ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 226

1st International Conference on Social Sciences (ICSS 2018)

East Java Human Development Index Modeling with Spatial Regression Approach

1st Wara Pramesti
Department of Statistics
University of PGRI Adi Buana
Surabaya, Indonesia
warapra@unipasby.ac.id

2nd Artanti Indrasetianingsih
Department of Statistics
University of PGRI Adi Buana
Surabaya, Indonesia
artanti.indra@unipasby.ac.id

***Abstract*—The success of the development of the quality of life in a country or region can be measured through an indicator called Human Development Index (HDI). HDI is formed by three basic dimensions, namely a long and healthy life, knowledge, and decent standard of living. Since 2010 East Java Province has an HDI value that continues increasing from year to year. East Java Province's HDI in 2016 reached 0.6974 or increased by 6.71% from the 2015 HDI. The purpose of this study is modeling the factors that are expected influencing HDI districts/cities of East Java Province 2016, namely per capita expenditure, morbidity, percentage of poor people and mean years of schooling. The analysis was carried out using a spatial regression approach. The results of the analysis obtained that the variables that significantly influence HDI discticts/cities of East Java Province in 2016 are per capita expenditure, the percentage of poor people and mean years of schooling. Modeling with spatial error regression results better when compared with spatial lag regression, because the spatial error regression model has a lower AIC(Akaike Information Criterion), which is 114.63, while the spatial lag regression is 119.98**

Keywords—AIC; HDI; spatial regression

I. INTRODUCTION

The success of the development of the quality of human life in a country or region can be measured through an indicator called the Human Development Index (HDI) [1]. The HDI explains how residents can access development results in obtaining income, health, education and etc. The HDI was introduced by the United Nations Development Programme (UNDP) in 1990 and published regularly in the annual Human Development Report (HDR) [1].The HDI is formed by 3 (three) basic dimensions, namely a long and healthy life, knowledge, and decent standard of living [1]. The HDI can determine the level or level of development of a region / country.

Based on Central Bureau of Statistics/BPS data , The HDI value of East Java Province in 2015 was 0.6895, an increase of 0.81 from 2014. East Java's HDI value in 2016 increased by 0.79 from 2015. East Java's HDI value is still in the medium category.

Tobler's Law says that locations that are close together have a higher relationship than distant locations/ regions. Based on Tobler's law, it can be interpreted that the HDI in districts/cities in East Java have close locations have a higher relationship. Therefore if we want to know the factors that influence the HDI, we need to pay attention to the location or region. There is a statistical method that can be used to model the factors that influence HDI, i.e. classical regression modeling. But classical regression modeling is not able to consider location aspects. Therefore, the development of classical regression analysis is used in this study, namely spatial regression analysis. There is a weighting matrix which is an indication of the relationship between an area and other regions. . In this spatial regression approach there is a weighting matrix which is an indication of the relationship between an area and other regions. Spatial modeling is done by an autoregressive process, which is indicated by a dependency relationship between a set of observations or locations [1].

Spatial effects can be divided into two parts, namely spatial autocorrelation and spatial heterogeneity. If the residual from the classical regression obtained is autocorrelation, this indicates a spatial dependency. This means that there are dependencies between locations / regions. Locations that are close together have a higher relationship than far locations / regions (Tobler's law). If the classical regression residual is declared not homogeneous or heteroscedasticity occurs, then there is a suspicion of heterogeneity between locations / regions. For further assure these allegations, Morans I was tested, which is a test to determine the existence of spatial dependencies. And the Breusch-Pagan Test is used to determine the heteroscedasticity or heterogeneity between regions. The next test is the Lagrange Multiplier (LM), used to detect the appropriate spatial regression models (Spatial Autoregressive Models (SAR), Spatial Error Models (SEM), or Spatial Autoregressive

Moving Average (SARMA)). the best model is chosen based on the smallest AIC value.

The following are some of the previous studies on the human development index. The ratio of students to teachers, participation rates of SMP/MTs, number of health facilities, percentage of RT with access to clean water, population density, labor force participation rate and GDP per capita had a positive and significant effect on the HDI of the districts/cities in East Java [2]. Life Expectancy, Literacy Rate and PPP had a positive and significant effect on the HDI of the districts/cities in Central Java [3]. The HDI from 15 countries that have HDI values in the category of very high, high and average level values have strong correlation values (above 0.7) with Gross National Index per capita (GNI), education index, innovation index, Information and Communication Technologies (ICT) index, life expectancy index. Regression modeling results between the HDI and the five factors indicate that the HDI was influenced by education index, index innovation, ICT and life expectancy index [4]. The HDI of North Sumatra was affected by unemployment, per capita income, education and health, the method used is Error Correction Model (ECM). The ECM results also showed that in the short and long-term variables used in this study the human development indeks [5]. A Bayesian factor analysis model was used as an alternative to the human development index. This study estimate human development with three auxiliary variables capturing environmental health and sustainability, income inequality, and satellite observed nightlight. The result found that the living standard dimension provided a greater contribution to human development than the official HDI suggests, while the longevity dimention provides a lower proportional contribution [6]

The purpose of this study is modeling the factors that are expected influencing the HDI districts/cities of East Java Province 2016, namely per capita expenditure, mortality rate, percentage of poor people and mean years of schooling. The analysis was carried out using a spatial regression approach.

II. RESEARCH METHODS

The data that published by BPS East Java used in this paper. The Human Development Index, Per Capita Expenditure, Morbidity, Percentage of Poor People and Mean Years of Schooling by district/city in East Java 2016, which consists of 29 districts and 9 cities.

The response variable in this study is the Human Development Index at the district / city level in East Java 2016. The predictor variables are per capita expenditures (x_1) (Rp 000), Morbidity (x_2), percentage of poor people (x_3), and mean years of schooling (x_4).

The research stage begins with classical regression modeling and residual testing. If the assumption of an independent residual is not fulfilled, it is said that there is an indication of spatial dependency, then proceed with the Moran's I test. But before conducting the Moran's I test, the weighting is given first. If the Moran's I test is significant, the Langrange Multiplier test is used to determine the appropriate spatial model, namely SAR, SEM or SARMA. The next step is the appropriate spatial modeling. Selection of the best model using AIC and the last interpretation of the model.

. III. RESULT AND DISCUSSION

A. *Result*

Descriptive statistics in Table I show that the average HDI of district/city in East Java 2016 is 69.792%. The lowest HDI value is 59.09% and the highest HDI is 80.46%. The average value per capita expenditure is 10,561 (Rp. 000). The average value of morbidity is 16.117. The average value of percentage of poor people is 12.138. The average value of mean years of schooling is 12.98 years.

Fig. 1. shows that the highest HDI value is the HDI of Malang City which is 0.8046, while the lowest HDI value is the HDI of Sampang district, which is 0.5909.

TABLE I. DESKRIPTIVE STATISTICS

Variable	Mean	Std.Dev.	Min	Max
HDI	69.792	5.377	59.090	80.460
x_1	10,561	2,123	7,846	16,295
x_2	16.117	3.237	9.843	25.093
x_3	12.138	5.262	4.330	25.000
x_4	12.981	0.904	11.370	15.380

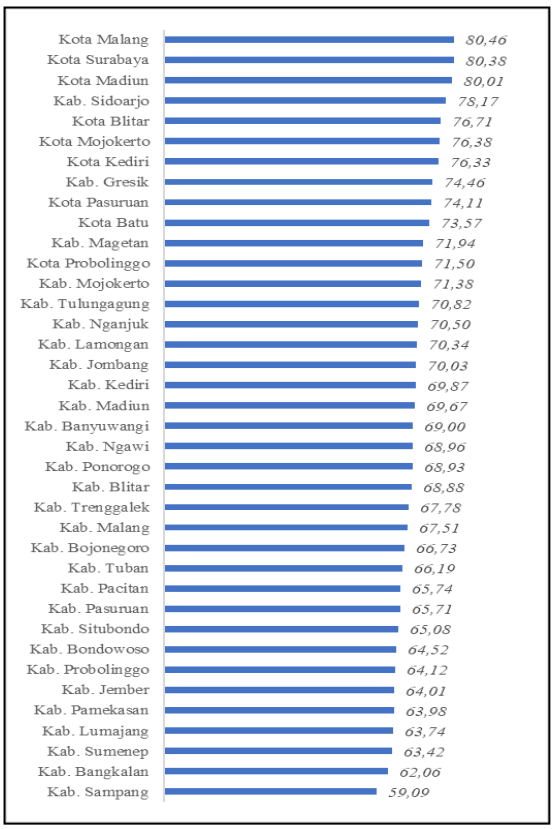

Fig. 1. The HDI districts/cities of East Jawa 2016

The classic regression analysis is used to determine the pattern of the relationship between the response variable and predictor variables. The results of the analysis can be seen in Table II.

TABLE II. PARAMETER ESTIMATION OF MULTIPLE LINEAR REGRESSION

Variable	Coef.	Standard error	Prob.
Constanta	38.1113	4.4719	0.00000
x_1	0.0013	0.0002	0.00000
x_2	-0.0660	0.0658	0.32290
x_3	-0.2929	0.0563	0.00001
x_4	1.7609	0.3667	0.00003

Based on Table II there is one insignificant predictor variable, morbidity (x_2) which has a probability value (0.3229) greater than the specified alpha value (0.05). Per capita expenditure variable (x_1), percentage of poor people (x_3) and mean years of schooling (x_4), each of which has a probability value less than 0.05, it can be said that these variables have a significant effect on the HDI . Per capita expenditure and mean years of schooling have a positive effect on the HDI, and the percentage of the poor people has a negative effect on the HDI. The determination coefficient value is 0.949 states that variations in the Human Development Index can be explained by predictor variables (per capita expenditure, morbidity, percentage of poor people and mean years of schooling) of 94.9%, the remaining 5.1% are explained by other variables not included in the model . The estimated regression model obtained:

$$\hat{y} = 38{,}1113 + 0{,}0012728x_1 - 0{,}0660293x_2 - 0{,}292922x_3 + 1{,}76086x_4$$

1. Residual assumption test and the spatial dependency test

The residual assumption test in regression includes homoskedasticity test, autocorrelation and normality testing. The test results show that only the assumption of autocorrelation is not met. Breusch-Pagan is used to test the assumption of variance homogeneity assumption, and obtained Breusch-Pagan value is 4.3055 with a probability value 0.36624 greater than 0.05, which means that the assumption of homogeneous residual variance is fulfilled. Autocorrelation test using Durbin Watson and DW value obtained at 1.17283 lies in the interval 0 <1.17283 <1.2614, which means that the assumption of residual autocorrelation is not met. Normality testing is used Kolmogorov Smirnov Test with a probability value 0.150 greater than 0.05, then the assumption of a normal distribution of residuals is met. The autocorrelation assumption is not met, it can proceed to the area spatial regression approach, by testing spatial dependencies to determine whether or not there is a spatial or location influence in the model.

TABLE III. LAGRANGE MULTIPLIER TEST

LM Test	LM	df	Prob.
LM lag (SAR)	13.8909	1	0.0002
LM error (SEM)	10.4146	1	0.0013

The spatial dependency test with Moran's I produces Moran's I value of 0.4413 with a probability value of 0.00017 which means significant. The Moran's I value obtained is greater than Io = -0.0270, then the HDI variation has a cluster pattern or it can be said that there is a positive autocorrelation. So it can be concluded that there is location autocorrelation or dependence between locations / regions, so it can be done modeling with an area spatial regression approach.

The Lagrange Multiplier (LM) Test was tested as the basis for selecting the appropriate spatial regression model. The Lagrange Multiplier test results are shown in Table III. Based on Table III, it is known that the value of LM lag (SAR) is 13.8909 with a probability value of less than α = 5% which means reject H_0. LM error is 10.4146 with a probability value of 0.00125 less than α = 5% which means reject H_0. So it can be concluded that there is a spatial lag autocorrelation in the response variable, and in the error. So that it can be continued modeling with Spatial Autoregressive Model (SAR), and Spatial Error Model (SEM).

2. Spatial Autoregressive Model (SAR)

The result of LM test show that there is a spatial lag autocorrelation on the response variable, so the Spatial Autoregressive Model (SAR) modeling is performed which results can be seen in Table IV.

Based on Table IV. obtained coefficient value ρ=0.235 with a probability value of 0.0002 less than α=0.05, so reject H_0. It means that ρ is significant. Because ρ is significant, it can be said that there is a dependency effect on an area with other adjacent areas.

From the results of the analysis on Table IV it is known that there is only one independent variable that is not significant, morbidity (x_2), which has probability value 0.91001 more than α=0.05. Variables per capita expenditure variables (x_1), the percentage of poor people (x_3), and mean years of schooling (x_4), each has a probability value 0.00, less than 0.05, so the variables have significant effect on the Human Development Index in the districts/cities of East Java 2016. The value of determination coefficient is 0.9649 indicating that the variation in the Human Development Index can be explained by per capita expenditure (x_1) , morbidity (x_2), percentage of poor people (x_3), and mean years of schooling (x_4), amounting to 96.49% and the remaining 3.51% is explained by other variables not included in the model. The Spatial Autoregressive Model (SAR) model equation obtained is:

$$\hat{y}_i = 20{,}6242 + 0{,}2348 \sum_{j=1;i\neq j}^{n} W_{ij}Y_j + 0{,}0011x_{1i} - 0{,}0061x_{2i} - 0{,}2525x_{3i} + 1{,}9197x_{4i}$$

3. Spatial Error Model (SEM)

Furthermore, an analysis of spatial regression approach was carried out using Spatial Error Model (SEM) modeling. Test results can be seen in Table V.

TABLE IV. PARAMETER ESTIMATION OF SAR

Variable	Coef.	Standard error	Prob.
ρ	0.235	0.0630	0.0002
constanta	20.624	5.8280	0.0004
x_1	0.001	0.0001	0.0000
x_2	- 0.006	0.0540	0.9100
x_3	- 0.252	0.0440	0.0000
x_4	1.919	0.2870	0.0000

Based on Table V, it is known that there is a spatial dependency on the error shown by λ=0.679 and has a probability value 0.000 less than α=0.05 . This means that there is a relationship between the Human Development Index in an area and other adjacent regions. Predictor variables that have no significant effect are only one i.e. morbidity (x_2), indicated by a greater probability value (0.128) than α=0.05 . Other predictor variables have a significant effect, because it has a probability value of less than α=0.05. It means that the Human Development Index in districts/cities in East Java in 2016 is influenced by per capita income (x_1), percentage of poor people (x_3), and mean years of schooling (x_4). The determination coefficient is 0.9724, indicating that variations in the Human Development Index in East Java can be explained by per capita income (x_1), morbidity (x_2), percentage of poor people (x_3), and mean years of schooling (x_4) i.e. 97.24% and the remaining 2.76% is influenced by other variables not included in the model. The Spatial Error Model (SEM) model is:

$$\hat{y}_i = 37{,}067 + 0{,}0012x_{1i} - 0{,}064x_{2i} - 0{,}208x_{3i} + 1{,}799x_{4i} + u_i$$

$$u_i = 0{,}679 \sum_{j=1, i \neq j}^{n} w_{ij} u_j + \varepsilon_i$$

From both models, Spatial Autoregressive Model and Spatial Error Model, the best model can be selected by comparing the value of Akaike's Information Criterion (AIC). The best model is the model with the smallest AIC value. The AIC value of the Spatial Autoregresive model is 119,981 and the AIC value of the Spatial Error Model is 114,629, smaller than the Spatial Autoregresive model. It can be concluded that Spatial Error Model is the best model.

TABLE V. PARAMETER ESTIMATION OF SEM

Variable	Coef.	Standard error	Prob.
λ	0.679	0.1090	0.0000
constanta	37,067	2.9520	0.0000
x_1	0.001	0.0001	0.0000
x_2	- 0.064	0.0420	0.1280
x_3	- 0.208	0.0410	0.0000
x_4	1.799	0.2310	0.0000

B. *Discussion*

HDI is an indicator used to see development progress in the long term. To see the progress of human development, there are two aspects that need to be considered, i.e. the speed and status of achievement. The average East Java HDI in 2016 was 0.69792 (69.792%) and included in the medium HDI category. There are eighteen districts / cities that have a HDI above the provincial average, while districts districts have an HDI below the provincial average. Human development achievements in an area at certain times can be categorized into four groups, i.e. very high : HDI ≥80, high : 70≤HDI<80, medium : 60≤HDI<70, and low : HDI<60. This grouping aims to organize regions into groups that are the same in terms of human development [7].

There are three cities that have a very high HDI category, i.e. Malang, Surabaya and Madiun. The total districts /cities in East Java which has a high category HDI is fourteen districts/cities. A moderate category of HDI districts/cities in East Java is twenty districts. One district with a low HDI category is Sampang district.

A long and healthy life dimension is represented by morbidity. The lowest morbidity of the people of East Java in 2016 was Gresik district, which was 9.843, while the highest morbidity was Sampang district, which was 25.093.

Knowledge dimension is represented by mean years of schooling. The lowest mean years of schooling of East Java public in 2016 was Sampang district, which was 11.37 years, while the highest was Malang city, which is 15.38 years

Dimensions that represent the quality of human life are decent standard of living represented by per capita expenditures (constant prices 2012) and percentage of poor people. In 2016, per capita expenditure of the people of East Java reached Rp. 10,561 million per year. The lowest per capita expenditure is Sumenep Regency, which is Rp 7.846 million per year, while the biggest per capita expenditure is the public expenditure in Surabaya, which is Rp 16.295 million per year. The lowest percentage of poor people in East Java in 2016 was Malang city, which was 4.33%, while the largest percentage of poor people was Bondowoso, which was 25%.

IV. CONCLUSION

Based on the results of the analysis it can be concluded that the best modeling for the district / city Human Development Index in East Java in 2016 is the Spatial Error Model because it has the AIC value smaller than the Spatial Autoregressive Model.

REFERENCES

[1] BPS, "Indeks Pembangunan Manusia Indonesia 2016," 2017.

[2] A. Melliana, and I. Zain, "Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten / Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel," *J. Sains dan Seni POMITS* , vol. 2, no. 2, pp. 237-242, 2013.

[3] D. W. Safitri, M. Y. Darsyah, T. W. Utami, "PEMODELAN SPATIAL ERROR MODEL (SEM) UNTUK INDEKS PEMBANGUNAN MANUSIA (IPM) DI PROVINSI JAWA TENGAH," *J. Statistika , vol*. 2, no. 2, pp. 9-14, 2014.

[4] R. P. Yakunina and G. A. Bychkov, "Correlation Analysis Of The Components Of The Human Development Index Across Countries," *Procedia Econ. Financ.*, vol. 24, no. July, pp. 766–771, 2015.

[5] E. W. Nugrahadi and M. Rinaldi, "Analysis of Human Development Index : Concept and Factors that Influence North Sumatera Province," vol. 46, no. Ebic 2017, pp. 27–31, 2018.

[6] Q. Qiu, J. Sung, W. Davis, and R. Tchernis, "Using spatial factor analysis to measure human development," *J. Dev. Econ.*, vol. 132, no. August 2016, pp. 130–149, 2018.

[7] BPS Provinsi Jawa Timur, "Berita resmi statistik," no. 25, pp. 1–7, 2017.