Live Expectancy Modelling using Spatial Durbin Robust Model

Arief Rachman Hakim, Budi Warsito, Hasbi Yasin,
Statistics Department of the Faculty of Science and Mathematics, Diponegoro University
Email: arief.rachman@live.undip.ac.id

Abstract. Spatial regression model is a model used to determine the relationship between response variables and predictor variables that have spatial influence in them. If the two variables have spatial influence, then the model that will be formed is the Spatial Durbin Model. One of the causes of the inaccuracy of the spatial regression model in predicting is observations of outliers. Removing outliers in spatial analysis can change the composition of spatial effects on the data. One method of settlement due to outliers in the spatial regression model is to use robust spatial regression. The application of the M-estimator parameter estimator principle is done in estimating the coefficient of spatial regression parameters that are robust to outliers. The results of modelling by applying the principle of M-estimator estimator on estimating the robust Spatial Durbin Model regression parameters are expected to be able to accommodate the existence of outliers in the spatial regression model. One example of the application of the Spatial Durbin Model Robust is the case of life expectancy modeling.

Keywords: Outlier, M-estimator, Spatial Durbin Model robust, Spatial Regression

1. Introduction
Regression analysis is a statistical analysis technique that aims to see the relationship between the response variable and the predictor variable so that it is able to predict the response variable if the predictor variable is known. When spatial (location) effects are found in data commonly referred to as spatial data, the regression analysis used is spatial regression. Modeling using spatial regression is one of them is the Spatial Durbin Model (SDM). SDM model is a spatial regression model that shows the existence of spatial effects in the response variable and predictor variables[1]. The existence of outliers (outliers) in the research data makes the estimation of parameters to be biased, to analyze data contaminated by outliers can be used a statistical analysis method that is robust regression. One estimation method in robust regression is M-estimator which is the simplest robust regression method both theoretically and computationally[2]. Besides found outliers, in the research data there will be cases where the data includes spatial data types, outliers in spatial data can be detected by the Moran’s Scatterplot method. To overcome the outliers in spatial data, a method called Spatial Durbin Model Robust was developed. One application of the model is in the case of Life Expectancy

2. Literature review
2.1 Spatial Regression
The concept of spatial regression is a development of the linear regression model with the addition of area-based spatial weights, spatial weighting is used to determine the weights between locations that are observed based on neighborhood relations between locations. Neighbors can be defined in several ways, there are [3]:

[1] [2] [3]
1. **Rook Contiguity**
   The area of observation is determined based on the sides that intersect and angles are not taken into account.

2. **Bishop Contiguity**
   The area of observation is determined based on the intersecting angles and sides are not taken into account.

3. **Queen Contiguity**
   The area of observation is determined based on the sides that intersect and the angle is also taken into account.

The general model of spatial regression can be written as follows\(^\text{[3]}\):

\[
y = \rho Wy + X\beta + u
\]

\[
u = \lambda Wu + \varepsilon \quad : \quad N(0, \sigma^2 I_n)
\]

with:
- \(y\) = Vector dependent variable size \(n \times 1\)
- \(\rho\) = Coefficient of spatial lag parameters of the dependent variable
- \(W\) = Spatial weighting matrix \(n \times n\)
- \(X\) = Matrix of independent variables measuring \(n \times (k+1)\)
- \(\beta\) = Vector coefficient of regression parameter size \((k+1) \times 1\)
- \(\lambda\) = Coefficient of spatial parameter error
- \(u\) = Error vector which has a spatial effect of size \(n \times 1\)
- \(\varepsilon\) = vector error of size \(n \times 1\)

From the general spatial regression model equation (1), several other models can be formed as follows\(^\text{[1]}\):

1. If \(\rho=0\) and \(\lambda=0\) then it is called the classic linear regression model with the formed equation is:
   \[
y = X\beta + \varepsilon
\]

2. If \(\rho\neq0\) and \(\lambda=0\) is called the Spatial Autoregressive Model (SAR) regression with the formed equation is:
   \[
y = \rho Wy + X\beta + \varepsilon
\]

3. If \(\rho=0\) and \(\lambda\neq0\) are called Spatial Error Model (SEM) regressions with the formed equation is:
   \[
y = X\beta + u
   \]

4. If \(\rho\neq0\) and \(\lambda\neq0\) are called Spatial Autoregressive Moving Average (SARMA) with the formed equation is:
   \[
y = \rho Wy + X\beta + u
   \]

\[u = \lambda Wu + \varepsilon\]

2.2 **Spatial Durbin Model (SDM)**

Spatial Durbin Model (SDM) is a spatial regression model that has a shape such as the Spatial Autoregressive Model (SAR) which has spatial lag in the response variable (Y) as in the equation (3). It's just that SDM has the characteristic spatial lag in the predictor variable (X)\(^1\). The SDM model has the following form of equation\(^\text{[12]}\):

\[
y = \rho Wy + \alpha I_n + X\beta + WX\theta + \varepsilon \quad : \quad N(0, \sigma^2 I_n)
\]

can be written as follows:

\[
y = \rho Wy + Z\delta + \varepsilon
\]
With: \(Z = [1_n \ X \ WX] \); \(\delta = \begin{bmatrix} \alpha \\ \beta \\ \theta \end{bmatrix}\)

- \(\alpha\) = Constant Parameter
- \(\theta\) = Vector spatial lag parameter predictor with size \(k \times 1\)
- \(1_n\) = Vector containing the number 1 measuring \(n \times 1\)

The likelihood function is formed through the error equation (\(\varepsilon\)) which is normally distributed as follows:

\[
y = \rho Wy + Z\delta + \varepsilon
\]

\[
\varepsilon = y - \rho Wy - Z\delta
\]

\[
\varepsilon = (1_n - \rho W)y - Z\delta
\]

The SDM model parameters can be estimated using the Maximum Likelihood Estimation (MLE) method, the SDM model parameter estimates are obtained as follows:

\[
\frac{1}{\lambda_{min}} < \rho < \frac{1}{\lambda_{max}}
\]

\[
\hat{\delta} = (Z^T Z)^{-1} Z^T y - \hat{\rho} (Z^T Z)^{-1} Z^T Wy = \delta_h - \hat{\rho} \hat{\delta}_d
\]

\[
\hat{\sigma}^2 = \frac{(e_h - \hat{\rho} e_d)^T (e_h - \hat{\rho} e_d)}{n}
\]

2.3 Testing Regression Models

2.3.1 Model Match Test

Testing the suitability of the SDM model, the following procedure is used to test the hypothesis \([5]\):

- \(H_0: \rho = \beta_j = \theta_j = 0\), with \(j = 1, 2, \ldots, k\)
- \(H_1: \rho \neq 0\) or at least one that \(\beta_j \neq 0, \theta_j \neq 0\), with \(j = 1, 2, \ldots, k\)

Statistics test:

\[
F_{Count} = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{MSR}{MSE}
\]

Decision: \(H_0\) is rejected if the value \(F_{Count} > F_{\alpha,k,n-k-1}\) or \(p-value < \alpha\).

2.3.2 Significance of Parameters Test

Testing the significance of the spatial modeling parameters using the Wald test \([1]\), to test the \(\rho\) parameters used the following hypothesis \([5]\):

- \(H_0: \rho = 0\) ;
- \(H_1: \rho \neq 0\)

Statistics Test: \(Wald_\rho = \frac{\hat{\rho}^2}{\text{var}(\hat{\rho})}\)  

(12)

For testing parameters of \(\beta\):

- \(H_0: \beta_j = 0\);
- \(H_1: \beta_j \neq 0, j = 1, 2, \ldots, k\)
Statistics Test: \( Wald_{\beta} = \frac{\hat{\beta}_j^2}{\text{var}(\hat{\beta}_j)} \) (13)

For parameter \( \theta \) use the following hypothesis:

\[ H_0 : \theta_j = 0 ; \]
\[ H_1 : \theta_j \neq 0 , j = 1,2,...,k \]

Statistics Test: \( Wald_{\theta} = \frac{\hat{\theta}_j^2}{\text{var}(\hat{\theta}_j)} \) (14)

Decision making criteria is \( H_0 \) is rejected if the value \( Wald > \chi^2_{\alpha,1} \)

### 2.4 Robust Regression

Robust regression is a regression method used when the distribution of residuals is not normal or there are some outliers that affect the model [6]. This method is an important tool for analyzing data contaminated with outliers and can provide results that are resistant to outliers. One estimation method in robust regression is Robust M-estimator.

#### 2.4.1 Robust M-Estimator

Robust M-estimator actually minimizes objective functions [6]:

\[
\min_{\beta} \sum_{i=1}^{n} \rho(u_i) = \min_{\beta} \sum_{i=1}^{n} \rho \left( \frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{s} \right) = \min_{\beta} \sum_{i=1}^{n} \rho \left( \frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{s} \right)
\]

where \( s \) is a robust estimation scale. The estimate \( s \) used is:

\[
s = \frac{\text{MAD}_{\text{u}}}{0.6745} = \frac{\text{median}[e] - \text{median}[e]}{0.6745}
\]

A value of 0.6745 makes \( s \) an unbiased estimator of \( \sigma \) if \( n \) is large and the error is normally distributed[7]. To get the parameter estimate by minimizing the equation (15).

With the first partial derivative of \( \rho \) with respect to \( \beta_j \) \((j = 0,1, ..., k)\) it is equal to 0, so that:

\[
\sum_{i=1}^{n} x_{ij} \left( \frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{s} \right) = 0, \quad j = 0,1,K ,k
\]

(17)

Given a solution by defining the weighting function:

\[
w(u_i) = \left( \frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{s} \right)
\]

(18)

and \( w_i = w(u_i) \). Then the estimation of equation (17) can be written:

\[
\sum_{i=1}^{n} x_{ij} w_i \left( y_i - \sum_{j=0}^{k} x_{ij} \beta_j \right) = 0, \quad j = 0,1,K k
\]

(19)

Equation (19) is solved by Iteratively Reweighted Least Square (IRLS). In matrix notation, equation (19) can be written:

\[
\hat{\beta} = (X^tWX)^{-1} X^t Wy
\]

(20)

Iteration stop if \( \hat{\beta}_j \) convergent, which is the difference in value \( \hat{\beta}_{j}^{(m+1)} \) and \( \hat{\beta}_{j}^{(m)} \) close to 0.

#### 2.4.2 Objective Functions

The function used to find the weighting function in robust regression is an objective function [8]. One of the weighting functions that can be used is the Tukey Bisquare weighting function as follows:
The $c$ value is called the tuning constant, and the tuning constant for the Tukey Bisquare weighting function in the M-estimator estimation method is $c = 4.685^{[8]}$.

### 2.5 Life Expectancy

According to the Indonesian Central Statistics Agency $^{[9]}$, Life Expectancy at an age $x$ is the average life year that will still be lived by someone who has succeeded in reaching age $x$ in a given year and in the prevailing mortality situation in his community. There are several factors that have a significant influence on Life Expectancy including education, health and economic factors. The average length of school variables are used as variables to explain educational factors. The variable percentage of households behaving clean and healthy and the number of posyandu used as variables to explain health factors. The variable percentage of poor population and adjusted per capita expenditure are used as variables to explain economic factors.

### 3. Research Methodology

#### 3.1. Data Sources and Research Variables

The data used in this study are secondary data obtained from the catalog of Central Java Province in Figures 2018 issued by the Central Statistics Agency (BPS) of Central Java Province and the 2017 Central Java Provincial Health Profile book issued by the Central Java Provincial Health Office. The observation units in this study were 35 districts and cities in Central Java Province. The variables used in this study are Life Expectancy data as response variables and average length of schooling ($x_1$), percentage of households with clean and healthy living behavior ($x_2$), number of posyandu ($x_3$), percentage of poor population ($x_4$) and adjusted per capita expenditure ($x_5$) as a predictor variable.

#### 3.2. Analysis steps

The analysis steps that will be carried out in this study are as follows:

1. Obtain AHH data along with five factors that are thought to influence it.
2. Determine the spatial weighting matrix based on queen contiguity.
3. Perform spatial autocorrelation test with Moran’s I test.
4. Detect outlier with Moran’s Scatterplot.
5. Estimating parameters with Robust Spatial Durbin Model M-estimator regression.
6. Conduct model compatibility test and parameter significance test.
7. Calculate adjusted coefficient of determination (Adjusted R2) and MSE

### 4. Result

#### 4.1 Parameter Estimate Regression Spatial Durbin Model Robust

To obtain the estimated spatial parameters of the spatial durbin model Robust, an estimate is made using the least squares method. The general Spatial Durbin Model equation is as follows:

$$ y = \rho Wy + Z\delta + \varepsilon $$

with:

$$ Z = [1_n, X, WX] $$

$$\delta = [\alpha, \beta, \theta]$$

\[(I - \rho W)y = Z\delta + \varepsilon\]

\[\varepsilon = (I - \rho W)y - Z\delta, \varepsilon \sim N(0, \sigma^2 I_n)\]

To obtain the estimated spatial parameters of the spatial durbin model Robust, an estimate is made using the least squares method. Estimation of parameters in the least squares method can be obtained by minimizing the amount of residual squares as follows:

\[\sum e_i^2 = \sum [(I - \rho W)y_i - Z_i\delta_i - Z_{i2}\delta_{i2} - K - Z_{i4}\delta_{i4}]^2\]  \hspace{1cm} (21)

In matrix notation can be written:

\[e^T e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = e_1^2 + e_2^2 + K + e_n^2 = \sum e_i^2\]  \hspace{1cm} (22)

From equation (9) substituted into equation (22), it is obtained:

\[e^T e = [(I - \rho W)y - Z\delta]T [(I - \rho W)y - Z\delta]\]

\[= [(I - \rho W)y]^T - \delta^T Z^T \left( (I - \rho W)y \delta Z^T + \delta^T Z^T Z \delta \right)\]

\[= [(I - \rho W)y]^T - (I - \rho W)y \delta^T Z^T - (I - \rho W)y \delta^T Z^T + \delta^T Z^T Z \delta\]

\[= [(I - \rho W)y]^T - 2(I - \rho W)y \delta^T Z^T + \delta^T Z^T Z \delta\]  \hspace{1cm} (23)

To minimize equation (21), it is differentiated from \(\delta^T\)

\[\frac{\partial}{\partial \delta^T} e^T e = \frac{\partial}{\partial \delta^T} [(I - \rho W)y]^T (I - \rho W)y - 2(I - \rho W)y \delta^T Z^T + \delta^T Z^T Z \delta\]

\[= 0 - 2(I - \rho W)y Z^T + 2Z^T Z \delta\]

\[= 2Z^T Z \delta - 2(I - \rho W)y Z^T + 2Z^T Z \delta\]

\[\delta_{ols} = (Z^T Z)^{-1} Z^T (I - \rho W)y\]  \hspace{1cm} (24)

So equation (9) can be written as:

\[\varepsilon = (I - \rho W)y - Z\delta_{ols}\]  \hspace{1cm} (25)

To get the effect function, equation (24) can be written as:

\[\hat{\delta}_{ols} = \left(Z^T \psi Z\right)^{-1} Z^T \psi (I - \rho W)y\]  \hspace{1cm} (26)

From the influence function, the weighting function can be defined as follows:

\[b_i = b(u_i) = \frac{\psi(u_i)}{u_i}\]

where \(u_i\) is the i residual which is standardized against the estimated standard deviation, it is obtained \(u_i = \frac{e_i}{s}\).
To obtain the $u_i$ value, first calculate the robust $s$ estimation scale. The value of $s$ can be obtained by means of:

$$
s = \frac{\text{MAD}}{0.6745} = \frac{\text{median}(|\varepsilon - \text{median}(\varepsilon)|)}{0.6745}
$$

Or can write $u = \frac{(I - \rho W)y - Z\hat{\delta}_{OLS}}{\text{MAD}}$, so the weighting function can be written as:

$$
\psi = \left(\frac{(I - \rho W)y - Z\hat{\delta}_{OLS}}{\text{MAD}} \right)
$$

$$
b_i = b(u) = \left\{ \begin{array}{ll}
1 - \left(\frac{u_i}{c}\right)^2, & \text{untuk } |u_i| \leq c \\
0, & \text{untuk } |u_i| > c
\end{array} \right.
$$

With $c = 4.685$ (Fox, 2002).

Because $\psi$ function isn’t linier, then the parameter estimation is solved by the iterative weighting least squares estimation method known as Iteratively Reweighted Least Square (IRLS) (Fox, 2002).

For parameters where $m$ is the number of iterations, the initial estimate $\hat{\delta}^{(0)}$ is as follows:

$$
\hat{\delta}^{(0)} = (Z^T B Z)^{-1} Z^T (I - \rho W)y
$$

With $B^{(0)}$ is an initial weighting matrix of size $n \times n$ which contains the initial weighting $b_1^{(0)}, b_2^{(0)}, \ldots, b_n^{(0)}$ and estimator $\hat{\delta}^{(0)}$, then the next estimator can be obtained by:

$$
\hat{\delta}^{(1)} = (Z^T B^{(0)} Z)^{-1} Z^T B^{(0)} (I - \rho W)y
$$

Then the weighting is recalculated from $b_i^{(1)}$, but $\hat{\delta}^{(1)}$ as a substitute $\hat{\delta}^{(0)}$, so it is obtained:

$$
b_i^{(1)} = \psi \left( \frac{(I - \rho W)y - Z\hat{\delta}^{(1)}}{s} \right) \left( \frac{(I - \rho W)y - Z\hat{\delta}^{(0)}}{s} \right)
$$

so obtained:

$$
\hat{\delta}^{(2)} = (Z^T B^{(1)} Z)^{-1} Z^T B^{(1)} (I - \rho W)y
$$
So on so that it is obtained:

\[
(1 - \rho W)y - Z\hat{\delta}^{(m-1)}
\]

\[
\frac{1}{s}
\]

then obtained: \( \hat{\delta}^{(m)} = \left( Z^T B^{(m-1)} Z \right)^{-1} Z^T B^{(m-1)} (1 - \rho W)y \)

Then for \( b_i^{(m)} \) weighted given, the estimator is obtained as follows:

\[
\hat{\delta}^{(m+1)} = \left( Z^T B^{(m)} Z \right)^{-1} Z^T B^{(m)} (1 - \rho W)y
\]

The calculation is carried out repeatedly until a convergent estimator is obtained, ie when the difference in value \( \hat{\delta}^{(m+1)} \) and \( \hat{\delta}^{(m)} \) close to 0 with \( m \) is the number of iterations.

Then the estimator \( \hat{\delta}^{(m+1)} \) will show unbiased, the estimator \( \hat{\delta}^{(m+1)} \) is said to be unbiased if \( E(\hat{\delta}^{(m+1)}) = \delta \)

\[
E\left( \hat{\delta}^{(m+1)} \right) = E\left[ \left( Z^T B^{(m)} Z \right)^{-1} Z^T B^{(m)} (1 - \rho W)y \right]
\]

\[
= E\left[ \left( Z^T B^{(m)} Z \right)^{-1} Z^T B^{(m)} \right] E\left[ (1 - \rho W)y \right]
\]

\[
= \left( Z^T B^{(m)} Z \right)^{-1} Z^T B^{(m)} \delta
\]

\[
= \delta
\]

From this description it is evident that \( \hat{\delta}^{(m+1)} \) it is an unbiased estimator.

4.2 Moran’s I Test

Obtained Z-hit value along with Moran’s I value as follows:

| Variable | Moran’s I | Z-hit | Conclusion |
|----------|-----------|-------|-------------|
| \( y \)  | 0.5752232 | 5.2978 | Rejected H_0 |
| \( x_1 \) | 0.19489697 | 1.9503 | Fail to Rejected H_0 |
| \( x_2 \) | -0.12105565 | -0.7779 | Fail to Rejected H_0 |
| \( x_3 \) | 0.06377452 | 0.8069 | Fail to Rejected H_0 |
| \( x_4 \) | 0.35389376 | 3.2687 | Rejected H_0 |
| \( x_5 \) | 0.17116800 | 1.7432 | Fail to Rejected H_0 |

From Table 1 we get the results that the variables \( y \) and \( x_1 \) have the value \( |Z_{hit\ unq}| > 1.96 \), which means that H_0 is rejected. This shows that there is spatial autocorrelation between locations in the response variable and the predictor variable. The response variable and at least one of the predictor variables indicate spatial autocorrelation between locations, the SDM model can be used in research.

4.3 Residual Plot

In Figure 1, visually there are some spatial outliers in residual data. To ascertain which data is included outlier, formal outline detection is done using formulas \( Z\left[ f(i) \right] x \sum_j \left( W_{ij} Z\left[ f(j) \right] \right) < 0 \).
From the analysis results, there were 19 spatial outliers detected, namely regions 1, 2, 9, 10, 11, 14, 16, 17, 18, 20, 23, 24, 27, 29, 30, 31, 32, 33 and 34.

4.4 Estimation of Spatial Durbin Model Robust Regression

The iteration process to achieve convergent $\delta$ values is performed using R software. Based on the results of robust spatial durbin M-estimator output models with Tukey Bisquare weighting, it can be seen that the iteration is performed 18 times to obtain convergent $\delta$. So the regression model obtained is as follows:

$$\hat{y}_i = 0.63839 \sum_{j=1}^{n} w_{ij} y_j + 28.76139 + 0.67309 x_{i1} + 0.03036 x_{i2} + 0.00045 x_{i3}$$

$$-0.11198 x_{i4} - 0.00025 x_{i5} - 1.70228 \sum_{j=1}^{n} w_{ij} x_{j1} + 0.05127 \sum_{j=1}^{n} w_{ij} x_{j2}$$

$$+ 0.00214 \sum_{j=1}^{n} w_{ij} x_{j3} - 0.09736 \sum_{j=1}^{n} w_{ij} x_{j4} + 0.00025 \sum_{j=1}^{n} w_{ij} x_{j5}$$

$$\hat{y}_i = 0.63839 \sum_{j=1}^{n} w_{ij} y_j + 23.70977 + 0.61799 x_{i1} + 0.02843 x_{i2} - 2.6744 \times 10^{-6} x_{i3}$$

$$- 0.19592 x_{i4} - 2.6278 \times 10^{-4} x_{i5} - 0.84814 \sum_{j=1}^{n} w_{ij} x_{j1} - 0.01467 \sum_{j=1}^{n} w_{ij} x_{j2}$$

$$+ 0.00138 \sum_{j=1}^{n} w_{ij} x_{j3} + 0.11522 \sum_{j=1}^{n} w_{ij} x_{j4} + 6.111 \times 10^{-4} \sum_{j=1}^{n} w_{ij} x_{j5}$$
4.4.1 Testing Hypothesis
Model Match Test
Hypothesis
\[ H_0 : \rho = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 0 \]
\[ H_1 : \rho \neq 0 \text{ or there is at least one } \beta_j \neq 0, \theta_j \neq 0, j=1, 2, 3, 4, 5 \]
Based on computational results, a value \( F_{hitung} = 40,55417 \) greater than \( F_{(0.05;5;35)} = 2.49 \). Then obtained \( p \)-value = 0.000. So the decision is obtained that \( H_0 \) is rejected at the 5% significance level.

So, it can be concluded that there is a relationship between AHH response variables with predictor variables together.

Significance SDM Robust Parameters Test
Hypothesis
\[ H_0 : \rho, \beta_j, \theta_j = 0, \text{ where } j=1,2,3,4,5 \]
\[ H_1 : \rho, \beta_j, \theta_j \neq 0, \text{ where } j=1,2,3,4,5 \]

Statistics Test:

| Parameter | Wald       | Keputusan          |
|-----------|------------|--------------------|
| \( \rho \) | 31,23823   | Rejected \( H_0 \) |
| Intercept | 15,97262   | Rejected \( H_0 \) |
| \( \beta_1 \) | 20,98586   | Rejected \( H_0 \) |
| \( \beta_2 \) | 7,89157    | Rejected \( H_0 \) |
| \( \beta_3 \) | 5,49284    | Rejected \( H_0 \) |
| \( \beta_4 \) | 13,14545   | Rejected \( H_0 \) |
| \( \beta_5 \) | 8,71631    | Rejected \( H_0 \) |
| \( \theta_1 \) | 13,49787   | Rejected \( H_0 \) |
| \( \theta_2 \) | 4,83751    | Rejected \( H_0 \) |
| \( \theta_3 \) | 13,41105   | Rejected \( H_0 \) |
| \( \theta_4 \) | 1,71848    | Fail Rejected \( H_0 \) |
| \( \theta_5 \) | 1,84298    | Fail Rejected \( H_0 \) |

Based on Table 2 it can be seen that the spatial lag parameters (\( \rho \)) and the parameters \( \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \theta_1, \theta_2 \) and \( \theta_3 \) have a significant effect on the model because the value of the wald obtained \( \chi^2_{0.05;1}=3,841 \), while the parameters \( \theta_4 \) and \( \theta_5 \) no significant effect on the model due to the value of the wald obtained \( \chi^2_{0.05;1}=3,841 \). So it can be concluded that the spatial lag parameters, the average length of school variables, the percentage of households behaving clean and healthy, the number of posyandu, the percentage of poor population, adjusted per capita expenditure, the spatial lag variable of the average predictor variable of school length, percentage households behave in a clean and healthy life and the number of posyandu has a significant effect, while for the spatial lag variable the percentage of poor population and per capita expenditure are adjusted insignificantly to AHH in Central Java Province.

4.4.2 Model Match Size With Adjusted \( R^2 \) and Mean Square Error (MSE)
The results of the analysis obtained values \( R^2_{Adj.k} = 1 - \frac{SS_G}{SS_T/(n-k-1)} = 0.9369 \), meaning that life expectancy is influenced by the average length of schooling, the percentage of households behaving clean and healthy, the number of posyandu, the percentage of poor population and per capita expenditure adjusted as much as 93.69% and the remaining 6.31% is influenced by other factors. MSE
value obtained = \frac{SS_{E}}{n-k-1} = 0.12551, \text{MSE values close to 0 indicate that the SDM Robust model formed is a good model.}

5. Conclusion

Based on the results and discussions that have been carried out, a number of conclusions can be drawn as follows:

1. The Durbin Spatial Model (SDM) Robust obtained is

\[ y_i = 0.63839 \sum_{j=1}^{n} w_{ij} y_j + 28.76139 + 0.67309 x_{i1} + 0.03036 x_{i2} + 0.00045 x_{i3} \]
\[ -0.11198 x_{i4} - 0.00025 x_{i5} - 1.70228 \sum_{j=1}^{n} w_{ij} x_{j1} + 0.05127 \sum_{j=1}^{n} w_{ij} x_{j2} \]
\[ +0.00214 \sum_{j=1}^{n} w_{ij} x_{j3} - 0.09736 \sum_{j=1}^{n} w_{ij} x_{j4} + 0.00025 \sum_{j=1}^{n} w_{ij} x_{j5} \]

The better model used to explain life expectancy in Central Java Province in 2017 is the spatial durbin model robust, because the value of Adjusted R^2 obtained is greater and the MSE value obtained is smaller.

2. Factors that have a significant influence on life expectancy in Central Java Province in 2017 are average length of schooling, percentage of households behaving clean and healthy, number of posyandu, percentage of poor population and adjusted per capita expenditure.

References

[1] Anselin L 1988 Spatial Econometrics: Methods and Models Dordrecht: Kluwer Academic Publishers
[2] Chen C 2002 Robust Regression and Outlier Detection with The ROBUSTREG Procedure 265-27 SAS Institute Inc., Lary, NC
[3] Wuryandari T, Hoyyi A, Kusumawardani D and Rahmawati D 2014 Identification of Spatial Autocorrelation in the Number of Unemployment in Central Java Using Moran’s Index Media Statistika Journal 7(1) 1-10 Diponegoro University
[4] LeSage J P 1999 The Theory and Practice of Spatial Econometrics Ohio: Department of Economics, University of Toledo
[5] Ramadani I R, Rahmawati R and Hoyyi A 2013 Analysis of Factors Affecting Malnutrition of Toddlers in Central Java Using the Spatial Durbin Model Method Gaussian Journal 2(4) 333-342. Diponegoro University, Semarang
[6] Draper N R and Smith H 1998 Applied Regression Analysis (3rd ed.) New York: John Wiley and Sons
[7] Montgomery D C and Peck E A 1992 Introduction To Linier Regression Analysis New York: John Wiley and Sons, Inc
[8] Fox J 2002 Robust Regression : Appendix to An R and S-Plus Companion to Applied Regression
[9] Badan Pusat Statistik (BPS) 2018 Central Java in Numbers 2018 Semarang: BPS, Central Java