Modeling Spatial Error Model (SEM) on Human Development Index (IPM) in Central Java 2018

Research Article

Aprilia Dwi Anggara Wati[1*], Laelatul Khikmah [iD][2]

[1,2]Department of Statistics, Akademi Ilmu Statistika Muhammadiyah Semarang, Semarang 50185, Indonesia

[*]apriliaanggara11@gmail.com (coresponden author)
aisyah.salsabila17@gmail.com

ABSTRACT

The Human Development Index (HDI) is a human development index that is used to achieve the development outcomes of a region. HDI is formed by 3 basic dimensions, namely the health dimension as seen from the indicator of life expectancy at birth, the dimension of knowledge seen from a combination of indicators of average length of schooling and expectation of school years and dimensions of decent living standards as seen from the indicator of average per capita expenditure has been adjusted. The development of HDI in Central Java shows an increase every year. In 2018 the HDI figure for Central Java Province reached 71.12% and increased by 0.6% from the previous year. This is because the large HDI figures in an area are influenced by the large HDI numbers in adjacent areas. The location / area factor is thought to have a spatial dependence effect on the HDI figure. This problem can be overcome by using spatial regression by including the relationship between regions into the model. The spatial regression approach used in this study is the Spatial Error Model (SEM). The weighting matrix used in this study is Queen Contiguity (intersection between sides and corners). This study provides results that the variables that significantly influence HDI are poverty and school enrollment rates.

Keywords : Human development index, queen contiguity, spatial error model (SEM), spatial regression.

1. INTRODUCTION

Human Development Index (HDI) or HDI is an index of human development that is used to achieve the development results of an area (Nisa and Karim 2017). The development of HDI figures provides an indication of an increase or decrease in human development performance in an area (Sari and Priati 2016). Nationally, Central Java Province is one of the regions with an HDI in the middle category, ranking 13 out of 34 provinces in Indonesia. The development of HDI in Central Java shows an increase every year. In 2016 the HDI figure was 68.98%. in 2017 it was 70.52%, an increase of 1.54% and in 2018 it was 71.12%, an increase of 0.6%. The results of BPS publications show that the HDI in Central Java has developed. which reflects the improving human development in Central Java.

One of the statistical methods that can be used to determine the variables that affect the human development index is regression analysis. Regression analysis is a method that is often used to see whether the dependent variables affect an independent variable. The relationship between the dependent and independent variables is often influenced by the location of an area or the presence of spatial effects (Wardani, Handajani, and Zukhronah 2019). According to (Lispani, Sumarjaya, and Sukarsa 2018) spatial regression is one of the statistical methods used to determine the relationship between the dependent variable and the independent variable by considering the spatial influence between regions.

Previous research has been conducted by (Novitasari and Khikmah 2019b) about "Application of the Spatial Regression Model to the Human Development Index (HDI) in Central Java in 2016" using the Spatial Durbin Model (SDM) regression method. The weighting matrix used in this study is Queen Contiguity (the intersection between the angles). Variables per capita expenditure, number of health facilities, poverty, and school-student ratios (SD, SMP and SMA) have an effect on HDI figures in Central Java Province in 2016. The strength of this research is that the HR method provides better test results and information accuracy. spatial influence. In 2018 (Novitasari and Khikmah 2019a) conducted research on "Application of the Spatial Regression Model to the Human Development Index (HDI) in Central Java in 2017" which compares the Spatial Autoregressive Model (SAR) and the Spatial Error Model (SEM) and the SEM model is better than the SAR model with an AIC value of - 1405.6, the weighting matrix used in this study is Queen Contiguity (the intersection of the sides and corners). This study provides results that the SEM model is better than the SAR model. While the variables of the number of health facilities, poverty, per capita expenditure, and inflation have an effect on the HDI figure in Central Java Province in 2017. The strength of this study is that it uses 2 models in spatial regression by comparing the two,

which aims to find the best model.

The object used in this research is the HDI in Central Java Province. Because the units of observation in this study are in the form of regions or locations, namely districts and cities in Central Java Province. This study covers 29 districts and 6 cities in Central Java Province with a time period of 2018 to see the effect of independent variables on the dependent variable by considering spatial or regional factors. In spatial data, observations at one location often depend on observations at other nearby locations (Safitri, Darsyah, and Utami 2014). In overcoming this problem, a spatial error dependency effect approach is used, namely using the Spatial Error Model (SEM). The purpose of this study is to find out what variables affect HDI in Central Java Province in 2018 using the Spatial Error Model (SEM).

2. THE MATERIAL AND METHOD

2.1. Spatial Regression

Spatial regression is the result of the development of the classical linear regression method. The development is based on the influence of place or spatial on the analyzed data (Astuti, Yasin, and Sugito 2013). The general model of spatial regression is stated as follows :

$$y = \rho W_1 y + X\beta + u \tag{1}$$
$$u = W_2 u + \varepsilon$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

where:

y : The response variable vector is n x 1
X : The predictor variable matrix of size n x (k + 1)
β : The parameter vector of the regression coefficient is (k + 1) x 1
ρ : The spatial lag coefficient parameter is the predictor variable
u : The error vector is nx1
ε : The error vector is nx1
W_1, W_2 : Weight matrix of size nxn
I : Identity matrix of size nxn

3.2. Spatial Error Model (SEM)

Spatial Error Model is a spatial error model where the error has spatial correlation (Laswinia and Chamid 2016). The advantage of the SEM model is that it provides a better model for interconnected observations. The spatial error model is formed when $W_1 = 0$ and $\rho = 0$. So this model assumes that the autoregressive process is only the error model. The general SEM model is shown by the following equation:

$$y = X\beta + u$$
$$u = \lambda W_u + \varepsilon$$
$$Y = X\beta + \lambda W_u + \varepsilon \tag{2}$$

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

Modeling Spatial Error Model (SEM) on Human Development Index (IPM) in Central Java 2018
pISSN: 2715-6923, eISSN: 2721-9186, Vol. 1, No. 2, September 2020, *pp. 50-55*

52

where Y is dependent variable; λ is error spatial coefficient parameter; X is independent variable; β is the parameter of the regression coefficient; ε is the error; and W_u is spatial weighting matrix error.

The spatial error model parameter (λ) indicates the spatial effect of the error from a ith region on the jth region around it, if there is no spatial dependence on the error between the ith region and the jth region, the spatial error model parameter (λ) will be is zero.

3. RESULT AND DISCUSSION

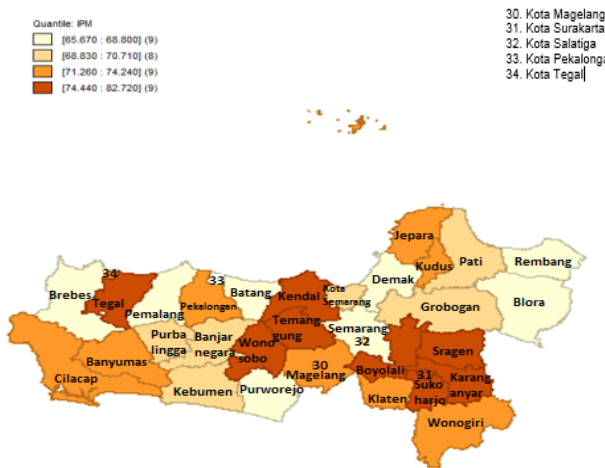### 3.1. IPM Deployment Thematic Map

Fig. 1. IPM deployment thematic map.

Based on Figure 1 above, it can be seen that the Thematic Map of the Distribution of HDI is grouped into four categories, namely very high category, high category, medium category and low category based on HDI numbers. The areas with the highest HDI numbers are marked with the darkest color. There are 9 areas that are in the very high category, namely Tegal Regency, Kendal Regency, Wonosobo Regency, Temanggung Regency, Boyolali Regency, Surakarta City, Sragen Regency, Sukoharjo Regency and Karanganyar Regency. Meanwhile, the regions in the lowest category are marked with the lightest color. There are 9 regions, namely Brebes Regency, Pemalang Regency, Pekalongan City, Batang Regency, Purworejo Regency, Semarang Regency, Demak Regency, Blora Regency and Rembang Regency. The following is a description of the Regency / City with each category based on the HDI figures.

Table 1.List of districts with the highest IPM based on each category according to thematic maps.

IPM Percentage Number (%)	Regency / City
74.440 : 82.720	Tegal Regency, Kendal Regency, Wonosobo Regency, Temanggung Regency, Boyolali Regency, Surakarta City, Sragen Regency, Sukoharjo Regency and Karanganyar Regency
71.260 : 74.240	Cilacap Regency, Banyumas Regency, Pekalongan Regency, Magelang Regency, Klaten Regency, Kudus

IPM Percentage Number (%)	Regency / City
	Regency, Tegal City, Jepara Regency and Magelang City
68.830 : 70.710	Purbalingga Regency, Banjarnegara Regency, Kebumen Regency, Semarang City, Grobogan Regency and Pati Regency, Salatiga Regency and City
65.670 : 68.800	Brebes Regency, Pemalang Regency, Pekalongan City, Demak Regency, Blora Regency and Rembang Regency

Table 1, shows the list of districts / cities based on each category according to the thematic map in Figure 4.2 and it can be seen that districts / cities that have HDI are in the very high category, namely 74,440: 82,720 in nine districts / cities. Districts / Cities that have HDI in the high category are 71,260: 74,240 in eight districts / cities. Districts / Cities that have HDI in the medium category, namely 68,830: 70,710 in nine districts / cities and districts / cities that have an HDI in the low category, namely 65,670: 68,800 in nine districts / cities.

3.2. Spatial Weighted Matrix

In this study, using a spatial weighted matrix queen contiguity. This queen contiguty weighting matrix has the concept of the intersection of sides and angles, where the adjacent areas and vertices meet other regions, then the value is $W_{ij} = 1$. Meanwhile, other regions are given a value of $W_{ij} = 0$. Making a queen contiguity weighting matrix is done by compiling a matrix measuring 35 x 35 which is based on the number of districts / cities in Central Java Province, totaling 35 districts / cities. The following is the histogram to find out the number of neighbors in each district / city in Central Java Province.

from	1	2	3	4	5	6	7	8
to	2	3	4	5	6	7	8	9
#obs	2	5	7	6	8	1	5	1
% of total	5.714	14.286	20.000	17.143	22.857	2.857	14.286	2.857
sd from mean	-1.171	-0.632	-0.092	0.000	0.447	0.986	1.525	2.064

min: 1, max: 8, median: 4, mean: 4,17143, s.d.: 1,85481, #obs: 35
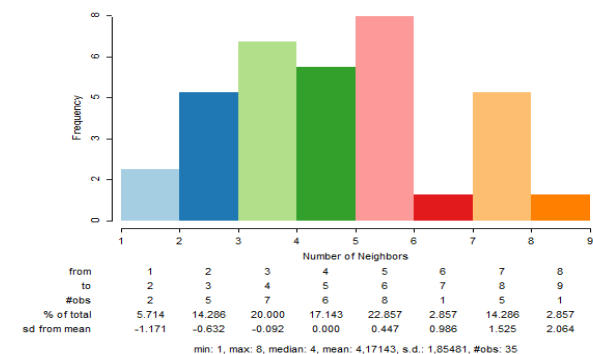
Fig. 2. Number of neighborhood histograms.

Based on the spatial weighted matrix as in Figure 2, the histogram above for the horizontal section shows how many neighbors each district / city has. Meanwhile, the vertical section shows the number of districts / cities. From all regions / locations, they were divided into 8 groups. Those with the highest number of neighbors were 8 regions, while those with the least number of neighbors were 1 region.

3.3. Spatial Autocorrelation Test

To find out whether or not there is spatial autocorrelation between regions, you can use the Moran's Index Test (Morran's test). If there is spatial autocorrelation, it is carried out using a spatial regression model.

Hypothesis:

$H_0: I = 0$ (no spatial autocorrelation between locations)
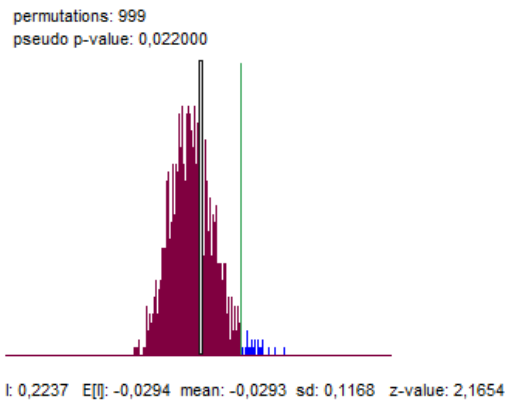$H_1: I \neq 0$ (there is spatial autocorrelation between locations)

Fig. 3. Moran's index test.

From Figure 3, the morans index test is obtained on the Human Development Index data in Central Java Province in 2018 as in Table 2 below.

Table 2. Moran's index test.

Test	P-value	Z-Value	Morans'i	E(I)
IPM	0.022000	2.1654	0.2237	-0.0294

Based on Table 2, it can be seen that the P-value = 0.022000 is smaller than α = 10%, so it can be concluded that rejecting H0 means that there is spatial autocorrelation in the Human Development Index in Central Java in 2018. The moran index statistical value obtained is 0.2237 while the expected value of the daily index is -0.029. means $I > I_0$ (data with positive autocorrelation or grouped data). It can be said that in general the Human Development Index in Central Java Province in 2018 experienced a spatial grouping process. The point is that the high HDI in districts / cities of Central Java Province in 2018 tends to be surrounded by high district / city HDI as well. and vice versa.

To see the pattern of HDI grouping in Central Java Province in 2018. Then you can use a scatterplot. Through this scatterplot, it can classify districts / cities into four different types of quadrants. Quadrant I and quadrant III explain the existence of positive spatial autocorrelation or grouping an area that has an HDI that is directly proportional. Meanwhile, in quadrant II and quadrant IV, it explains the existence of negative spatial autocorrelation or classifies an area that has an HDI that is inversely proportional.

Based on Figure 4, it can be seen that most of the areas in Central Java Province tend to spread out in quadrant I and quadrant III. This identifies that the regions in Central Java Province tend to experience grouping or positive spatial autocorrelation, or in quadrant I and quadrant III there is a grouping of regions with an HDI which is

directly proportional. It can be said that an area that has the highest HDI will tend to be surrounded by other areas that have the highest HDI. Likewise, the areas with the lowest HDI will tend to be surrounded by areas with the lowest HDI.
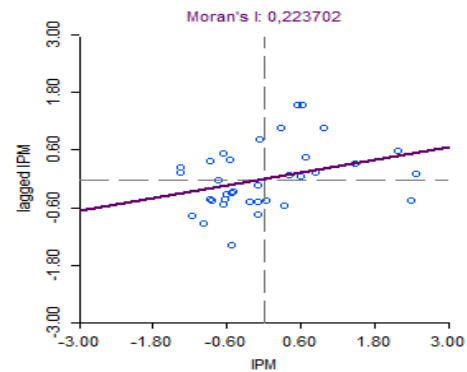

Fig. 4. Morans' scatterplot.

3.4. Lagrange Multiplier test

Based on the results of the previous Moran index test, it was known that there was spatial autocorrelation in the data. So it is necessary to do the next research, namely the Lagrange Multiplier test. Below are the results of the Lagrange Multiplier test using R studio 3.6.0.

Table 3. Lagrange multiplier test results.

Test	Statistic Parameter	Probability
Lagrange Multiplier (lag)	3.59189	0.05806

Based on table 3 it can be seen that the error is less than α of 0.05868. so an error dependency is detected. Therefore. Analysis can be done using the Spatial Error Model (SEM) model. In this study, the spatial error dependency effect approach is used, namely by using the Spatial Error Model (SEM).

3.5. Spatial Error Model (SEM)

From the results of the previous lagrange multiplier test, it was known that the analysis was carried out using the spatial error model (SEM). so the next step is to test with the SEM model. Below are the results of parameter estimation using the SEM model.

Table 4. SEM model parameter estimation.

Variable	Estimation	Std. Error	Z-Value	P-Value
IPM (ρ)	99.332	13.252	7.496	6.595e-14
Poverty	-0.052	0.009	-5.721	1.059e-14
Teacher Pupil Ratio	-0.024	0.152	-0.160	0.873
School Participation Rate	-0.196	0.078	-2.530	0.011
Inflation	-2.064	1.528	-1.351	0.177
Open unemployment	0.241	0.350	0.688	0.491

Modeling Spatial Error Model (SEM) on Human Development Index (IPM) in Central Java 2018
pISSN: 2715-6923, eISSN: 2721-9186, Vol. 1, No. 2, September 2020, *pp. 50-55*

54

Variable	Estimation	Std. Error	Z-Value	P-Value
λ	0.678			0.008

From Table 4, the output of the Spatial Error Model (SEM) is obtained so that it can be seen that the independent variables that have an effect on the dependent variable that has a p-value less than alpha 10% are poverty (X1) and school participation rate (X3). Meanwhile, the teacher student ratio variable (X2). inflation (X4) and open unemployment (X5) do not affect the dependent variable.

So that the SEM model equation obtained is based on significant variables, namely:

$$\hat{Y} = 99.332 + 0.678W_u - 0.052X_1 - 0.196X_3$$

From the SEM model obtained, the two variables that have a negative effect on the human development index are poverty and school enrollment rates. So that increasing poverty and school enrollment rates will reduce the human development index in Central Java in 2018.

The coefficient results can be seen that the value of lambda (λ) is 0.67784, it will increase the percentage of human development index in a district / city by 0.67784. Meanwhile, the normality assumption test for the residuals uses the Kolmogorov-Smirnov test with a p-value 0.11649> α = 10%, which means that the residuals are normally distributed. The identical assumption test uses the Breusch-Pagan test (BP test) with a p-value of 0.1612 which is significant for α = 10%, which means that it shows identical residuals. To test the independent assumption using the residual Morans'I which shows that the p-value = 0.2971> α = 10% means that the model fulfills the independent assumption.

4. CONCLUSION

There are 9 regions with the distribution of HDI numbers that fall into the very high category, namely Tegal Regency, Kendal Regency, Wonosobo Regency, Temanggung Regency, Boyolali Regency, Surakarta City, Sragen Regency, Sukoharjo Regency and Karanganyar Regency. While the areas that fall into the lowest category are 9 regions, namely Brebes Regency, Pemalang Regency, Pekalongan City, Batang Regency, Purworejo Regency, Semarang Regency, Demak Regency, Blora Regency and Rembang Regency. From the HDI modeling results and the variables that influence it, it can be concluded that the poverty variable (X1) and the school participation rate (X3) have a negative effect on the human development index figure in Central Java Province in 2018. Meanwhile, the student teacher ratio variable (X2), inflation (X4) and open unemployment (X5) do not have a significant effect on the human development index figure in Central Java province in 2018. Spatial Autoregressive (SAR) and Spatial Autoregressive Moving Average (SARMA) approaches are needed to model Central Java's HDI.

REFERENCES

Astuti, Restu, Hasbi Yasin, and Sugito Sugito. 2013. "Aplikasi Model Regresi Spasial Untuk Pemodelan Angka Partisipasi Murni Jenjang Pendidikan Sma Sederajat Di Provinsi Jawa Tengah." *Jurnal Gaussian* 2 (4): 375–84.

Laswinia, Vonesa Devi, and Mutiah Salamah Chamid. 2016. "Analisis Pola Hubungan Persentase Penduduk Dan Sosial Di Indonesia Menggunakan Regresi Spasial." *Jurnal Sains Dan Seni ITS* 5 (2): 1–6. https://doi.org/10.5152/jaem.2014.16769.

Lispani, Ni Made Lasti, I Wayan Sumarjaya, and I Komang GDE Sukarsa. 2018. "Pemodelan Jumlah Tindak Kriminalitas Di Provinsi Jawa Timur Dengan Analisis Regresi Spatial Autoregressive and Moving Average." *E-Jurnal Matematika* 7 (4): 346. https://doi.org/10.24843/mtk.2018.v07.i04.p224.

Nisa, Imaroh Izzatun, and Abdul Karim. 2017. "Pemodelan Spatial Durbin Error Model (SDEM) Pada Data Indeks Pembangunan Manusia (IPM) Di Jawa Tengah." *Jurnal Statistika Universitas Muhammadiyah Semarang* 5 (1): 33–40.

Novitasari, Dina, and Laelatul Khikmah. 2019a. "Penerapan Model Regresi Spasial Pada Indeks Pembangunan Manusia (IPM) Di Jawa Tengah Tahun 2017." *STATISTIKA Journal of Theoretical Statistics and Its Applications* 19 (2): 123–34. https://doi.org/10.29313/jstat.v19i2.5068.

———. 2019b. "Penerapan Model Regresi Spasial Pada Indeks Pembangunan Manusia (IPM) Di Jawa Tengah Tahun 2017." *STATISTIKA Journal of Theoretical Statistics and Its Applications* 19 (2): 123–34. https://doi.org/10.29313/jstat.v19i2.5068.

Safitri, Diana Wahyu, Moh Yamin Darsyah, and Tiani Wahyu Utami. 2014. "Pemodelan Spatial Error Model (SEM) Untuk Indeks Pembangunan Manusia (IPM) Di Provinsi Jawa Tengah." *Jurnal Statistika* 2 (2): 9–14.

Sari, Betha Nurina, and Priati Priati. 2016. "Identifikasi Keterkaitan Variabel Dan Prediksi Indeks Pembangunan Manusia (IPM) Provinsi Jawa Barat Menggunakan Dynamic Bayesian Networks." *JURNAL INFOTEL - Informatika Telekomunikasi Elektronika* 8 (2): 150. https://doi.org/10.20895/infotel.v8i2.123.

Wardani, Iis Kusuma, Sri Sulistijowati Handajani, and Etik Zukhronah. 2019. "Pemodelan Tingkat Pengangguran Terbuka Provinsi Jawa Tengah Menggunakan Model Spatial Autoregressive Moving Average." *Prosiding Sendika* 5 (2): 199–205.

Aprilia Dwi Anggara Wati (Time New Roman 9 pt, Bold) received the Diploma 3 statistict (A.Md. Stat.) degree in Statistic from Akademi Statistika (AIS) Muhammadiyah Semarang in 2020. His research interests include spatial regression.

Laelatul Khikmah received the Master of Science (M.Si) degree in Applied of Statistics from IPB University in 2017. Her research interests include linear regression and categorical data analysis.