Human Development Index Modelling In Indonesia Using Spatial Error Model Approach

Asmadhini Rahma
{ahr@umkt.ac.id}

Muhammadiyah University of East Kalimantan, Indonesia

Abstract. Human Development Index (HDI) is the benchmark of the success of a country in establishing the quality of human life. One way to strengthen the HDI is to create a model based on several things that are expected to influence the increase in each province in Indonesia based on the three dimensions that make HDI, i.e Life Expectancy, Expected Years of Schooling, Number of Poor People, per Capita Expenditure, and Open Unemployment Rate. To determine the best model by weighing that a region can affect its neighboring regions, so that the spatial model is selected. After implementing data analysis, through the Lagrange Multiplier test it was obtained that the best model needed to determine what strategy to do to strengthen HDI per province in Indonesia is a spatial error model.

Keywords: Human Development Index, Life Expectancy, Expected Years of Schooling, Number of Poor People, per Capita Expenditure, Open Unemployment Rate, Spatial Error Model.

1. Introduction

Indonesia is a country consisting of many islands, which are often also called archipelago countries. Currently, Indonesia is divided into 34 provinces. The number of people spread across Indonesia and the area of territory that has made Indonesia become one of the countries with the biggest 4 (four) population in the world after China, India and the United States. [1] Indonesia's current population is estimated at 268 million people with a population growth rate of 1.3% annually which means there is a population increase of 3, 5 million annually. The interesting thing is, with the unique state of Indonesia, whether Indonesians live in good quality?

To answer these questions, we can see Human Development Index Indonesia (HDI). One important indicator of measuring the success of a country or region in building the quality of human life is through HDI. [2] HDI can determine the rank or level of development of a region or country. In Indonesia, HDI is a strategic data that is used as one of the determination of the General Allocation Fund allocator. HDI is measured by BPS based on 3 (three) basic dimensions, i.e. longevity & healthy Living, knowledge and decent living standards [3].

HDI Indonesia is currently ranked 6th in Southeast Asia. Head of UNDP Indonesia representative Christophe Bahuet said, the inclusion of Indonesia in the category is due to a strong national commitment to build human beings in terms of health, education, and welfare of the people [4]. However, Indonesia still needs to reduce and anticipate inequality in the future. This means that it is necessary to strengthen the HDI to overcome the problems of the inequality in the future.

One of the ways to strengthen HDI is to form a model based on some things that allegedly affect the increasing HDI in Indonesia. [5] To observe this HDI pattern, it is not enough to use only a double linear regression analysis, because the influence of the location of a province has an influence on the HDI pattern in other provinces which is a spatial effect, this is called spatial dependence consisting of 2 models, namely spatial Autoregressive (SAR) and spatial Error Model (SEM).

To form a spatial model, it takes many variables that affect the HDI. Based on the 3 dimensions that make HDI, a Life Expectancy is chosen which represents the dimension of longevity and healthy living, Expected Years of Schooling representing the dimension of knowledge, then, [3] Number of Poor People, per Capita Expenditure, and Open Unemployment Rate representing the standard of decent living. These five variables are expected to form the best model for calculating HDI per province in Indonesia.

2. Method

2.1 Data sources and research variables

This research data is a secondary data taken from the website of Badan Pusat Statistik (BPS) of Indonesia, which is www.bps.go.id . The variables used were chosen based on the theoretical study of previous studies and also the elements that formed the HDI in the form of a basic dimension i.e. longevity & healthy Living, knowledge and standards of decent living. Based on these two reasons, the independent variables in this study are life expectancy (x_1), expected years of schooling (x_2), number of poor people (x_3), per capita expenditure (x_4), and open unemployment rate (x_5)

2.2 Analysis Methods

The stages that are performed to get the best relationship pattern as a model, in search of the variables that make HDI, are as follows:

First. Collecting the necessary data.

Second. Inputting data on Geoda software.

Third. Create thematic maps to define neighboring criteria based on quantile to group HDI variables by value.

Forth. Determining the associated spatial matrix. The formation of related spatial matrix or matrix W uses 3 kinds of techniques of weighted technique, namely contiguity, Distance and General. Contiguity is a matrix formed based on the touch of the boundaries, so it is stated that spatial interaction occurs between regions that have the touch of the boundaries, in which case the region is a province. The main metric and diagonal is zero-value. There are three types of spatial interactions: Rook Contiguity, Bishop Contiguity and Queen Contiguity. In this study, the Queen Contiguity was chosen which is a side and Vertek point, between one region and another [6].

Fifth. Spatial dependencies test or autocorrelation between regions by using Moran's I for each variable. The effect of spatial dependencies occurs due to dependencies in regional data. Spatial dependencies test in a model error can be done using the statistics of Moran's I. Moran's I is a test to identify a location from a spatial grouping that sees a correlation between

variables and themselves based on space. With the expectation value of Moran's I are as follows [6].

$$E(I) = I_o = -\frac{1}{n-1} \qquad (1)$$

if $I > I_o$ then, autocorrelation have a positive value and Data patterns form the group, while $I = I_o$ means there is no autocorrelation, and $I < I_o$ means the autocorrelation value is negative, so the data pattern spreads.

Sixth. Test the Lagrange Multiplier for more specific identification of spatial effects i.e. dependency lag, error or both (lag and error) between provinces. In making a decision, reject H_0 if $LM > \chi^2(1)$ or $p - value < \alpha$. Thus, there can be a spatial dependency. The results of this test are used to determine the choice of spatial models [7].

Seventh. Test spatial heterogeneity with the Breusch Pagan test to see the presence or absence of diversity between regions. In making a decision, reject H_0 if $p - value < \alpha$. Thus, it can mean that there is diversity between provincial areas [8].

Eighth. Create equations by using the choice of spatial models, in this case using the model of the Spatial Error Model (SEM). The SEM model is as follows [9]:

$$Y = X\beta + u \qquad (2)$$

Where

$$u = \lambda W_2 u + \varepsilon \qquad (3)$$

The SEM Model appears due to the dependency of the error value of a location with errors in other locations associated with it. This occurs when there are variables that affect the value of the dependent variable but are not included in the model, correlated between locations [10].

3. Result And Discussion

The neighboring criteria using thematic maps are applied to the HDI data to divide the data into 3 groups by sharing the region.
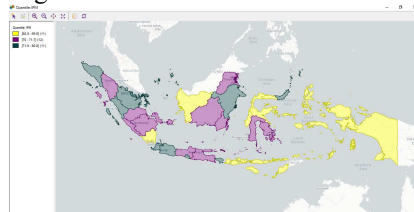

Fig.1. The Neighboring Criteria

Fig. 1. is a thematic map that classify provinces by the number of HDI. Based on the geographical location of the thematic map of Fig. 1, it is apparent that each province in the region tends to be adjacent. Geographically, it is indicated that there is spatial influence on HDI data per province in Indonesia. Group 1 is a provincial group that has the least number of HDI, while Group 3 is the provincial group that has the highest number of HDI. Based on the provincial data that has the most small HDI is Gorontalo and the greatest is West Java.

The neighboring criteria that can be used are border contact. Border contact is one of the criteria used to define neighbors for a region. The trick is to see the territories that are directly bordered (land) with other territories. Territories that directly border with other territories are assumed to have more significant influence. In other words, directly bordered territories can be assumed to be neighboring. In this study was used by Queen Contiguity in the

determination of neighbors, so that neighboring territories were determined based on boundary-side contact or border crossing with other regions. By using Queen Contiguity then gained that Aceh only has 1 neighbor, namely North Sumatra, Bali has 2 neighbors namely East Java and NTB, Bengkulu has 4 neighbors namely West Sumatra, Jambi, South Sumatera, Lampung, and so on.

The weighted matrix with the neighboring criteria of Queen continguity to be used is the matrix W which is the normality form of the W_{Queen}. The matrix is as follows :
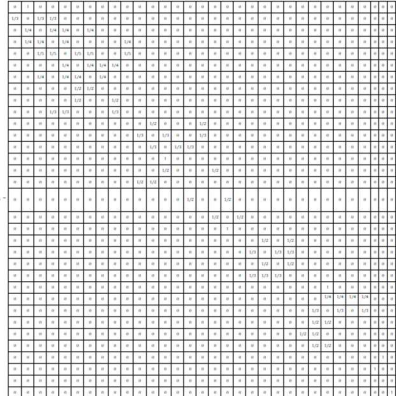

Fig.2. Matrix of Neighboring Criteria

Implement spatial effect testing to get whether the data of each variable has a spatial effect on the site. Test spatial dependencies by using Moran's I. After knowing the value of Moran's Ieach variable then calculates the value of I_0 is $-0,030303$. Overall the value of the variable is greater than I_0 which means the variable has a positive autocorrelation. Then it can be said spatial autocorrelation in all variables is strong. Positive spatial autocorrelation indicates the similarity of values and adjacent territories and tends to be grouped. In this discussion is the province of Indonesia, so it can be concluded that inter-provincial one with the other has a similarity of value or number of HDI is relatively the same.

The selection of spatial models was conducted with the LM test as an initial identification. Lagrange Multiplier is used to detect more specific spatial dependencies i.e. dependency lag, error or both (lag and error). According to LM test results., it is known that the value of the probability smaller than α, is Test Lagrange Multiplier (Error) and test Lagrange Multiplier (SARMA), which can mean that the reject H_0 which means there is spatial dependencies in the regression Error, so in this case it needs to be continued on creating a Spatial Error Model (SEM) model.

Furthermore, the testing of spatial effects is the test of spatial heterogeneity that is by using the Breusch Pagan test H_0 is that there is no diversity between regions. In this case the prob value of the Breusch Pagan test was 0.70465, so obtained a decision to receive H_0 which means that there is no diversity or heterorogentity between the provincial areas.

Table 1. *Spatial Error Model* (SEM)

Variable	Coefficient	Prob
Constanta	-1,38988	0,00689
Life Expetancy	0,585085	0,00000
Expected Years of Schooling	1,45777	0,00000

Number of Poor People	-0,00033	0,00571
per Capita Expenditure	0,001035	0,00003
Open Unemployment Rate	-0,291356	0,00000
Lambda	0,175485	0,00297

Based on table 1. it is obtained that Constanta and all variables including lambda values (λ) significantly affect the HDI because the value of its prob is smaller than 0.05. The coefficient of λ, which is significant suggests that if a region is surrounded by another region as much as N, then the influence of each area that surrounds it can be measured by 0.175485. Therefore formed an equation of model of the Spatial Error Model (SEM) as follows:

$$Y_i = -1,38988 + 0,585085X_{1i} + 1,45777X_{2i} - 0,000333X_{3i} + 0,00103585X_{4i} - 0,291356X_{5i} + U_i$$

Where,

$$U_i = 0,175485 \sum_{i,j=1}^{34} W_{ij}U_j + \varepsilon_i$$

Description :
Y_i : HDI in Province to i
X_{1i} : Life Expetancy in Province to i
X_{2i} : Expected Years of Schooling in province to i
X_{3i} : Number of poor people in province to i
X_{4i} : Per Capita expenditure in province to i
X_{5i} : Open unemployment rate in province to i
W_{ij} : Spatial weighing matrix
U_i : Residual Spatial from province to i
ε_i : Residual from Province to i

The inter-provincial spatial error is correlated by 0.175485, so the magnitude of spatial interaction between 34 provinces in Indonesia has a area of 0.175485. The equation resulted in every province having the same neighbor number would have the same model of equation. Provinces that have a neighbouring number of 2, there are 12 provinces. Thus, the model of the equation of one of the provinces of Bali that has the number of neighbors 2, is as follows :

$$Y_{Bali} = -1,38988 + 0,585085X_{1Bali} + 1,45777X_{2Bali} - 0,000333X_{3Bali} + 0,00103585X_{4Bali} - 0,291356X_{5Bali} + 0,0877425(U_{Jawa\ Timur} + U_{NTB})$$

In this case, HDI a province will be negative if all variables are life expectancy, expected years of schooling, number of poor people, per capita expenditure, and open employment rate has a value of 0. However, it will increase if life expectancy, expected years of schooling and per capita expenditure increase, the reverse, will decrease if number of poor people and open unemployment rate increases. Meanwhile, the error value of East Java and NTB affects the spatial error model number HDI in Bali amounted to 0.0877425.

The effort to calculate the human development index to the smallest level is very important because of the decentralized process (regional autonomy) in Indonesia. This, will surrender most of the development process to local governments and communities, which is expected to

understand more about the local conditions and supported by the data that is more adequate for the province concerned [5].

Areas with high development will provide a variety of benefits especially in terms of service. The service will continue to increase with the increasing development, both economic related services (trade, job field), as well as education and health areas. Based on the theory, it can be understood that when a region has high development it can also provide services to other regions surrounding it. Furthermore, the adjacent area will also increase the HDI [11].

Based on the model it is confirmed that if a province wants to improve HDI, then increase life expectancy, expected years of schooling and per capita expenditure and leave the number of poor people and open unemployment rate. If it is abused, this is true because a province will have a good HDI if the community in the province has a high expectancy life, because with a high expectancy life means society is living well and is in a healthy environment. Likewise, in expected years of schooling, if the average community has a good education, then the level of life will be much better than the lack of education. This also has a lot of impact on the province, namely with reduced unemployment, because the average community is absorbed in various jobs or entrepreneurial well because it has the knowledge as adequate provision.

Per capita expenditure had a positive impact on the HDI, although actually it was like a double-edged knife, on the one hand it would have a good impact, but on the other side would have a bad impact if it was too excessive. An area would be assessed to have a good economic turnover if the per capita expenditure is high, because it means that if their expenditure is high, then their standard of living is good, their life needs a lot and certainly their income is good. In contrast, adverse effects will arise if this state is excessive, because it can be assessed that the area is too consumptive, so it will appear jammed credit and other economic impacts are bad. HDI will certainly decrease if number of people increases. Obviously it is seen if an area is experiencing increased poverty, meaning that the area is not in good condition, so is the open unemployment rate.

With this spatial model, then to calculate the HDI a province can not be called flatten the model, because each province has its own uniqueness. This is in line with the situation that Indonesia has diversity in every region. In this spatial error model, every model per province depends on the error value of each neighbor. Therefore, each province has a close relationship that affects each other against a neighboring province that is in contact with a point or line with it.

4. Conclusion

Based on the results of the analysis it can be concluded that the best modeling for the Human Development Index per provence in Indonesia is the Spatial Error Model. Therefore formed an equation of model of the Spatial Error Model (SEM) as follows:

$$Y_i = -1{,}38988 + 0{,}585085X_{1i} + 1{,}45777X_{2i} - 0{,}000333X_{3i} + 0{,}00103585X_{4i} - 0{,}291356X_{5i} + U_i$$

Where,

$$U_i = 0{,}175485 \sum_{i,j=1}^{34} W_{ij}U_j + \varepsilon_i$$

References

[1]   www.bps.go.id, "STATISTIK INDONESIA 2020," vol. 1101001, 2020, [Online]. Available: https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html.

[2]   www.bps.go.id, "Indeks Pembangunan Manusia." https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html#subjekViewTab1.

[3]   W. Pramesti and A. Indrasetianingsih, "East Java Human Development Index Modeling with Spatial Regression Approach," vol. 226, no. Icss, pp. 1494–1498, 2020, doi: 10.2991/icss-18.2018.312.

[4]   UNDP., "Briefing note for countries on the 2019 Human Development Report: Indonesia.," *Hum. Dev. Rep. 2019 Inequalities Hum. Dev. 21st Century*, pp. 1–9, 2019, [Online]. Available: http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/NZL.pdf.

[5]   F. Samallo, Y. A. Lesnussa, A. Z. Wattimena, and M. Y. Matdoan, "Pemodelan Spatial Error Model (SEM) Untuk Mengidentifikasi Indeks Pembangunan Manusia di Provinsi Maluku Tahun 2016," *J. Mat.*, vol. 8, no. 1, p. 75, 2018, doi: 10.24843/jmat.2018.v08.i01.p99.

[6]   H. M. Pohan, "Ekonometrika Spasial :," *Univ. Parahiyangan*, 2019.

[7]   W. Sanusi, H. Ihsan, and H. Syam, "Model Regresi Spasial dan Aplikasinya dalam Menganalisis Angka Putus Sekolah Usia Wajib Belajar di Provinsi Sulawesi Selatan," vol. 1, no. 2, pp. 183–192, 2018.

[8]   E. Hari Samadi, Yudiantri Asdi, "Penerapan Model Regresi Spasial Dalam Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan," *J. Mat. UNAND*, vol. VI, no. 4, pp. 80–89, 2017.

[9]   Y. Sari, N. K. Dwidayati, and P. Hendikawati, "Estimasi Parameter pada Regresi Spatial Error Model (SEM) yang Memuat Outlier menggunakan Iterative Z Algorithm," *Prisma*, vol. 1, pp. 456–463, 2018.

[10]   L. Mubatdiah, "Estimasi Parameter Model Regresi Spasial Error Dengan Metode Maximum Likelihood Estimation," UIN Maulana Malik Ibrahim, Malang, 2011.

[11]   O. Hamdan, "RASIO TENAGA PENDIDIK, RASIO TENAGA KESEHATAN, DAN CAPAIAN PEMBANGUNAN MANUSIA DI INDONESIA DALAM ANALISIS SPASIAL," *J. Ekon. dan Kebijak. Publik Indones.*, vol. Volume 6, p. No. 2, ISSN: 2442-7411, 2019.