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

Achi Rinaldi¹, Yuni Susianto², Budi Santoso³, Wahyu Kusumaningtyas⁴

¹ Universitas Islam Negeri Raden Intan Lampung, Indonesia
² Badan Pusat Statistik, Indonesia
³ STKIP Kumala Metro, Indonesia

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.

Copyright (c) 2021 Al-Jabar : Jurnal Pendidikan Matematika
The Research Methods

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

\[ E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \]  
\[ \text{Var}(Y|X) = \sigma^2 \]

\( \beta_i \) for \( i = 0, 1, 2, \ldots, p \) or \( \sigma^2 \) 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:

\[
Y = \begin{pmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{pmatrix}, \quad X = \begin{pmatrix}
  1 & x_{11} & \cdots & x_{1p} \\
  1 & x_{21} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  1 & x_{n1} & \cdots & x_{np}
\end{pmatrix}
\]

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

\[
\beta = \begin{pmatrix}
  \beta_0 \\
  \beta_1 \\
  \vdots \\
  \beta_p
\end{pmatrix}, \quad e = \begin{pmatrix}
  e_1 \\
  e_2 \\
  \vdots \\
  e_n
\end{pmatrix}
\]

The multiple regression matrix can be written as follows:

\[ Y = X\beta + \epsilon \]  

(2)

Assuming the model: \( \epsilon \overset{iid}{\sim} N(0, \sigma^2 I_n) \)

The least squares method is used to estimate the parameters \( \beta \) by minimizing the sum of the residual squares.

\[
JK(\epsilon) = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - x_i^T \beta)^2
\]

By deriving the above equation with the parameter \( \beta \), the estimator will be:

\[
\hat{\beta} = (X'X)^{-1}(X'Y) \quad \text{whereas for} \quad \text{Var}(\hat{\beta}) = (X'X)^{-1}
\]

The assumptions in the classical regression model are:

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

\[
E(y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}
\]

2. \( \text{Var}(\epsilon_i) = \sigma^2 \), for \( i = 1, 2, \ldots, n \), or equal to \( \text{Var}(y_i) = \sigma^2 \)

3. \( \text{Cov}(\epsilon_i, \epsilon_j) = 0 \), for \( i \neq j \).
General Spatial Regression Model (GSM)

The general model of spatial regression is:

\[ Y = \rho W y + X\beta + u \]  \hspace{1cm} (3)

\[ u = \lambda W u + \varepsilon \]  \hspace{1cm} (4)

\[ \varepsilon \sim N(0, \sigma^2 I) \]

y is the dependent variable of \( n \times 1 \), X is the matrix of independent variables of \( (n \times (p+1)) \), \( \beta \) is the vector of regression parameter coefficients of \( p \times 1 \), \( \rho \) is a spatial lag autoregression coefficient, \( \lambda \) 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:

\[ y - \rho W y = X\beta + u \] or 

\[ I - \rho W y = X\beta + u \] \hspace{1cm} (5)

Equation (3) can be expressed as:

\[ (I - \rho W) u = \varepsilon \] or \[ u = (I - \rho W)^{-1}\varepsilon \] \hspace{1cm} (6)

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

\[ (I - \rho W) y = X\beta + (I - \rho W)^{-1}\varepsilon \]

\[ (I - \rho W)^{-1}\varepsilon = (I - \rho W) y - X\beta \]

If all sides are multiplied by \( (I - \rho W) \), then:

\[ \varepsilon = (I - \rho W)^{-1}(I - \rho W) y - X\beta \] \hspace{1cm} (7)

The value of the probability function of the variable \( \varepsilon \) is:

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

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

\[ L (\sigma^2; \varepsilon) = c(\varepsilon)\sigma^{2n} \exp \left[ \frac{1}{2\sigma^2} \varepsilon^t \varepsilon \right] \] \hspace{1cm} (9)

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

\[ J = \left| \frac{\partial \varepsilon}{\partial y} \right| |I - \rho W| |I - \rho W| \]

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

\[ L (\rho, \lambda, \sigma^2, \beta; y) = c(y) (\sigma^2)^{-n/2} |I - \lambda W| |I - \rho W| \]

\[ \exp \left[ -\frac{1}{2\sigma^2} \begin{bmatrix} (I - \rho W)(I - \rho W) y - X\beta \end{bmatrix}^T (I - \rho W) \begin{bmatrix} (I - \rho W)(I - \rho W) y - X\beta \end{bmatrix} \right] \]

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

\[ \ln L (\rho, \lambda, \sigma^2, \beta; y) = \ln(c(y)) - n/2 \ln(\sigma^2) + \ln|I - \lambda W| + \ln|I - \rho W| \]

\[ -\frac{1}{2\sigma^2} \begin{bmatrix} (I - \rho W)(I - \rho W) y - X\beta \end{bmatrix}^T (I - \rho W) \begin{bmatrix} (I - \rho W)(I - \rho W) y - X\beta \end{bmatrix} \]

Suppose the square of the weighting matrix \( (I - \rho W)^T(I - \rho W) \) is denoted as \( \Omega \) and estimator \( \beta \) is obtained by maximizing the log probability function in equation (10), then the parameter estimator \( \beta \) is:

\[ \hat{\beta} = (X'\Omega X)^{-1}X'\Omega(I - \lambda W)Y \]
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:

$$
y = \rho W y + X\beta + \varepsilon \quad (11)
$$

$\varepsilon \sim N(0, \sigma^2)$

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), $\varepsilon_i$ is assumed to be normally distributed, stochastically free, identical, with a mean of zero, and a variance of $\sigma^2$. $\varepsilon_i$ is the error at the location of $i$.

The probability density function of $\varepsilon_i$:

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

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

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

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

$$\varepsilon = y - \rho W y - X\beta
$$

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

$$f(y) = f(\varepsilon) |J|
$$

$$= \exp \left[ -\frac{\varepsilon^T \varepsilon}{2\sigma^2} \right] |I - \rho W|
$$

The probability function of the response variable $y$:

$$L(\beta, \rho, \sigma^2; y) = f(y; \beta, \rho, \sigma^2)
$$

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

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).

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

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

The estimation for $\sigma^2$, $\beta$, and $\rho$ is obtained by maximizing the log probability function in equation (13). The estimator for $\sigma^2$ is:

$$\hat{\sigma}^2 = \frac{(y - \rho W y - X\beta)^T(y - \rho W y - X\beta)}{n} \quad (14)$$
Equation (14) can be written as:
\[
\sigma^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{JKG}{n}
\]
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 \(\beta\) is:
\[
\hat{\beta} = (X^TX)^{-1}X^Ty
\]
and the estimator for \(\rho\) is:
\[
\hat{\rho} = (y^TW^Wy)^{-1}y^TW^y
\]
2.4. Error Spatial Regression Model (SEM)
If \(\rho=0\) and \(\lambda\neq0\), 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:
\[
y = X\beta + u, \ u = \lambda Wu + \varepsilon
\]
where is assumed \(\varepsilon \sim N(0, \sigma^2 I)\)
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 \(\varepsilon_i\):
\[
f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-\varepsilon_i^2}{2\sigma^2}\right]
\]
Joint probability density function of \(n\) random variables 1, 2, ..., \(n\)
\[
f(\varepsilon) = f(\varepsilon_1). f(\varepsilon_2) \ldots f(\varepsilon_n)
\]
\[
= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n} \varepsilon_i^2}{2\sigma^2}\right)
\]
\[
= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T \varepsilon}{2\sigma^2}\right]
\]
The density function with the response variable \(y\) is obtained by transforming an \(n\)-dimensional \(\varepsilon\) space into an \(n\)-dimensional \(y\) space. From equation (15), it can be obtained:
\[
u = y - X\beta \text{ dan}
\]
\[
\varepsilon = (I - \lambda Wu) u
\]
Therefore, \(\varepsilon = (I - \lambda Wu) (y - X\beta)\)
The joint probability density function of \(n\) response variables \(y\):
\[
f(y) = f(\varepsilon)|I|
\]
\[
= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T \varepsilon}{2\sigma^2}\right] |I - \lambda W|\]
The probability function of the variable \(y\):
\[
L (\beta, \lambda, \sigma^2; y) = f(y; \beta, \lambda, \sigma^2)
\]
\[
= (2\pi\sigma^2)^{-\frac{n}{2}} |I - \lambda W| \exp\left[-\frac{(I - \lambda W)(y - X\beta)^T (I - \lambda W)(y - X\beta)}{2\sigma^2}\right]
\]
The logarithm of the probability function above is:
\[
\ln (L (\beta, \lambda, \sigma^2; y))
\]
\[= \ln\left( \frac{2\pi \sigma^2}{n} \right) - \ln |I - \lambda W| \exp \left\{ - \frac{\left( (I - \lambda W)(y - X\hat{\beta}) \right)^T (I - \lambda W)(y - X\hat{\beta})}{2\sigma^2} \right\} \]

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

The estimation for \( \sigma^2 \), \( \beta \) and \( \lambda \) is obtained by maximizing the log-likelihood function in equation (17).

The estimator for \( \sigma^2 \) 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 \( \beta \) 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 \( \lambda \) parameter, a numerical iteration is needed to get an estimator for \( \lambda \) that maximizes the probability log function (Bivand et al., 2008).

2.5. Spatial Weighted Regression Model (GWR)

The GWR model equation is:

\[y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{P} \beta_k(u_i, v_i)x_{ik} + \epsilon_i \] (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:

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

\(W(u_i, v_i)\) is a diagonal weighting matrix of \(n \times n\), 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).

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

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:

\[ LM = E^{-1} \{(Ry)^2 T - 2RyR e_T + (D+T)\} \sim \chi^2(q) \]

Where:

- \( Ry = e^T Wy / \sigma^2 \)
- \( Re = e^T W e / \sigma^2 \)
- \( M = I - X(X^T X)^{-1} X^T \)
- \( T_{ij} = \text{tr}\{W_iW_j + W_i^T W_j\} \)
- \( D = \sigma^{-2}(WX\beta)TM(WX\beta) \)
- \( E = (D+T)T - (T)^2 \)
- \( q: \text{Number of spatial parameters} \)
- \( T = \text{tr}\{(W^T + W)W\} \)

LM test criteria:

\[ \begin{cases} \leq \chi^2(q), & H_0 \text{ is accepted} \\ > \chi^2(q), & H_0 \text{ is rejected} \end{cases} \]

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)
- \( H_1: \text{There is at least one } \sigma_i^2 \neq \sigma^2 \) (there is variation between regions)

The Breusch-Pagan (BP) test statistic is:

\[ BP = (1/2) h^T Z Z^T Z^{-1} h \sim \chi^2(p) \]

The vector element \( h_i \) 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 x 1 that has been standardized for each observation.

BP test criteria:

\[ \begin{cases} \leq \chi^2(q), & H_0 \text{ is accepted} \\ > \chi^2(q), & H_0 \text{ is rejected} \end{cases} \]
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

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t_observed</th>
<th>p-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.508</td>
<td>9.816</td>
<td>0.97</td>
<td>0.340</td>
<td></td>
</tr>
<tr>
<td>x1</td>
<td>-0.251</td>
<td>0.903</td>
<td>-0.28</td>
<td>0.783</td>
<td>2.1</td>
</tr>
<tr>
<td>x2</td>
<td>0.512</td>
<td>0.134</td>
<td>3.83</td>
<td>0.001**</td>
<td>3.8</td>
</tr>
<tr>
<td>x3</td>
<td>0.026</td>
<td>0.098</td>
<td>0.27</td>
<td>0.789</td>
<td>13.4</td>
</tr>
<tr>
<td>x4</td>
<td>-0.006</td>
<td>0.110</td>
<td>-0.05</td>
<td>0.957</td>
<td>9.2</td>
</tr>
<tr>
<td>x5</td>
<td>0.209</td>
<td>0.095</td>
<td>2.19</td>
<td>0.036*</td>
<td>6.4</td>
</tr>
<tr>
<td>x6</td>
<td>-0.104</td>
<td>0.074</td>
<td>-1.41</td>
<td>0.170</td>
<td>7.5</td>
</tr>
<tr>
<td>x7</td>
<td>-19.900</td>
<td>18.160</td>
<td>-1.10</td>
<td>0.282</td>
<td>1.6</td>
</tr>
</tbody>
</table>

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 \((x_2)\) and the underemployment rate \((x_5)\). 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>(t_{\text{observed}})</th>
<th>P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.102</td>
<td>1.379</td>
<td>2.25</td>
<td>0.031*</td>
<td></td>
</tr>
<tr>
<td>(x_2)</td>
<td>0.490</td>
<td>0.117</td>
<td>4.19</td>
<td>0.000**</td>
<td>3.0</td>
</tr>
<tr>
<td>(x_5)</td>
<td>0.158</td>
<td>0.064</td>
<td>2.47</td>
<td>0.019*</td>
<td>3.0</td>
</tr>
<tr>
<td>R2: 0.778</td>
<td>R2(adj): 0.765</td>
<td>AIC :192.41</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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](image)

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>
<thead>
<tr>
<th>Model</th>
<th>Statistics</th>
<th>Parameter</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMerr</td>
<td>4.858</td>
<td>1</td>
<td>0.028 *</td>
</tr>
<tr>
<td>LMI lag</td>
<td>1.997</td>
<td>1</td>
<td>0.158</td>
</tr>
<tr>
<td>GSM</td>
<td>4.859</td>
<td>2</td>
<td>0.088</td>
</tr>
</tbody>
</table>

   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>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Coefficient</th>
<th>Error Standard</th>
<th>Z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Konstanta</td>
<td>3.249</td>
<td>1.278</td>
<td>2.54</td>
<td>0.011*</td>
</tr>
<tr>
<td>x2</td>
<td>0.488</td>
<td>0.121</td>
<td>4.05</td>
<td>0.000**</td>
</tr>
<tr>
<td>x5</td>
<td>0.151</td>
<td>0.057</td>
<td>2.62</td>
<td>0.009**</td>
</tr>
</tbody>
</table>
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 fulfills the normality assumption.

![Figure 3. The Normality Test on SEM model error with Two Explanatory Variables using Kolmogorov-Smirnov](image)

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>median</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>Konstanta</td>
<td>0.637</td>
<td>2.605</td>
<td>3.016</td>
<td>4.055</td>
<td>13.590</td>
<td>3.102</td>
</tr>
<tr>
<td>x2</td>
<td>0.280</td>
<td>0.532</td>
<td>0.580</td>
<td>0.659</td>
<td>0.774</td>
<td>0.490</td>
</tr>
<tr>
<td>x5</td>
<td>-0.325</td>
<td>0.099</td>
<td>0.143</td>
<td>0.165</td>
<td>0.266</td>
<td>0.159</td>
</tr>
</tbody>
</table>

R²: 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

<table>
<thead>
<tr>
<th>Sources</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F&lt;sub&gt;observed&lt;/sub&gt;</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS Residuals</td>
<td>3.000</td>
<td>285.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWR Improvement</td>
<td>11.807</td>
<td>162.21</td>
<td>13.739</td>
<td></td>
<td>0.026</td>
</tr>
<tr>
<td>GWR Residuals</td>
<td>23.193</td>
<td>122.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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](image)
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


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


