Spatial Modeling for Poverty: The Comparison of Spatial Error Model and Geographic Weighted Regression

Achi Rinaldi[1], Yuni Susianto[2], Budi Santoso[3], Wahyu Kusumaningtyas[4]
[1] Universitas Islam Negeri Raden Intan Lampung, Indonesia
[2,3] Badan Pusat Statistik, Indonesia
[4] STKIP Kumala Metro, Indonesia

Article Info	Abstract
	This study aims to analyze poverty using spatial models. The researchers also compared the Spatial Error Model (SEM) and Geographically Weighted Regression (GWR). The comparison of the two models was based on the estimation evaluation criteria and the constructed spatial associations. Spatial regression is considered very appropriate to be used to model the relationship pattern between poverty and explanatory variables when the observed data has a spatial effect caused by the proximity between the observation areas. The spatial dependence of errors on observational data can be overcome using SEM, while the effect of heterogeneity of spatial variance can overcome using GWR. Keywords: Geographically Weighted Regression, poverty, regression, spatial dependence, Spatial Error Model.

Introduction

Poverty is one of the fundamental problems concerned by all countries in the world (Atkinson, 1987; Ferezagia, 2018; Nurwati, 2008), including Indonesia. Poverty can be overcome by correctly identifying the variables that have a real effect on poverty. The poverty of a region cannot be separated from the influence of the poverty of other regions around it. This problem requires special attention to spatial effects in modeling poverty data (Djuraidah & Wigena, 2012). Spatial regression is an approach that can be used to model spatial influence data, both spatial dependence and spatial heterogeneity (Anselin, 2009).

Spatial Autoregressive (SAR) model and the Spatial Error Model (SEM) can be used on spatial dependency cases (Kelejian & Prucha, 2010; Lee & Yu, 2010). On the other hand, the Geographically Weighted Regression (GWR) model can be used on spatial heterogeneity cases.(Charlton et al., 2009). Based on these considerations, the poverty data for districts or cities in East Java can be modelled using the Spatial Regression Model. By applying the Spatial Regression Model in modelling the poverty data, a complete picture of the variables that affect poverty can be illustrated.

Several research has been done on spatial modeling, whether it's SAR, SEM, and even those involving extreme values (Kelejian & Prucha, 2010; Lee & Yu, 2010; Putra et al., 2020; Rinaldi et al., 2017; Zhang et al., 2021). The GWR spatial model has also been studied and researched (Griffith, 2008; Zhou et al., 2019; Zhu et al., 2020). However, previous research was focused on separate modelling. Therefore, this study aims to examine the spatial models and compare the advantages of the models as a whole, even in analytical form.

The Research Methods

In general, the form of the mean and variance functions for multiple regression is denoted as follows:

$$E(Y|X) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} X_1 + \boldsymbol{\beta_2} X_2 + \cdots + \boldsymbol{\beta_p} X_p \tag{1}$$

$$\mathbf{Var}(Y|X) = \boldsymbol{\sigma^2}$$

β_i for $i = 0, 1, 2, \dots, p$ or σ^2 are unknown parameters and must be estimated (Draper & Smith, 1998; Fox & Weisberg, 2018; Montgomery et al., 2021; Myers, 1990; Timm, 2002; Weisberg, 2013).

For example, for n observational data, the response variables and explanatory variables are defined in the form of vectors and matrices:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Y is a vector of n x 1 and X is a matrix of n x (p+1). The regression coefficient is denoted as $\boldsymbol{\beta}$, which is a vector of (p + 1) x 1, where e is a residual vector of (n 1).

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta_0} \\ \boldsymbol{\beta_1} \\ \vdots \\ \boldsymbol{\beta_p} \end{pmatrix} \qquad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \vdots \\ \boldsymbol{e_n} \end{pmatrix}$$

The multiple regression matrix can be written as follows:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

Assuming the model: $\boldsymbol{\varepsilon} \underset{iid}{\sim} \boldsymbol{N(0, \sigma^2 I_n)}$

The least squares method is used to estimate the parameters $\boldsymbol{\beta}$ by minimizing the sum of the residual squares.

$$\boldsymbol{JK(e)} = \sum \boldsymbol{e_i}^2 = \sum (\boldsymbol{y_i} - \widehat{\boldsymbol{y_i}})^2 = \sum (\boldsymbol{y_i} - \boldsymbol{x_i'\beta})^2$$
$$= (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

By deriving the above equation with the parameter $\boldsymbol{\beta}$, the estimator will be:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X'Y}) \text{ whereas for } \boldsymbol{Var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X'X})^{-1}$$

The assumptions in the classical regression model are:

1. $E(\varepsilon_i) = 0$, for i = 1, 2, …, n; therefore, the expected value becomes:

$E(\mathbf{y}_i) = \beta_0 + \beta_1 \mathbf{X}_{i1} + \beta_2 \mathbf{X}_{i2} + \dots + \beta_p \mathbf{X}_{ip}$

2. $Var(\varepsilon_i) = \sigma^2$, for i = 1, 2, …, n, or equal to $Var(\mathbf{y}_i) = \sigma^2$

3. $Cov(\varepsilon_i, \varepsilon_j) = 0$, for i≠j.

General Spatial Regression Model (GSM)

The general model of spatial regression is:

$$\mathbf{Y} = \rho\mathbf{Wy} + \mathbf{X\beta} + \mathbf{u} \tag{3}$$

$$\mathbf{u} = \lambda\mathbf{Wu} + \mathbf{\varepsilon} \tag{4}$$

$$\mathbf{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

y is the dependent variable of $n \times 1$, X is the matrix of independent variables of (n x (p+1)), $\boldsymbol{\beta}$ is the vector of regression parameter coefficients of $p \times 1$, ρ is a spatial lag autoregression coefficient, λ is an autoregression coefficient spatial error value of $|\lambda| < 1$, u is an error vector which is assumed to contain autocorrelation of $n \times 1$, W is the spatial weighting matrix of $n \times n$, n is the number of observations (Anselin, 2009; Blangiardo & Cameletti, 2015; LeSage & Pace, 2009). The estimation parameter in GSM model is obtained using maximum probability estimation method (Arbia & Baltagi, 2008; LeSage & Pace, 2009; Schabenberger & Gotway, 2017). Based on equation (2), it can be expressed as:

$$\mathbf{y}\text{-}\rho\mathbf{Wy} = \mathbf{X\beta} + \mathbf{u} \text{ or}$$

$$\mathbf{I} - \rho\mathbf{W} \, \mathbf{y} = \mathbf{X\beta} + \mathbf{u} \tag{5}$$

Equation (3) can be expressed as:

$$(\mathbf{I}\text{-}\lambda\mathbf{W})\,\mathbf{u} = \mathbf{\varepsilon} \text{ or } \mathbf{u} = (\mathbf{I}\text{-}\lambda\mathbf{W})^1\mathbf{\varepsilon} \tag{6}$$

By substituting equation (6) into equation (5):

$$(\mathbf{I}\text{-}\rho\mathbf{W})\,\mathbf{y} = \mathbf{X\beta} + (\mathbf{I}\text{-}\lambda\mathbf{W})^1\mathbf{\varepsilon}$$

$$(\mathbf{I}\text{-}\lambda\mathbf{W})^1\mathbf{\varepsilon} = (\mathbf{I}\text{-}\rho\mathbf{W})\,\mathbf{y}\text{-}\mathbf{X\beta}$$

If all sides are multiplied by $(\mathbf{I}\text{-}\lambda\mathbf{W})$, then:

$$\mathbf{\varepsilon} = (\mathbf{I}\text{-}\lambda\mathbf{W})\,(\mathbf{I}\text{-}\rho\mathbf{W})\,\mathbf{y}\text{-}\mathbf{X\beta} \tag{7}$$

The value of the probability function of the variable ε is:

$$L\,(\sigma^2; \mathbf{\varepsilon}) = c(\mathbf{\varepsilon})|\mathbf{V}|^{-1/2}\exp\left[\frac{-1}{2\sigma^2}\mathbf{\varepsilon}^t\mathbf{\varepsilon}\right] \tag{8}$$

Where V is the ε covariance matrix of $\mathbf{V} = \sigma^2\mathbf{I}$. The determinant of the matrix V is $\sigma^2 n$ and the reciprocal of the covariance matrix of $\mathbf{V}^{-1} = 1/(\sigma^2\mathbf{I})$. By substituting the value of $|V|$ and $\mathbf{V}^{-1}$ in equation (8), it is obtained that:

$$L\,(\sigma^2; \mathbf{\varepsilon}) = c(\mathbf{\varepsilon})\sigma^{2n}\exp\left[\frac{-1}{2\sigma^2}\mathbf{\varepsilon}^t\mathbf{\varepsilon}\right] \tag{9}$$

From the ε and y relationship in equation (7), the Jacobian value is:

$$J = \left|\frac{\partial\varepsilon}{\partial y}\right| |\mathbf{I}\text{-}\lambda\mathbf{W}| \, |\mathbf{I}\text{-}\rho\mathbf{W}|$$

By substituting equation (7) into equation (9), the probability function for y will be:

$$L\,(\rho, \lambda, \sigma^2, \mathbf{\beta}; \mathbf{y}) = c(\mathbf{y})\,(\sigma^2)^{-n/2}\,|\mathbf{I}-\lambda\mathbf{W}|\,|\mathbf{I}-\rho\mathbf{W}|$$
$$\exp\left[-\frac{1}{2\sigma^2}\{(I-\rho W)[(I-\rho W)y - X\beta]\}^T\{(I-\rho W)[(I-\rho W)y - X\beta]\}\right]$$

The log-likelihood function obtains the following equation (10):

$$\ln L\,(\rho, \lambda, \sigma^2, \mathbf{\beta}; \mathbf{y}) = \ln(c(\mathbf{y})) - n/2\ln(\sigma^2) + \ln|\mathbf{I}-\lambda\mathbf{W}| + \ln|\mathbf{I}-\rho\mathbf{W}|$$
$$-\frac{1}{2\sigma^2}\{(I-\rho W)[(I-\rho W)y - X\beta]\}^T\{(I-\rho W)[(I-\rho W)y - X\beta]\}$$

Suppose the square of the weighting matrix $(\mathbf{I}\text{-}\rho\mathbf{W})^{\mathbf{T}}(\mathbf{I}\text{-}\rho\mathbf{W})$ is denoted as Ω and estimator β is obtained by maximizing the log probability function in equation (10), then the parameter estimator β is:

$$\widehat{\mathbf{\beta}} = (\mathbf{X}'\mathbf{\Omega}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}(\mathbf{I}-\lambda\mathbf{W})\mathbf{Y}$$

Spatial Lag Regression (SAR) Model

If $\rho \neq 0$ and $\lambda = 0$, then equation (3) is the general form of the spatial regression model into a spatial lag regression model:

$$\mathbf{y} = \rho \mathbf{Wy} + \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{11}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

Response variables in the SAR model are spatially correlated. The maximum likelihood method can be used to estimate the parameters of this model (Lawson, 2013; Ver Hoef et al., 2018; Zhang et al., 2021).

In equation (11), ε_i is assumed to be normally distributed, stochastically free, identical, with a mean of zero, and a variance of σ^2. ε_i is the error at the location of i.

The probability density function of ε_i:

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right]$$

Joint probability density function of n random variables $1, 2, \ldots, \varepsilon_n$:

$$f(\boldsymbol{\varepsilon}) = f(\varepsilon_1) . f(\varepsilon_2) \ldots f(\varepsilon_n)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n} \varepsilon_i^2}{2\sigma^2}\right)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T \varepsilon}{2\sigma^2}\right]$$

The density function with the response variable y is obtained by transforming an n-dimensional $\boldsymbol{\varepsilon}$ space into an n-dimensional y space. From equation (11), it is obtained that:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \rho \mathbf{Wy} - \mathbf{X\beta}$$

The joint probability density function of n response variables y is:

$$f(\mathbf{y}) = f(\boldsymbol{\varepsilon})|J|$$
$$= \exp\left[-\frac{\varepsilon^T \varepsilon}{2\sigma^2}\right]\left|\frac{d\varepsilon}{dy}\right|$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{(\mathbf{y}-\rho\mathbf{Wy}-\mathbf{X\beta})^T(\mathbf{y}-\rho\mathbf{Wy}-\mathbf{X\beta})}{2\sigma^2}\right]|\mathbf{I} - \rho\mathbf{W}|$$

The probability function of the response variable y:

$$L(\boldsymbol{\beta}, \rho, \sigma^2; \mathbf{y}) = f(\mathbf{y}; \boldsymbol{\beta}, \rho, \sigma^2)$$
$$= |I-\rho\mathbf{W}|(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{(\mathbf{y}-\rho\mathbf{Wy}-\mathbf{X\beta})^T(\mathbf{y}-\rho\mathbf{Wy}-\mathbf{X\beta})}{2\sigma^2}\right]|\mathbf{I} - \rho\mathbf{W}| \tag{12}$$

The estimation of the model parameters is obtained by maximizing the probability function which is equivalent to maximizing the logarithm of the probability function in equation (12).

$$\ln(L(\boldsymbol{\beta}, \rho, \sigma^2; \mathbf{y}) = \ln\left\{|I - \rho W|(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{(y-\rho Wy-X\beta)^T(y-\rho Wy-X\beta)}{2\sigma^2}\right]|I - \rho W|\right\}$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|I - \rho W| - \frac{(y-\rho Wy-X\beta)^T(y-\rho Wy-X\beta)}{2\sigma^2} \tag{13}$$

The estimation for 2, β, and ρ is obtained by maximizing the log probability function in equation (13). The estimator for 2 is:

$$\hat{\sigma}^2 = \frac{(y-\rho Wy-X\beta)^T(y-\rho Wy-X\beta)}{n} \tag{14}$$

Equation (14) can be written as:

$$\sigma^2 = \sum \frac{(y_i - \hat{y}_i)^2}{n} = \frac{JKG}{n}$$

Where y_i is the response variable at location i, $\hat{y}_i$ is the estimator value of the dependent variable at location i, n is the number of observations, and JKG is the number of squared errors. Estimator for β is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} - (\mathbf{X}^T\mathbf{X})^{-1}\hat{\rho}\mathbf{Wy}$$

and the estimator for ρ is:

$$\hat{\rho} = (\boldsymbol{y}^T\boldsymbol{W}^T\boldsymbol{Wy})^{-1}\boldsymbol{y}^T\boldsymbol{W}^T\boldsymbol{y}$$

2.4. Error Spatial Regression Model (SEM)

If $\rho=0$ and $\lambda \neq 0$, then equation (3) which is the general form of the spatial regression model becomes the form of the spatial error regression model which can be written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \boldsymbol{u} = \lambda\boldsymbol{Wu} + \boldsymbol{\varepsilon} \qquad\qquad (15)$$

where is assumed $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\sigma^2 I})$

The spatial error model is a linear regression model in which the error variable has a spatial correlation due to the existence of explanatory variables that are not included in the linear regression model. Therefore, it will be calculated as an error and that variable is spatially correlated with errors in other locations. The spatial error parameters model can be estimated using the maximum likelihood method (Anselin & Florax, 2012; Anselin & Kelejian, 1997; Baltagi & Li, 2001).

The probability density function of $\boldsymbol{\varepsilon_i}$:

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right]$$

Joint probability density function of n random variables 1, 2…, n

$$f(\boldsymbol{\varepsilon}) = f(\varepsilon_1).\, f(\varepsilon_2)\dots f(\varepsilon_n)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n} \varepsilon_i^2}{2\sigma^2}\right)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T\varepsilon}{2\sigma^2}\right]$$

The density function with the response variable y is obtained by transforming an n-dimensional $\boldsymbol{\varepsilon}$ space into an n-dimensional y space. From equation (15), it can be obtained:

$\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{X\beta}$ dan

$\boldsymbol{\varepsilon} = (\mathrm{I} - \lambda\boldsymbol{Wu})\,\boldsymbol{u}$

Therefore, $\boldsymbol{\varepsilon} = (\boldsymbol{I} - \lambda\boldsymbol{Wu})\,(\boldsymbol{y} - \boldsymbol{X\beta})$

The joint probability density function of n response variables y:

$$f(\boldsymbol{y}) = f(\boldsymbol{\varepsilon})|J|$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T\varepsilon}{2\sigma^2}\right]\left|\frac{d\varepsilon}{dy}\right|$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{[(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X\beta})]^T[(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X\beta})]}{2\sigma^2}\right]|\boldsymbol{I} - \lambda\boldsymbol{W}|$$

The probability function of the variable y:

$$L(\boldsymbol{\beta}, \lambda, \sigma^2; \boldsymbol{y}) = f(\boldsymbol{y}; \boldsymbol{\beta}, \lambda, \sigma^2)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} |I - \lambda W| \exp\left[-\frac{[(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X\beta})]^T[(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X\beta})]}{2\sigma^2}\right] \qquad (16)$$

The logarithm of the probability function above is:

$$\ln(L(\boldsymbol{\beta}, \lambda, \sigma^2; \boldsymbol{y})$$

$$= \ln\left\{(2\pi\sigma^2)^{-\frac{n}{2}} |I - \lambda W| \exp\left[-\frac{[(I-\lambda W)(y-X\beta)]^T[(I-\lambda W)(y-X\beta)]}{2\sigma^2}\right]\right\}$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln |I - \lambda W| - \frac{[(I-\lambda W)(y-X\beta)]^T[(I-\lambda W)(y-X\beta)]}{2\sigma^2} \qquad (17)$$

The estimation for σ^2, $\boldsymbol{\beta}$ and λ is obtained by maximizing the log-likelihood function in equation (17).

The estimator for σ^2 is:

$$\hat{\sigma}^2 = \frac{\left[(I - \lambda W)(y - X\hat{\beta})\right]^T(I - \lambda W)(y - X\hat{\beta})]}{n}$$

The estimator for $\boldsymbol{\beta}$ is:

$$\hat{\beta} = [(X - \hat{\lambda}WX)^T(X - \hat{\lambda}WX)]^{-1}(X - \hat{\lambda}WX)^T(y - \hat{\lambda}Wy)$$

To estimate the λ parameter, a numerical iteration is needed to get an estimator for λ that maximizes the probability log function (Bivand et al., 2008).

2.5. Spatial Weighted Regression Model (GWR)

The GWR model equation is:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \qquad (18)$$

Where (u_i, v_i) represents the coordinates (longitude, latitude) of the i-th location.
The GWR parameters can be estimated using the weighted least squares approach(Griffith, 2008; Tasyurek & Celik, 2020; Zhu et al., 2020), obtained:

$$\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i)X]^{-1}XW(u_i, v_i)y$$

$W(u_i, v_i)$ is a diagonal weighting matrix of n x n, where the diagonal element is the weighting of the i-th location whose value is determined by the distance between observation locations based on coordinates (longitude, latitude).

The initial stage of GWR modeling is to determine the optimum bandwidth value that minimizes the value of cross validation (CV).

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(b)]^2$$

In this study, the weighting function used is the Gaussian kernel function:

$$W(u_i, v_i) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right]$$

dij: Distance between i-th and j-th
b: Optimum bandwidth.
The Analysis of Variance (ANOVA) is used to determine the effectiveness of the GWR model on classical regression.

2.6. Model Assumption Test
To check the assumptions of the first model, the Kolmogorov Smirnov statistical test is used (Lilliefors, 1967).

Several test methods can be used to determine the spatial effects (spatial dependence and spatial diversity on the data). In this study, the spatial dependence test used the Lagrange multiplier test, while the Breusch-Pagan test was used to test the spatial diversity. The spatial dependence is tested by the Lagrange Multiplier test (Anselin, 1988).

The Lagrange Multiplier (LM) statistic:

$$LM = E^{-1} \{(Ry)^2 T - 2R_yR_eT + (D+T)\} \sim \chi^2_{(q)}$$

Where:

Ry: $\mathbf{e}^T\mathbf{Wy}/\sigma^2$

R_e: $\mathbf{e}^T\mathbf{W}_e/\sigma^2$

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

T_{ij}: tr $\{\mathbf{W}_i\mathbf{W}_j + \mathbf{W}_i^T\mathbf{W}_j\}$

D: $\sigma^{-2}(\mathbf{WX\beta}) \, T \, M \, (\mathbf{WX\beta})$

E: $(D+T) \, T - (T)^2$

q: Number of spatial parameters

$T = $ tr $\{(W^T + W) \, W\}$

LM test criteria: $\begin{cases} \leq \chi^2(\mathbf{q}), & \mathbf{H_0} \text{ is accepted} \\ > \chi^2(\mathbf{q}), & \mathbf{H_0} \text{ is rejected} \end{cases}$

To test the spatial diversity, the Breusch-Pagan test is used (Breusch & Pagan, 1979). The tested hypotheses are:

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (same variance)

H_1: There is at least one $\sigma_i^2 \neq \sigma^2$ (there is variation between regions)

The Breusch-Pagan (BP) test statistic is:

$BP = (1/2) \, \mathbf{h}^T\mathbf{Z} \, \mathbf{Z}^T\mathbf{Z}^{-1}\mathbf{Z}^T\mathbf{h} \sim \chi^2_{(p)}$

The vector element $\mathbf{h}$ is:

$$h_i = \left(\frac{e_i^2}{\sigma^2}\right) - 1$$

ei is the square of the error for the i-th observation and Z is the y vector of $n \times 1$ that has been standardized for each observation.

BP test criteria $\begin{cases} \leq \chi^2(\mathbf{q}), & \mathbf{H_0} \text{ is accepted} \\ > \chi^2(\mathbf{q}), & \mathbf{H_0} \text{ is rejected} \end{cases}$

The Results of the Research and the Discussion

The classical regression model is an appropriate approach when the observed responses are independent and normally distributed. This model can be used to model poverty data based on the percentage of poor people as the response variable that has a normal distribution as indicated by the significance value (p-value) of the Kolmogorov-Smirnov, which is greater than 0.05 (Figure 1). Therefore, at the initial stage, the classical regression model is used to see how

well this model explain the pattern of the relationship between the responses, namely the percentage of poor people (y) and the explanatory variables consisting of the open unemployment rate (x1), the percentage of the illiterate population (x2), the percentage of the population working in the informal sector (x3), the maximum education level is junior high school/equivalent (x4), the underemployment rate (x5), the gross regional domestic product of primary sector (x6), and income inequality of Gini coefficient (x7).
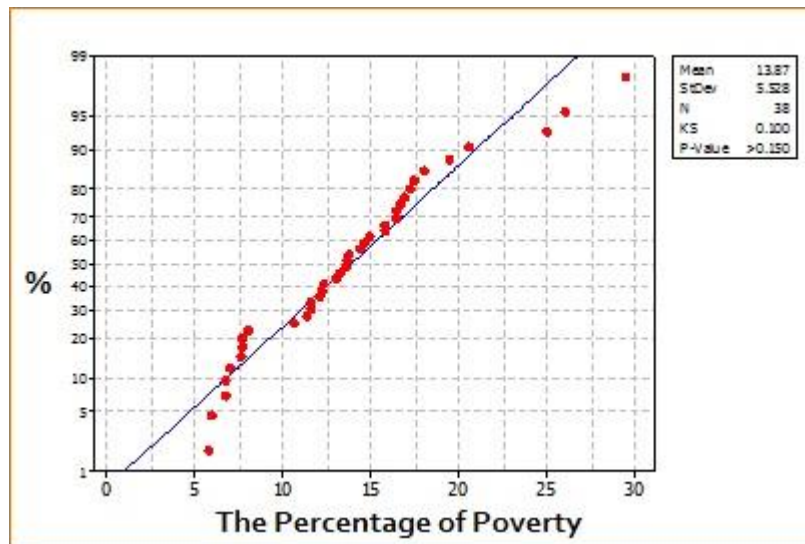

Figure 1. The Normality Test of the Variable Percentage of Poor Population Using Kolmogorov-Smirnov

Furthermore, based on the results of the model estimation parameters (Table 1), this model can explain 80% of the percentage diversity of the poor indicated by an R2 value of 0.80. However, judging from the significance value of the t-test, only two of the seven explanatory variables that have significant relationship at 0.05, namely the percentage of the illiterate population (x2) and the underemployment rate (x5). Besides, multicollinearity presents among the explanatory variables as indicated by the VIF of the percentage of the population working in the informal sector whose value is more than 10. This result shows that the classical regression model with seven explanatory variables is not the best model.

Table 1. Alleged Classical Regression Model with Six Explanatory Variables

Explanatory Variables	Coefficient	Standard Error	$t_{observed}$	p-value	VIF
Constant	9.508	9.816	0.97	0.340	
x1	-0.251	0.903	-0.28	0.783	2.1
x2	0.512	0.134	3.83	0.001**	3.8
x3	0.026	0.098	0.27	0.789	13.4
x4	-0.006	0.110	-0.05	0.957	9.2
x5	0.209	0.095	2.19	0.036*	6.4
x6	-0.104	0.074	-1.41	0.170	7.5
x7	-19.900	18.160	-1.10	0.282	1.6
R2: 0.800	R2(adj): 0.753				

Significance * 0.05 ** 0.01

At the next stage, the best regression model was selected using the Backward method (see Appendix 3). Based on the best selected classical regression model, two explanatory variables that had significant effects were the percentage of the illiterate population (x2) and the underemployment rate (x5). By using only two explanatory variables, this model can explain 77.8% of the dependent variable variability. The remaining 22.2% was explained by other variables outside the model. Since the VIF of each variable below 10, the two independent variables that build this model were also free from multicollinearity.

Table 2. The Best Classical Regression Model with Two Explanatory Variables

Explanatory Variable	Coefficient	Standard Error	$t_{observed}$	P-value	VIF
Constant	3.102	1.379	2.25	0.031*	
x2	0.490	0.117	4.19	0.000**	3.0
x5	0.158	0.064	2.47	0.019*	3.0
R2: 0.778	R2(adj): 0.765	AIC :192.41			

Significance * 0.05 ** 0.01

Classical Assumption Test

After obtaining the best classical regression model, it was necessary to test the model assumptions to see the feasibility of the model. The tests consisted of the normality of error, freedom of error, and homogeneity of variance error. The normality of error was estimated using the Kolmogorov-Smirnov test (Figure 2), where the significance value was above 0.15. The result indicated that the error distribution had met the normality assumption.
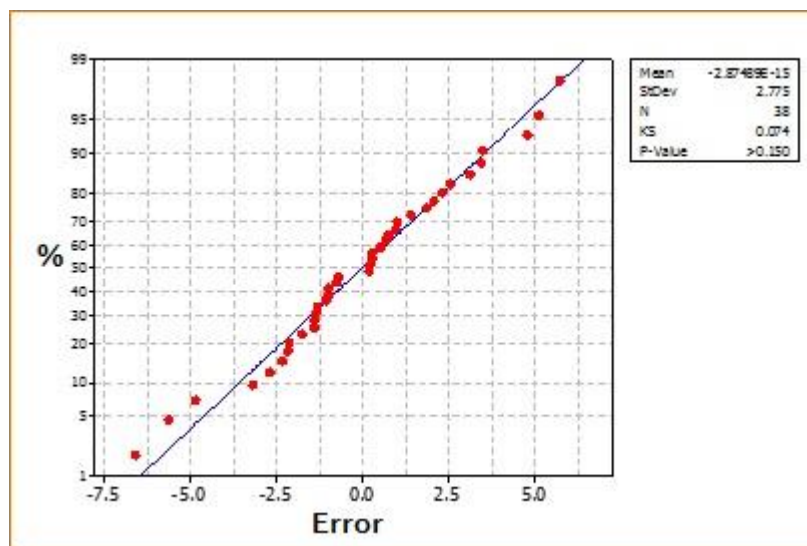

Figure 2. The Normality of Error Test on the Best Classical Regression Model Using Kolmogorov-Smirnov

The freedom of error was tested using the Durbin Watson (DW) test. The obtained DW value was 1.55 with a significance value of 0.05. This result shows that the freedom of error assumption of the classical regression was not fulfilled. On the other hand, the Breusch-Pagan (BP) statistic values of 6.42 and 0.04 indicated that there was a violation of the assumption of homogeneity variance error. Since it does not fulfill some of the assumptions, the classical

regression model is not considered appropriate for modeling poverty of East Java's cities and regencies.

1.   **Spatial Regression Model**
      Identifying Spatial Influence
      The assumption violation of the freedom of error and the homogeneity of variance error in classical regression of regional observations shows that there is a spatial influence on spatial dependence and spatial heterogeneity that have not been handled in the model.

      Moran Index statistic is used to generally identify spatial dependence. Based on Appendix 4, the Moran Index statistic value is 0.0978 with a significance value of 1.341e-05. This result shows that there is a spatial dependence of the poor people percentage in adjacent areas. This spatial dependence is also supported by the local Moran Index statistics (Appendix 5) and the Moran Index's scatter diagram plot (Appendix 6).

      Slightly different from the Moran Index, the Lagrange Multiplier (LM) test is used to identify spatial dependencies, either the spatial dependence in lag or error. From the results of the LM test (Table 3), the statistical significance value of the LM-SEM model (LMerr) is 0.028. It indicates that there is a spatial dependence of error, so that the formation of the SEM model can be done. Furthermore, the statistical significance values of LM-SAR model (LMerr) and GSM model are more than 0.05, which indicates that there is no spatial lag dependence or the combination between error and lag. Therefore, there is no need to build SAR and GSM models.

Table 3. Spatial Dependency Test with Lagrange Multiplier

Model	Statistics	Parameter	P-value
LMerr	4.858	1	0.028 *
LMlag	1.997	1	0.158
GSM	4.859	2	0.088

Significance * 0.05

2.   *Spatial Error Regression Model (SEM)*
      The Estimation of SEM Model Parameters
      SEM model is a spatial approach to overcome spatial effects, especially error dependence. This model can handle the dependence of error, as shown by the coefficient of 0.828, which is higher than the coefficient of determination of the classical regression model (Table 4). Besides explaining 82.8 percent of the variance in the percentage of poor people, this model's AIC and log likelihood are lower than the classical regression model (187.83 and -88.92). By judging the estimated model parameters, all SEM parameters are at the significant level of 0.05. In other words, the two explanatory variables (the percentage of the illiterate population and the underemployment rate) statistically have significant influences on the percentage of the poor population.

Table 4. Estimated Parameters of the SEM Model with Two Explanatory Variables

Explanatory Variables	Coefficient	Error Standard	Z-value	P-value
Konstanta	3.249	1.278	2.54	0.011*
x2	0.488	0.121	4.05	0.000**
x5	0.151	0.057	2.62	0.009**

Lambda	0.339	0.108	3.14	0.002**

R^2: 0.828   AIC: 187.83

Significance * 0.05 ** 0.01

SEM Model Assumption

Similar with the classical regression model, several classical assumptions must also be met in the SEM model, including the normality of error, freedom of error, and homogeneity of variance of error. Based on the error distribution plot of the SEM model (Figure 3), this model's Kolmogorov-Smirnov statistical significance value is greater than 0.15. Therefore, the SEM error flfills the normality assumption.
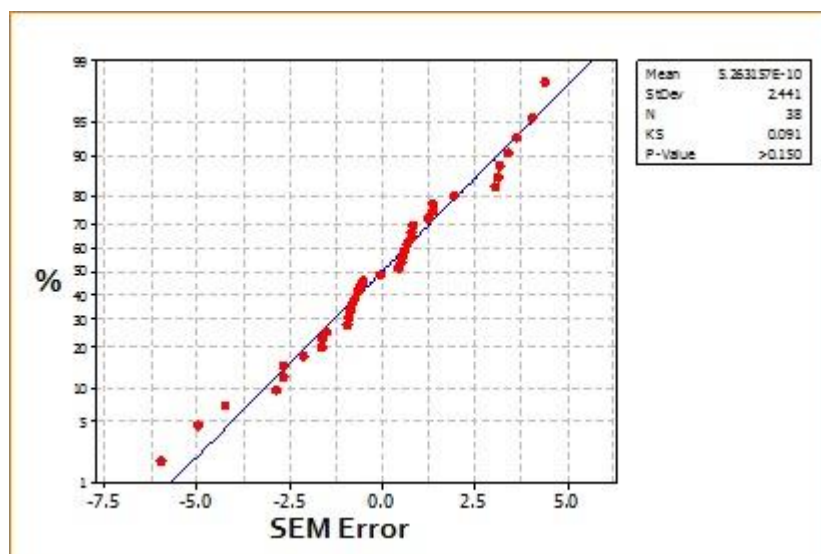

Figure 3. The Normality Test on SEM model error with Two Explanatory Variables using Kolmogorov-Smirnov

The examination of the assumption of freedom on SEM model error can be seen from the Moran Index statistics, which is 0.065 and the significance value of 0.474. It means that SEM model error fulfills the assumption of freedom or there is no spatial dependence. Besides, the Breusch-Pagan statistic's value is 4.31 with a significance value of 0.116. It indicates that the SEM model error has fulfilled the assumption of homogeneity of variance error.

By fulfilling several classical assumptions, the SEM model with two explanatory variables is considered capable of overcoming the spatial effects, both error dependence and heterogeneity of variance error. Thus, this model is suitable to model East Java's poverty data.

1.   Geographically Weighted Regression Model (GWR)
GWR Model Parameter Estimation

Geographically Weighted Regression Model or GWR is an approach to overcome the variety of errors caused by spatial influences. GWR is basically a development of the classical regression model into a geographically weighted regression model. The classical regression model produces estimates of global parameters that are generally applicable to all observed locations. However, the GWR model produces local parameter estimates in each observed location.

The initial stage of GWR modeling is to determine the optimum bandwidth value that minimizes the cross validation (CV) value using the Gaussian Kernel weighting function. After a certain number of iterations, the minimum CV is 269.42 and the optimum bandwidth is 0.4584. With this optimum bandwidth value, the parameter estimation of the GWR model is carried out. The summary of the estimated results is displayed in Table 5.

Table 5. The Summary of GWR Model Parameter Estimates

Variable	Min.	1st Qu.	median	3rd Qu.	Max.	Global
Konstanta	0.637	2.605	3.016	4.055	13.590	3.102
x2	0.280	0.532	0.580	0.659	0.774	0.490
x5	-0.325	0.099	0.143	0.165	0.266	0.159
R^2: 0.904	AIC: 163.81					

Based on the Anova analysis on the effectiveness of the GWR model on the classical regression model (Table 6), a significance value of 0.026 was obtained. It means that the Geographically Weighted Regression model is more effective in describing the relationship between the response variables and the explanatory variables.

Table 6. The Effectiveness Variance Analysis of the GWR on Classical Regression

Sources	df	SS	MS	$F_{observed}$	P-value
OLS Residuals	3.000	285.01			
GWR Improvement	11.807	162.21	13.739		
GWR Residuals	23.193	122.80	5.295	2.595	0,026

GWR Model Assumption Check

Kolmogorov Smirnov's statistical significance value of the GWR error is greater than 0.15. It means that the GWR error has fulfilled the normality assumption.
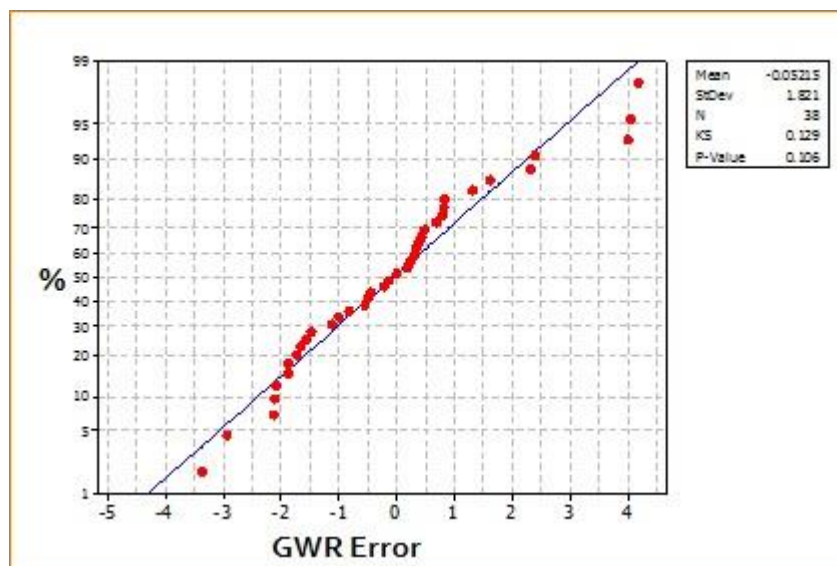

Figure 4. The Normality Test of GWR Model Error with Two Explanatory Variables Using Kolmogorov-Smirnov

The assumption of freedom of the SEM model error can be seen from the Moran Index statistics. The result of the calculation is 0.065 with a significance value of 0.474. It means that the SEM model error fulfills the assumption of freedom or there is no spatial dependence. Furthermore, the Breusch-Pagan statistic's value is 4.31 with a significance value of 0.116. It indicates that the SEM model error has met the assumption of homogeneity of variance.

2.   The Comparison of Classical Regression Model, SEM, and GWR

The coefficient determination value of the GWR model is higher than the classical regression model and SEM. It indicates that this model is better in explaining the poor people percentage diversity as a response than the classical regression model and the SEM model. Also, the GWR model's low AIC statistics indicates that this model can reduce the spatial effect of the data observed regionally.

Conclusion and Suggestion

Based on the results and discussions, the following conclusions can be drawn: 1) the spatial regression approach is considered very appropriate to be used to model the relationship pattern between the response and the explanatory variables when the observed data has a spatial effect caused by the proximity between the observation areas; 2) The effect of spatial dependence of error on observational data can be overcome by using the Spatial Error Regression Model (SEM), while the effect of spatial variance heterogeneity can be overcome by the Geographically Weighted Regression model (GWR).

If the spatial dependence and spatial heterogeneity influence the observation data simultaneously, a hybrid model from several models is considered worthy to be applied. In the case when the observation data has the effect of dependence and spatial heterogeneity at the same time, the use of a hybrid model from the SEM and GWR models is recommended for further research.

References

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*(1), 1–17.

Anselin, L. (2009). Spatial regression. *The SAGE Handbook of Spatial Analysis*, *1*, 255–276.

Anselin, L., & Florax, R. (2012). *New directions in spatial econometrics*. Springer Science & Business Media.

Anselin, L., & Kelejian, H. H. (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review*, *20*(1–2), 153–182.

Arbia, G., & Baltagi, B. H. (2008). *Spatial econometrics: Methods and applications*. Springer Science & Business Media.

Atkinson, A. B. (1987). On the measurement of poverty. *Econometrica: Journal of the Econometric Society*, 749–764.

Baltagi, B. H., & Li, D. (2001). LM tests for functional form and spatial error correlation. *International Regional Science Review*, *24*(2), 194–225.

Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data*

analysis with R (Vol. 747248717). Springer.

Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287–1294.

Charlton, M., Fotheringham, S., & Brunsdon, C. (2009). Geographically weighted regression. *White Paper. National Centre for Geocomputation. National University of Ireland Maynooth*, *2*.

Djuraidah, A., & Wigena, A. H. (2012). Regresi Spasial untuk Menentuan Faktorfaktor Kemiskinan di Provinsi Jawa Timur. *Statistika*, *12*(1).

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons.

Ferezagia, D. V. (2018). Analisis Tingkat Kemiskinan di Indonesia. *Jurnal Sosial Humaniora Terapan*, *1*(1).

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.

Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, *40*(11), 2751–2769.

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*(1), 53–67.

Lawson, A. B. (2013). *Statistical methods in spatial epidemiology*. John Wiley & Sons.

Lee, L., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*(2), 165–185.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2). Duxbury press Belmont, CA.

Nurwati, N. (2008). Kemiskinan: Model pengukuran, permasalahan dan alternatif kebijakan. *Jurnal Kependudukan Padjadjaran*, *10*(1), 1.

Putra, A. S., Tong, G., & Pribadi, D. O. (2020). Spatial Analysis of Socio-Economic Driving Factors of Food Expenditure Variation between Provinces in Indonesia. *Sustainability*, *12*(4), 1638.

Rinaldi, A., Djuraidah, A., Wigena, A. H., Mangku, I. W., & Gunawan, D. (2017). Spatial Extreme Models with Copula to Determine Extreme Rainfall Zone. *Applied Mathematical Sciences*, *11*(27), 1327–1336. https://doi.org/10.12988/ams.2017.74124

Schabenberger, O., & Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. CRC press.

Tasyurek, M., & Celik, M. (2020). RNN-GWR: A geographically weighted regression approach for frequently updated data. *Neurocomputing*, *399*, 258–270.

Timm, N. H. (2002). *Applied multivariate analysis*. Springer.

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*(1), 36–59.

Weisberg, S. (2013). *Applied linear regression*. John Wiley & Sons.

Zhang, Y., Zhang, D., & Miller, E. J. (2021). Spatial Autoregressive Analysis and Modeling of Housing Prices in City of Toronto. *Journal of Urban Planning and Development*, *147*(1), 5021003.

Zhou, Q., Wang, C., & Fang, S. (2019). Application of geographically weighted regression (GWR) in the analysis of the cause of haze pollution in China. *Atmospheric Pollution Research*, *10*(3), 835–846.

Zhu, C., Zhang, X., Zhou, M., He, S., Gan, M., Yang, L., & Wang, K. (2020). Impacts of urbanization and landscape pattern on habitat quality using OLS and GWR models in Hangzhou, China. *Ecological Indicators*, *117*, 106654.