accuracy of 83.97%. This achievement is superior to that of the RIS Simulator technique, which only yields an accuracy of 74.05%. These results emphasize the advantages of the efficiency and accuracy of the Fisher Scoring approach in representing spatial dependencies in the poverty phenomenon in Sumatra.

Keywords: fisher scoring; MLE; poverty; RIS simulator; spatial probit

1. INTRODUCTION

The initial foundation of probit regression was introduced by Bliss in 1934, offering a nonlinear analytical framework to connect predictors with binary response variables. It is a nonlinear analytical instrument for evaluating the relationship between predictor variables (numerical, categorical, or mixed) and a binary categorical response variable. In this model, a value of 1 represents the existence of a specific attribute, whereas a value of 0 indicates the absence of that attribute. Empirically, categorical response data in this model often exhibit the influence of adjacent regions, known as spatial autocorrelation (Anselin, 1988). Spatial autocorrelation refers to a phenomenon in which observations at a specific location tend to depend on or are influenced by the values in the surrounding geographical areas. If this geographical relationship is ignored in a standard probit model specification, the resulting parameters will be both biased and inconsistent. This bias phenomenon refers to the estimator's inability to approach the true population parameter value owing to unaccommodated distortions in spatial patterns.

The main challenge in SAR Probit modeling lies in its likelihood function, which does not have a closed-form solution, thus requiring a numerical optimization approach. Standard approaches, such as the Newton-Raphson algorithm, iteratively use the Hessian matrix (second derivative), which often experiences convergence failure or matrix instability when dealing with high-dimensional spatial data. To overcome this limitation, this study applies the Fisher Scoring numerical iteration. This algorithm modifies the Newton-Raphson approach by substituting the Hessian matrix with the Fisher information matrix. Mathematically, this substitution guarantees that the matrix is always positive definite, resulting in more stable convergence and more consistent parameter estimates, particularly when handling regional poverty data with high spatial variations (McMillen, 1992).

Parameter estimation plays a crucial role in obtaining unbiased estimated values and determining the covariance matrix of the model. In the spatial probit regression framework, this parameter estimation can be implemented through the Maximum Likelihood Estimation (MLE) method integrated with the Fisher scoring numerical iteration scheme. This iterative procedure is essential because the equation function in the model does not have an analytical solution or is non-closed when estimated with MLE. The Fisher scoring technique was developed as a variant of the Newton-Raphson method and is specifically designed to find numerical solutions for complex maximum likelihood functions. Unlike the Newton-Raphson algorithm, which utilizes a second-derivative-based Hessian matrix and often faces constraints in achieving convergence, Fisher scoring replaces this component with the Fisher information matrix to ensure the stability of the estimation process (Schworer & Hovey, 2004).

Therefore, this study aims to analyze parameter estimation in the Spatial Autoregressive (SAR) probit model using the Fisher Scoring algorithm on the 2022 Sumatra poverty data. The main contributions of this study are twofold. First, methodologically, it demonstrates the efficiency and stability of the Fisher Scoring algorithm in handling complex spatial dependencies compared with the Recursive Importance Sampling (RIS) simulator. Second, empirically, it provides novel insights into the spatial patterns of poverty in Sumatra and offers targeted policy recommendations. By proving that poverty in a region is significantly influenced by neighboring areas, this study suggests that poverty alleviation strategies in Sumatra must be conducted through synchronized cross-regional policies rather than isolated local efforts.

2. THEORETICAL FRAMEWORK

Before delving into the mathematical formulations, it is essential to understand the conceptual transitions that build the Spatial Probit model. Standard probit regression is used to evaluate binary outcomes, such as a region having either 'high' or 'low' poverty based on various predictors. However, socio-economic phenomena such as poverty are rarely isolated; a region's economic status often overflows into neighboring areas. This geographical relationship was captured by spatial regression. By integrating these two concepts, spatial probit regression is formed. This advanced framework not only models the binary categorical response but also intuitively accommodates the spatial influence of adjacent regions, ensuring that the estimated parameters remain unbiased and consistent across the study area.

2.1. Probit Regression

Referring to the thought of (Greene, 2008), the binomial probit regression framework begins with the formation of a latent variable y_i^* formulated through the following linear relationship:

$$y_i^* = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{\mathbf{i}} + \epsilon_i \quad (1)$$

Here, y_i^* is a latent response variable influenced by a vector of predictor variables $\mathbf{x}_{\mathbf{i}}$ and parameter coefficients $\boldsymbol{\beta}$. The error component ϵ_i is assumed to follow a normal distribution with a mean of zero and a variance of σ^2. The probability density function (PDF) of the variable y_i^* is as follows (Greene, 2008):

$$f(y_i^*) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i^* - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)^2\right) \quad (2)$$

The determination of the response variable category y_i is based on a threshold value δ, where a value of 1 is assigned if the latent variable exceeds that threshold. Based on this assumption, the probability of an event $(y_i = 1)$ occurring can be calculated using the standard normal cumulative distribution function Φ as follows:

$$\begin{aligned}
\mathrm{P}(y_i = 0) &= \mathrm{P}(y_i^* < \delta) \\
&= \mathrm{P}(\boldsymbol{\beta}^T\mathbf{x}_i + \epsilon_i < \delta) \\
&= P(\varepsilon_i < \delta - \boldsymbol{\beta}^T\mathbf{x}_i) \\
&= \Phi(\delta - \boldsymbol{\beta}^T\mathbf{x}_i) \quad (3)\\
\mathrm{P}(y_i = 1) &= \mathrm{P}(y_i^* \geq \delta) \\
&= 1 - \mathrm{P}(y_1^* < \delta) \\
&= 1 - \mathrm{P}(\boldsymbol{\beta}^T\mathbf{x}_i + \epsilon_i < \delta) \\
&= 1 - \mathrm{P}(\varepsilon_i < \delta - \boldsymbol{\beta}^T\mathbf{x}_i) \\
&= 1 - \Phi(\delta - \boldsymbol{\beta}^T\mathbf{x}_i) \quad (4)
\end{aligned}$$

2.2. Spatial Regression

The spatial regression approach for cross-sectional data was pioneered by Anselin (1988), one of which is through the Spatial Autoregressive (SAR) structure. For categorical response variables, the SAR model can be expressed as:

$$\boldsymbol{y}^* = \rho\mathbf{W}\boldsymbol{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (5)$$

Equation (5) involves the parameter ρ, which indicates the spatial lag coefficient, as well as the weight matrix $\mathbf{W}$, which represents the inter-regional interactions determined through the Queen Contiguity method. Queen Contiguity, or edge-and-corner contiguity, defines $w_{\{ii^*\}} = 1$ for entities that share an edge or a vertex with the region of interest, and $w_{ii^*} = 0$ for all other regions. For region (3), the values obtained are $w_{32} = 1$, $w_{34} = 1$, $w_{35} = 1$, while the remaining values are equal to zero. Figure 1 illustrates an example of five regions as they appear on a map.

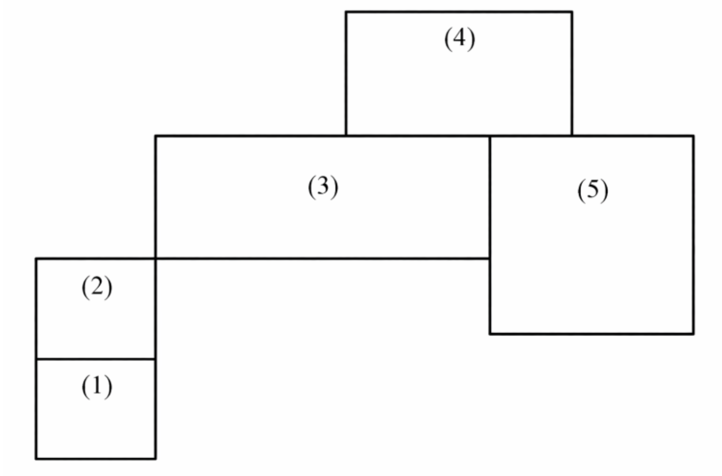

Figure 1. Illustration of Queen Contiguity

The selection of this edge-and-corner contiguity weight (Queen) is based on the geographical characteristics of the administrative boundaries of the districts/cities on Sumatra Island, which form highly irregular polygons. Consequently, even the slightest intersection at the boundary corners is accounted for as a spatial neighbor that potentially exerts a poverty spillover effect. It is important to note that the diagonal elements of matrix $\mathbf{W}$ must be zero to prevent self-correlation.

In spatial analysis, dependence between locations generally appears in two main formats: spatial lag and spatial errors. The spatial lag phenomenon indicates that the conditions in a region are directly influenced by its neighboring regions. Spatial error refers to the existence of inter-regional linkages that are not only limited to the response variable but also include disturbances in predictor variables (Anselin, 1988). To validate the presence of spatial lag interaction effects in the model, researchers can apply the Lagrange Multiplier (LM) test. This procedure is carried out by testing the significance of the parameter ρ through the following hypotheses:

$\mathbf{H_0}: \rho = 0$ (Assumption of no spatial lag dependence)

$\mathbf{H_1}: \rho \neq 0$ (Indication of spatial lag dependence)

The test statistic used is as shown in Equation (6):

$$LM = E^{-1}\left\{\left(R_y\right)^2 Tr_{22} - 2R_y R_e Tr_{12} + (R_e)^2(D + Tr_{11})\right\} \sim \chi^2_{(m)} \tag{6}$$

where :

$R_y = \boldsymbol{e}^{\mathbf{T}}\boldsymbol{W_1 y}/\sigma^2$

$R_e = \boldsymbol{e}^{\mathbf{T}}\boldsymbol{W_2 e}/\sigma^2$

$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$

$\mathrm{Tr}_{kk^*} = tr\{\mathbf{W}_k\mathbf{W}_{k^*} + \mathbf{W}_k^T\mathbf{W}_{k^*}\}$

$\mathrm{D} = \sigma^{-2}(\mathbf{W_1X\beta})^T\mathbf{M}(\mathbf{W_1X\beta})$

$E = (D + T_{11})T_{22} - (T_{12})^2$

The null hypothesis will be rejected if the LM statistic exceeds the critical value on the χ^2 distribution with m degrees of freedom. This condition indicates the influence of spatial lag dependence on the model. Specifically for the SAR model structure, the value of m is set to one because it only uses the coefficient ρ as a single spatial parameter.

2.3. Spatial Probit Regression

The synergy between the probit regression framework and spatial analysis methodology yields a spatial probit model. The development of this model is rooted in the research of McMillen (1992), who proposed the use of the EM algorithm to produce consistent parameters under autocorrelated conditions. In the Spatial Autoregressive (SAR) scheme, the specification of the spatial probit model can generally be expressed as follows:

$$\boldsymbol{y}^* = \rho \mathbf{W} \boldsymbol{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{7}$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n})$. In this context, the vector y^* represents the latent response variable, while inter-regional interactions are captured through the spatial parameter ρ and the weight matrix $\mathbf{W}$. For ease of estimation, this model can be transformed into a reduced form, as follows:

$$\boldsymbol{y}^* = (\mathbf{I} - \rho \mathbf{W})^{-\mathbf{1}} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-\mathbf{1}} \boldsymbol{\epsilon} \tag{8}$$

let $(\mathbf{I} - \rho \mathbf{W})^{-\mathbf{1}} = \boldsymbol{\Lambda}$ and $(\mathbf{I} - \rho \mathbf{W})^{-\mathbf{1}} \boldsymbol{\epsilon} = \boldsymbol{u}$, so that $\boldsymbol{y}^*$ forms the equation.

$$\begin{gathered} \boldsymbol{y}^* = \boldsymbol{\Lambda}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}, \\ \boldsymbol{u} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{T}) \end{gathered} \tag{9}$$

The variable y_i^* has a binary category defined as the variable y_i through the following equation:

$$y_i = \begin{cases} 0, & \text{jika } y_i^* < \delta \\ 1, & \text{jika } y_i^* \geq \delta \end{cases} \tag{10}$$

Then, the probability from Equation (10) can be written as follows:

$$\begin{aligned} \mathrm{P}(y_i = 0) &= \mathrm{P}(y_1^* < \delta) \\ &= \Phi\left(\frac{\delta - [\Lambda \mathrm{X}\beta]_\mathrm{i}}{\sqrt{\varsigma_{ii}}}\right) \end{aligned} \tag{11}$$

$$\begin{aligned} \mathrm{P}(y_i = 1) &= \mathrm{P}(y_i^* \geq \delta) \\ &= 1 - \Phi\left(\frac{\delta - [\Lambda \mathrm{X}\beta]_\mathrm{i}}{\sqrt{\varsigma_{ii}}}\right) \end{aligned} \tag{12}$$

From equations (11) and (12), the variable y_i will be obtained:

$$y_i = \begin{cases} 0, & \text{jika } u_i < \dfrac{\delta - [\Lambda \mathrm{X}\beta]_i}{\sqrt{\varsigma_{ii}}} \\ 1, & \text{jika } u_i \geq \dfrac{\delta - [\Gamma \mathrm{X}\beta]_i}{\sqrt{\varsigma_{ii}}} \end{cases} \tag{13}$$

The variable Y_i in equation (13) is Bernoulli distributed with the following probability function:

$$\mathrm{P}(Y_i = y_i) = \begin{cases} p(\boldsymbol{x_i})^{y_i}(1 - p(\boldsymbol{x_i}))^{1-y_i}, & y = 0{,}1; 0 < p(\boldsymbol{x_i}) < 1. \\ 0, & \text{lainnya.} \end{cases} \tag{14}$$

The mean and variance of Y_i are as follows:

$$\begin{gathered} E(Y_i) = p(\boldsymbol{x_i}), \\ Var(Y_i) = p(\boldsymbol{x_i})\big(1 - p(\boldsymbol{x_i})\big) \end{gathered} \tag{15}$$

where $p(x_i)$ is the probability of success of $P(y_i = 1)$. Since the variable $\mathbf{u}$ is assumed to have a multivariate normal distribution with a mean of 0 and a variance of $\mathbf{T}$, the probabilities for $P(y_i = 1)$ and $P(y_i = 0)$ based on equations (11) and (12) are as follows:

$$\mathrm{P}(y_i = 0|\mathbf{x}_i) = \left(1 - p(\boldsymbol{x_i})\right) = \Phi\left(\frac{\delta - [\Lambda \mathrm{X}\beta]_i}{\sqrt{\varsigma_{ii}}}\right) \tag{16}$$

$$\mathrm{P}(y_i = 1|\mathbf{x}_i) = p(\boldsymbol{x_i}) = 1 - \Phi\left(\frac{\delta - [\Lambda \mathrm{X}\beta]_i}{\sqrt{\varsigma_{ii}}}\right) \tag{17}$$

2.4. Fisher Scoring Parameter Estimation

In cases where the Maximum Likelihood Estimation (MLE) yields non-closed-form equations, an iterative procedure is required to find numerical solutions. One effective technique is the Fisher scoring algorithm, which modifies the Newton-Raphson method. According to Schworer and Hovey (2004), the iteration structure in this method is formulated to update parameter estimates until a certain convergence point is reached. The stability of this algorithm is guaranteed by the use of the Fisher information matrix in its updating process, which is mathematically expressed as

$$\boldsymbol{\beta}^{*(t+1)} = \boldsymbol{\beta}^{*(\mathbf{t})} + \left(\mathfrak{J}(\boldsymbol{\beta}^{*(\mathbf{t})})\right)^{-1} \boldsymbol{S}(\boldsymbol{\beta}^{*(\mathbf{t})})$$

where $\boldsymbol{\beta}^*$ is the vector of model parameter estimators, namely $\boldsymbol{\beta}^* = [\boldsymbol{\beta}, \rho]^T$ when applied to the spatial probit regression model. t is the iteration level. $\mathfrak{J}(\boldsymbol{\beta}^*)$ is the Fisher information matrix of $\boldsymbol{\beta}^*$, namely:

$$\mathfrak{J}(\boldsymbol{\beta}, \rho) = -E\left[\frac{\partial^2\left(l(\boldsymbol{\beta}, \rho)\right)}{\partial(\boldsymbol{\beta}, \rho)\partial(\boldsymbol{\beta}, \rho)^T}\right]$$

Meanwhile, $\boldsymbol{S}(\boldsymbol{\beta}^*)$ is the gradient vector to be estimated, having the following form:

$$\boldsymbol{S}(\boldsymbol{\beta}, \rho) = \frac{\partial\left(l(\boldsymbol{\beta}, \rho)\right)}{\partial(\boldsymbol{\beta}, \rho)}$$

The Fisher scoring iteration process stops if the convergence condition is met, namely $\left\|\boldsymbol{\beta}^{*(t+1)} - \boldsymbol{\beta}^{*(t)}\right\| < \Delta$, where Δ is a verll positive real value. If the criteria are met, the iteration process stops at the T-th iteration, so the parameter estimator is $\widehat{\boldsymbol{\beta}^*} = \boldsymbol{\beta}^{*(\boldsymbol{T})}$. After obtaining the value of the parameter estimator, parameter testing on the spatial probit model can be continued.

2.5. Simultaneous Parameter Testing

The significance of all predictor variables in the spatial probit model was evaluated using the Maximum Likelihood Ratio Test (MLRT). The operational basis of this procedure is to contrast the capability of the model that has been equipped with predictors against a model without independent variables (under the null hypothesis). Statistical decisions are made by comparing the G_{SP}^2 value against the χ^2 distribution at the appropriate degrees of freedom. The hypothesis framework for simultaneously evaluating the parameters is formulated as follows:

$\mathbf{H_0}$: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$

$\mathbf{H_1}$: at least one $\beta_j \neq 0$, for $j = 1,2, \dots, p$

Suppose $\mathrm{L}(\widehat{\boldsymbol{\Omega}}_{SP})$ is the likelihood function under the population with $\widehat{\boldsymbol{\Omega}}_{SP} = \{\rho, \boldsymbol{\beta}\}$, $L(\widehat{\boldsymbol{\omega}}_{SP})$ is the likelihood function under H_0 with $\widehat{\boldsymbol{\omega}}_{SP} = \{\rho_{\omega 0}, \beta_0\}$ and G_{SP}^2 is the test statistic used for simultaneous parameter significance testing, also called the likelihood ratio test. Then G_{SP}^2 can be written in equations (18) and (19).

$$V = \left[\frac{L(\hat{\boldsymbol{\omega}}_{SP})}{L(\hat{\boldsymbol{\Omega}}_{SP})}\right] \quad (18)$$

$$G_{SP}^2 = -2\ln[V_{SP}] = 2\left[\ln L(\hat{\boldsymbol{\Omega}}_{SP}) - \ln L(\hat{\boldsymbol{\omega}}_{SP})\right] \quad (19)$$

The decision to reject the null hypothesis (H_0) is taken if the statistical value of G_{SP}^2 is greater than the critical value of χ_{df}^2. The degrees of freedom (df) magnitude is obtained from the difference between the number of parameters in the population and the total parameters under the H_0 condition. The comparison value of χ_{df}^2 can be seen through the Chi-Square distribution table reference.

2.6. Partial Testing

After the overall model significance was proven, the next step was to dissect the contribution of each individual predictor. This procedure was performed using a partial test based on the Wald test statistic. The Z_{hitung} statistic is computed based on the comparison between the estimated parameter value and its standard error, which is then compared with the critical value in the standard normal distribution table. The hypotheses for conducting the partial tests were as follows:

$H_0: \beta_j = 0$, for $j = 1,2,\ldots,p$.

$H_1: \beta_j \neq 0$.

The test statistic used in the partial model parameter testing is as in Equation (20) below:

$$Z_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (20)$$

where $\hat{\beta}_j$ is the parameter estimator of β_j and $SE(\hat{\beta}_j)$ is $\sqrt{\widehat{var}(\hat{\beta}_j)}$, , with $\widehat{var}(\hat{\beta}_j)$ being the $(j+1)$-th diagonal of $\widehat{Cov}(\hat{\boldsymbol{\beta}}^*) \underset{n->\infty}{\cong} \mathfrak{I}^{-1}(\hat{\boldsymbol{\beta}}^*)$. The decision is to reject H_0 if $Z_{hitung} < -Z_{\alpha/2}$ or $Z_{hitung} > Z_{\alpha/2}$, where the value of $Z_{\alpha/2}$ can be obtained from the standard normal table.

3. RESEARCH METHODS

3.1. Research Data and Variables

This study utilizes secondary data from Badan Pusat Statistik (2022), covering 131 observation units of regencies/cities on Sumatra Island. The analyzed response variable was the binary poverty status with a threshold value of 9.54%. The factors tested for their influence included GRDP growth, unemployment rate, per capita expenditure, and expected years of schooling (see Table 1).

Table 1. Research Variables

Variabel	Tipe Data	Keterangan
Poverty Rate (Y)	Nominal	$0; y_i^* < 9{,}54$ $1; y_i^* \geq 9{,}54$
GRDP Growth Rate (X_1)	Ratio	-
Open Unemployment Rate (X_2)	Ratio	-
Average Per Capita Expenditure (X_3)	Ratio	-
Expected Years of Schooling (X_4)	Ratio	-

3.2. Research Stages

The research process was divided into several interconnected stages. The complete sequence for each stage is presented in the flowchart below (see Figure 2):

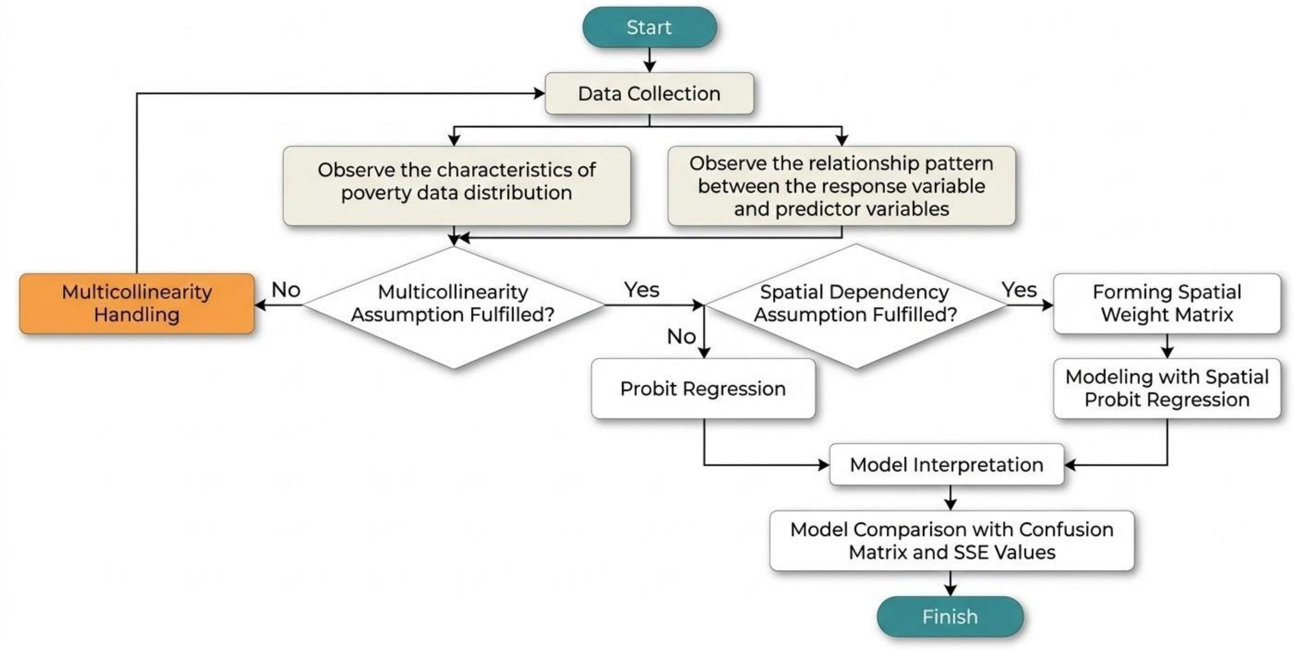

Figure 2. Research Phase Flowchart

4. RESULTS AND DISCUSSION

4.1. Overview of Poverty Rates in Regencies/Cities in Sumatra Island

A visualization of the distribution of poverty rates at the regency/city level on Sumatra Island for 2022 is presented in Figure 3. In the map, the low poverty rate classification is represented by the yellow zone, while areas with high poverty rates are marked in brown. Based on visual observation, there is a pattern where geographically adjacent regions show similarities in the degree of poverty, which explicitly indicates the presence of spatial dependency in poverty data in the Sumatra region.

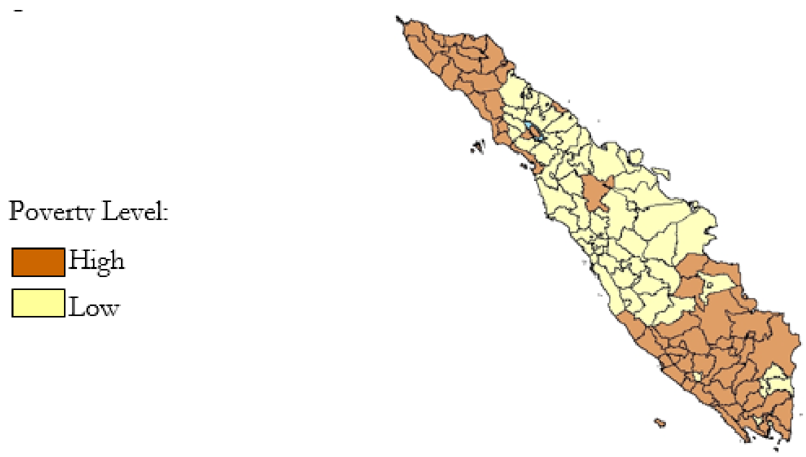

Figure 3. Poverty Rates by Regency/City on the Island of Sumatra in 2022

4.2. Pre-Modeling Analysis

The pre-estimation stage in the SAR Probit model requires verification of classic assumptions to ensure that the resulting model has high validity. One crucial aspect is ensuring the absence of multicollinearity, which is a phenomenon in which predictor variables correlate significantly with each other. As a preventive measure in modeling the poverty rate, early detection is performed by evaluating the Variance Inflation Factors (VIF) values. The following are the details of the VIF parameters for all the independent variables analyzed in this study.

Table 2. Predictor Variable Multicollinearity

Variable	VIF Value
$\mathbf{X_1}$	1,067145
$\mathbf{X_2}$	1,402539
$\mathbf{X_3}$	1,331065
$\mathbf{X_4}$	1,430545

Source: RStudio Software Processing Results

The data presented in Table 2 confirm that all Variance Inflation Factor (VIF) values are below the threshold of 10. These findings provide empirical evidence that there are no multicollinearity issues among the independent variables in this case study. With this assumption fulfilled, all predictor variables were eligible for integration into the subsequent statistical analysis stages.

The next step in this research procedure was to evaluate the presence of autocorrelation in the dependent variable through spatial dependency testing. The identification of inter-regional interactions is implemented using the Lagrange Multiplier (LM-Lag) test. The hypothesis formulation used to detect this spatial effect specifically refers to the theoretical framework.

$\mathbf{H_0}: \rho = 0$ (no spatial lag dependency)

$\mathbf{H_1}: \rho \neq 0$ (spatial lag dependency exists)

Referring to the statistical calculation results using the formula with a 5% error rate, an LM lag value of 19.51115 was obtained. This figure is greater than the Chi-square table value (3.8415) for one degree of freedom, so it is decided to reject H_0. This indicates strong evidence of inter-regional dependence in poverty patterns on Sumatra Island in 2022. Thus, the analysis process can proceed to the core modeling stage.

4.3. Poverty Rate Modeling with SAR Probit Method

The application of the spatial probit model in this study is intended to identify various variables that significantly contribute to poverty. The analysis process began with parameter estimation using the Fisher Scoring algorithm. This technique was chosen to overcome the complexity of non-closed-form functions in the Maximum Likelihood procedure. The parameter estimation results are as follows (see Table 3):

Table 3. Parameter Estimation Results Using Fisher Scoring Iteration

Variabel	Estimasi
Intercept (β_0)	3,38424
$X_1(\beta_1)$	-0,15899
$X_2(\beta_2)$	0,11361
$X_3(\beta_3)$	-0,16717
$X_4(\beta_4)$	-0,20512
ρ	2,5349

Source: RStudio Software Processing Results

From a spatial planning policy perspective, the presence of a significant spatial dependence effect indicates that poverty in Sumatra is closely linked. Policy interventions, such as the development of educational infrastructure, job creation, or economic stimulus, should not be implemented in isolation in each regency. Instead, local governments in Sumatra must adopt a coordinated, cross-regional approach, as economic improvements or setbacks in one district significantly impact the surrounding areas.

Based on the parameter estimation results using the Fisher Scoring approach, the SAR Probit regression model is as follows:

$$[\Lambda X\beta]_i = 3{,}384 + 2{,}535\left(\sum_{i=1, i^*\neq i}^{n} w_{ij}y_i^*\right) - 0{,}159X_{1i} + 0{,}114X_{2i} - 0{,}1677X_{3i} - 0{,}205X_{4i}$$

The analysis process then continues by verifying the parameters in the SAR Probit model through simultaneous and partial testing approaches. In the spatial probit regression model, simultaneous testing is conducted to evaluate the collective effect of the coefficient vector $\boldsymbol{\beta}$. This joint significance test utilizes the likelihood ratio test method based on the following hypothesis:

$\mathbf{H_0}$: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$,

$\mathbf{H_1}$: at least one $\beta_j \neq 0$, for $j = 1,2,3,4$.

Based on the collective parameter significance test results, a G_{SP}^2 statistical value of 44.30691 is obtained. This value is proven to be greater than the critical Chi-Square table value at a significance level of $\alpha = 0.05$ with 4 degrees of freedom, which is 9.487729, so the statistical decision taken is to reject the null hypothesis (H_0). This finding indicates that in the spatial probit regression model with a 95% confidence level, at least one independent variable has a tangible effect on the model as a whole. Based on these results, the analysis proceeded to the individual parameter testing stage (partial test) to identify which predictor variables specifically contributed significantly to the response variable.

The decision to reject H_0 in the simultaneous test requires a more in-depth analysis through partial hypothesis testing of the spatial probit model parameters. The focus of this test was to dissect the level of significance of the contribution of each predictor variable to the response variable separately. In its implementation, the Wald test statistic is used as the main instrument to verify the individual influence of each variable in the model.

$\mathbf{H_0}$: $\beta_j = 0$,

$\mathbf{H_1}$: $\beta_j \neq 0$, for $j = 1,2,3,4$.

The details of the individual parameter evaluation results with a significance level of $\alpha = 0.05$ are presented in Table 4. Based on these data, it is identified that the GRDP growth rate variable (X_1), open unemployment rate (X_2), average per capita expenditure (X_3), and expected years of schooling (X_4) have a partially significant effect on the poverty classification at the regency/city level in Sumatra Island. This finding confirms that these four economic and social indicators are important predictors of mapping poverty in the region.

Table 4. Partial Parameter Testing of Spatial Probit Model

Variable	Coeffisien	Std. Error	Z-value	P-value	Decision
Intercept	3,38424	-	-	-	-
X_1	-0,15899	0,0719	-2,2106	0,0289	Reject H_0
X_2	0,11361	0,0501	2,2689	0,0250	Reject H_0
X_3	-0,16717	0,0603	-2,7700	0,0065	Reject H_0
X_4	-0,20512	0,0987	-2,0773	0,0398	Reject H_0

Source: RStudio Software Processing Results

Table 4 shows that, in essence, the negative impact of GRDP growth rates and average per capita expenditure indicates that regions with higher economic growth and purchasing power tend to have a lower probability of experiencing high poverty rates, as these factors reflect a higher standard of living and the region's economic capacity. The positive impact of the open unemployment rate implies that a lack of job opportunities directly increases the risk of poverty owing to the loss of a stable source of income. Meanwhile, an increase in the expected years of education significantly reduces poverty, highlighting the critical role of education in enhancing human capital and productivity.

After the model structure is successfully formed, the next crucial stage involves calculating the predicted probability. The estimated probability for observation $y_i = 1$ in modeling the poverty rate across regencies/cities in Sumatra Island is formulated as follows:

$$\hat{P}(y_i = 1|X) = \hat{P}(y_i^* \geq 0|X) = 1 - \Phi\left(\frac{-[\Lambda X\hat{\beta}]_i}{\sqrt{\varsigma_{ii}}}\right)$$
$$= 1 - \Phi\left(\frac{-[(\mathbf{I} - 2{,}5349\mathbf{W})^{-1}X\hat{\beta}]_i}{\sqrt{\varsigma_{ii}}}\right)$$

As an illustration of the model's application, the calculation of the predicted probability value for Medan City, which occupies observation index $i = 49$, is carried out. The magnitude of the probability $y_i = 1$ in modeling the poverty rate in the region is determined through the following mathematical procedure.

$$[\Lambda X\hat{\beta}]_{49} = 3{,}38 + 2{,}54\left(\sum_{i=1, i^* \neq i}^{131} w_{49i^*} y_{49}^*\right) - 0.16X_{1;49} + 0{,}11X_{2;49} - 0{,}17X_{3;49} - 0{,}22X_{4;49}$$

Therefore, if substituted into Equation (17), the model becomes

$$P(y_{49} = 1|\mathbf{X}_{49}) = 1 - \Phi\left(\frac{-[\Lambda X\hat{\beta}]_{49}}{\sqrt{\varsigma_{49,49}}}\right)$$
$$= 0{,}096536984$$

The estimated predicted probability result for Medan City shows a value of 0.096536984, which implies that the probability of the region being in the high poverty category is only 9.65%, while the probability of remaining in the low poverty category is 90.35%. This finding indicates that the poverty rate in Medan City is influenced by internal indicators such as the economic growth rate, open unemployment rate, per capita income, and expected years of schooling. The existence of spatial linkages with surrounding regions, especially Deli Serdang Regency, which is also predicted to have a low poverty rate, makes a tangible contribution to Medan City's economic status. This reinforces the strong geographical dependence between the two bordering regions in the context of poverty distribution.

A visualization of the distribution of the predicted poverty results estimated using the SAR Probit model is presented in Figure 4. The findings from this modeling show that there are 66 regencies/cities included in the high poverty rate classification, while the other 65 regions are predicted to be in the low category.

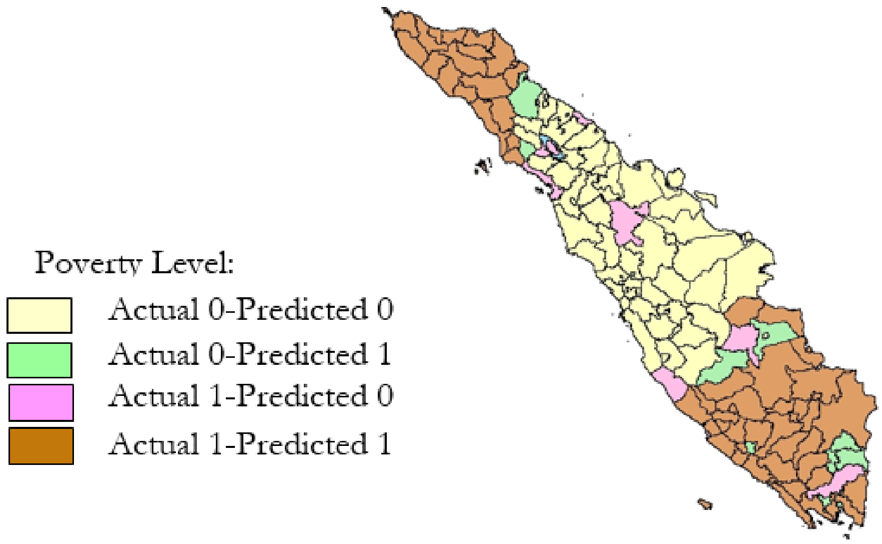

Figure 4. Map of Actual and Predicted Poverty Levels from The Spatial Probit Model

A more detailed explanation of the regencies/cities between the actual and prediction data is shown in Table 5.

Table 5. Results of Actual and Prediction Data Grouping by Regency/City

Actual	Predicted	Regency/City
0	0	Humbang Hasundutan, Kota Medan, Kota Dumai, Kampar, Kota Pekanbaru, Dairi, Labuhanbatu Selatan, Kota Padang Sidempuan, Padang Lawas, Kota Jambi, Tapanuli Selatan, Pasaman Barat, Kota Sungai Penuh, Bungo, Asahan, Tebo, Kota Banda Aceh, Kota Binjai, Serdang Berdagai, Labuhanbatu, Kota Padang Panjang, Siak, Tanah Datar, Kota Bukit Tinggi, Kota Padang, Padang Lawas Utara, Kota Payakumbuh, Labuhanbatu Utara, Kota Solok, Agam, Lima Puluh Kota, Karo, Pasaman, Sijunjung, Solok, Indragiri Hulu, Kuantan Singingi, Pariaman, Pesisir Selatan, Merangin, Dharmasraya, Kota Sawahlunto, Kota Pematangsiantar, Mandailing Natal, Pelalawan, Deli Serdang, Kerinci, Indragiri Hilir, Kota Pariaman, Rokan Hilir, Tapanuli Utara, Solok Selatan, Simalungun, dan Padang Toba Samosir.
0	1	Langkat , Kota Metro , Mesuji , Sarolangun , Tulang Bawang , Pakpak Bharat , Kota Bandar Lampung , Pringsewu , Muaro Jambi , Kota Pagar Alam , dan Tulang Bawang Barat.
1	0	Kota Tanjung Balai, Batanghari, Samosir, Kota Bengkulu, Tapanuli Tengah, Batu Bara, Lampung Tengah, Kota Tebing Tinggi, Muko Muko, dan Rokan Hulu.
1	1	Aceh Tengah , Ogan Komering Ilir , Musi Banyuasin , Lampung Selatan , Aceh Tamiang , Kota Lhokseumawe , Way Kanan , Empat Lawang , Pidie Jaya , Tanjung Jabung Timur , Bireuen , Ogan Ilir , Lampung Barat , Musi Rawas Utara , Aceh Barat , Nagan Raya , Kota Palembang , Muara Enim , Lampung Timur , Aceh Singkil , Bengkulu Tengah , Ogan Komering Ulu Selatan , Seluma , Lampung Utara , Kapahiang , Aceh Selatan , Aceh Timur , Gayo Luwes , Kota Langsa , Pesawaran , Penukal Abab Lematang Ilir , Aceh Barat Daya , Aceh Jaya , Banyuasin , Rejang Lebong , Bengkulu Bengkulu , Kota Prabumulih , Musi Rawas , Kota Subulussalam , Ogan Komering Ulu Timur , Tanjung Jabung Barat , Pidie , Aceh Utara , Lampung Tengah , Bener Meriah , Aceh besar , Lahat , Kaur , Lebong , Bengkulu Utara , Kota Lubuk Linggau , Tanggamus , Aceh Tenggara , Kota Sibolga , Ogan Komering Ulu , dan Pesisir Barat.

4.4. Evaluation and Model Comparison

After conducting spatial probit modeling using the Fisher scoring approach, a comparison was made with the spatial probit model using the RIS simulator approach. The models were compared based on their accuracy, sensitivity, and specificity. The comparison is also made by examining which SSE value is smaller for each model.

The results of comparing the classification capabilities of the SAR Probit model estimated using the Fisher Scoring approach and the Recursive Importance Sampling (RIS) simulator approach are presented in the following table.

Table 6. Comparison Results of Fisher Scoring Iteration Approach with RIS Simulator

Model Goodness Measure	Fisher Scoring	RIS Simulator
Accuracy	83.97%	74.05%
Sensitivity	84.61%	63.08%
Specificity	83.33%	84.85%
Sum of Squared Errors (SSE)	15.87	27.65

Source: RStudio Software Processing Results

The comparison results in Table 6 show that the Fisher Scoring approach is superior in terms of total accuracy and sensitivity. The much lower SSE value (15.87) indicates that this model has better estimation precision than the RIS Simulator in modeling poverty in Sumatra. Therefore, it can be concluded that the spatial probit model with the Fisher scoring approach is more effective and accurate in modeling poverty rates, providing predictions closer to the true values and producing better statistical results.

The pattern revealed by the SAR probit model should not be read only as a statistical dependency but as evidence that poverty in Sumatra is territorially structured. The coexistence of high-poverty clusters and low-poverty clusters implies that neighboring labor markets, commodity chains, and public-service environments may transmit deprivation across administrative borders. Similar interpretations are common in spatial growth and convergence research, where adjacent regions often move together because shocks and opportunities diffuse geographically (Rey & Montouri, 1999; Fingleton & López-Bazo, 2006). In practical terms, the findings suggest that district-level poverty programs will be less effective when they are implemented without coordination with surrounding jurisdictions.

This result is also consistent with the broader logic of spatial discrete-choice modeling. When poverty status is coded as a binary outcome, the probability that a district enters the high-poverty category is shaped not only by its own socioeconomic conditions but also by the configuration of neighboring areas. Earlier work on neighborhood influence and spatial probit estimation reached a similar conclusion: local choices and outcomes can become interdependent once interaction effects are admitted into the model (Case, 1992; LeSage, 2000). Therefore, the significant spatial parameter in this study should be understood as substantive spillover rather than a mere nuisance term.

From a methodological standpoint, the superior performance of the Fisher Scoring approach over the RIS simulator is important because it shows that estimation stability can materially affect substantive inference. A method that converges more reliably and yields smaller prediction error is especially valuable when researchers work with regional data characterized by uneven distributions and nontrivial spatial dependence. This interpretation aligns with the wider spatial econometric literature, which emphasizes that implementation details, optimization routines, and numerical stability shape model quality as much as theoretical specification does (Elhorst, 2010; Bivand & Piras, 2015). In this study, the higher sensitivity of Fisher Scoring is particularly useful because it improves the identification of districts that are genuinely vulnerable to high poverty.

At the same time, the current specification should be read as a parsimonious starting point rather than the final word on spatial poverty modeling in Sumatra. Recent spatial-econometric work has shown that alternative formulations, including models that more explicitly trace local spillovers or simplify likelihood computation, can provide complementary insights depending on the theoretical objective and data structure (LeSage & Pace, 2007; Halleck Vega & Elhorst, 2015). Future research could therefore compare SAR probit estimates with SLX-style or MESS-based formulations, especially when the aim is to separate direct local effects from broader neighborhood transmission mechanisms.

The policy implication is that poverty reduction on Sumatra Island should combine place-based targeting with intergovernmental coordination. Improvements in education, labor absorption, and household expenditure in one district may generate indirect benefits for adjacent districts, while persistent weakness in one area can slow progress in nearby locations. This reading is consistent with foundational discussions of indirect effects and spatial spillovers in regional econometrics, as well as later work on the large-sample behavior of alternative spatial specifications (LeSage & Pace, 2009; Debarsy et al., 2015). Accordingly, future studies should move beyond binary classification accuracy alone and estimate marginal effects, spillover magnitudes, and regional policy scenarios.

5. CONCLUSION

Based on the series of analyses and discussions, this study draws several conclusions. First, parameter estimation in the spatial probit regression model was conducted using the Maximum Likelihood Estimator (MLE) method. Because the resulting equation is non-closed form and does not have an analytical solution, the Fisher Scoring numerical iteration procedure was applied to obtain accurate parameter estimates. Second, the significance of the spatial probit model was tested using the Maximum Likelihood Ratio Test (MLRT), while the estimation process under both population conditions and the spatial model framework continued to rely on the Fisher Scoring algorithm to address the complexity of the equation. The hypothesis testing results on poverty data in Sumatra in 2022 demonstrate that per capita expenditure, unemployment rate, economic growth as measured by GRDP, and expected years of schooling significantly influence the classification of poverty rates, both simultaneously and partially. Third, the spatial probit model results show a high level of predictive accuracy, with 110 regencies or cities correctly classified in accordance with the observed data, while 21 regions were classified differently. Overall, the spatial probit model with the Fisher Scoring approach achieved an accuracy rate of 83.97 percent, indicating reliable model performance. Finally, comparative evaluation reveals that Fisher Scoring-based spatial probit regression performs better than the RIS Simulator approach in terms of accuracy, sensitivity, and efficiency, as reflected in its lower Sum of Squared Errors (SSE). However, the RIS Simulator approach produced a higher specificity. Taken together, these findings indicate that the Fisher Scoring approach is more effective for generating precise poverty predictions, whereas the RIS Simulator offers a particular advantage in identifying the negative category in poverty data.

Ethical Approval

Ethical approval was not required for this study as it utilizes secondary, publicly available data from Statistics Indonesia (BPS)

Informed Consent Statement

Informed consent was not obtained for this study, as it analyzed secondary aggregated data that did not involve direct interaction with human subjects.

Author's Contibutions

PJ conceptualized the study, developed the methodology together with NRTN, and conducted the formal analysis, along with NRTN. G, FM, and EA were responsible for validation, and EA and FM provided the necessary resources to support the research. PJ prepared the original draft of the manuscript, and G, NRTN, and FM contributed to the review and editing.

Disclosure Statement

The authors declare no conflicts of interest.

Data Availability Statement

The data presented in this study are available on request from the corresponding author due to privacy reasons.

Funding

This study did not receive any external funding.

Notes on Contributors

Panji Jiblathar
https://orcid.org/0009-0001-7492-9141
Panji Jiblathar is a Statistics Lecturer in the Faculty of Islamic Economics and Business at UIN Sulthan Thaha Saifuddin Jambi, where he specializes in teaching Econometrics. His primary research interests lie in advanced statistical modeling, particularly the application of spatial probit models and regression analyses. As the lead author of this study, he was responsible for the conceptualization, performing the core data analysis using the Fisher Scoring numerical iteration approach, and drafting the original manuscript.

Germansah
https://orcid.org/0009-0005-1184-8984
Germansah is a Lecturer at the Faculty of Islamic Economics and Business, UIN Sulthan Thaha Saifuddin Jambi. His academic focus includes economic theory and regional studies. In this study, he contributed to the validation of economic indicators used in the modeling process and played an active role in the comprehensive review and editing of the manuscript.

Fadhlul Mubarak
https://orcid.org/0000-0001-7394-3817
Fadhlul Mubarak serves as a Lecturer at the Faculty of Islamic Economics and Business, UIN Sulthan Thaha Saifuddin Jambi. His expertise includes public policy evaluation and economic development research. His contributions to this study involved providing supporting literature resources and analyzing the socioeconomic implications derived from the results of poverty modeling.

Ellys Agustina
Ellys Agustina is a professional practitioner and civil servant at the Regional Financial and Revenue Management Agency of Kerinci Regency. She has extensive practical experience in regional financial management and data governance. For this study, she provided critical support in terms of empirical data resources and validated the relevance of the model's findings against practical field conditions.

Nabilla Rida Tri Nisa
Nabilla Rida Tri Nisa is a Lecturer at the State Media Polytechnic. Her primary expertise is in applied statistics and computational simulations. She provided significant methodological contributions to the advanced formal analysis, specifically regarding the model comparison using the RIS Simulator technique, and assisted in the final editing of the research paper.

REFERENCES

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-7799-1

Badan Pusat Statistik. (2022). *Data dan informasi kemiskinan kabupaten/kota tahun 2022* (Data and information on regency/municipality poverty, 2022). https://www.bps.go.id/id/publication/2022/11/30/3b084878f782dfa44e0025e0/data-dan-informasi-kemiskinan-kabupaten-kota-tahun-2022.html

Bivand, R., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, *63*(18), 1–36. https://doi.org/10.18637/jss.v063.i18

Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics*, *22*(3), 491–508. https://doi.org/10.1016/0166-0462(92)90041-X

Debarsy, N., Jin, F., & Lee, L.-F. (2015). Large sample properties of the matrix exponential spatial specification with an application to FDI. *Journal of Econometrics*, *188*(1), 1–21. https://doi.org/10.1016/j.jeconom.2015.02.046

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, *5*(1), 9–28. https://doi.org/10.1080/17421770903541772

Fingleton, B., & López-Bazo, E. (2006). Empirical growth models with spatial effects. *Papers in Regional Science*, *85*(2), 177–198. https://doi.org/10.1111/j.1435-5957.2006.00074.x

Greene, W. H. (2008). *Econometric analysis* (6th ed.). Pearson/Prentice Hall. https://books.google.com/books/about/Econometric_Analysis.html?id=b541vgAACAAJ

Halleck Vega, S., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339–363. https://doi.org/10.1111/jors.12188

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, *32*(1), 19–35. https://doi.org/10.1111/j.1538-4632.2000.tb00413.x

LeSage, J. P., & Pace, R. K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, *140*(1), 190–214. https://doi.org/10.1016/j.jeconom.2006.09.007

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press. https://doi.org/10.1201/9781420064254

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science, 32*(3), 335–348. https://doi.org/10.1111/j.1467-9787.1992.tb00190.x

Rey, S. J., & Montouri, B. D. (1999). US regional income convergence: A spatial econometric perspective. *Regional Studies*, *33*(2), 143–156. https://doi.org/10.1080/00343409950122945

Schworer, A., & Hovey, P. (2004). Newton-Raphson versus Fisher scoring algorithms in calculating maximum likelihood estimates. *Electronic Proceedings of Undergraduate Mathematics Day, 1*, Article 7, 1-11. https://ecommons.udayton.edu/epumd/vol1/iss1/7