Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Spatial panel data model on human development index at Central Java

To cite this article: S Yulianti *et al* 2021 *J. Phys.: Conf. Ser.* **1722** 012090

Spatial panel data model on human development index at Central Java

S Yulianti*, Y Widyanigsih, and S Nurrohmah

Department of Mathematics, Universitas Indonesia, Kampus Baru UI, Depok, 16424, West Java, Indonesia

*salsabila.yulianti@sci.ui.ac.id

Abstract. Human development is one of the ways to create human resources that have the ability to drives the economy. It is become the focus of Indonesia's government attention and listed on several programs implemented at the global and national levels. Human development is measured by one composite index known as the Human Development Index (HDI). HDI is often used as a measure of government success in human development. Human development in Indonesia could vary because it depends on the policies of each region, so in a region, it could give an influence between regencies or cities. In this study, focus of attention is Central Java because HDI in this region has always increased from 2017 to 2019. Therefore, this study aims to analyze the factors that affect HDI on regencies or cities in Central Java from 2017 to 2019. Factors used in this study based on previous research are gross regional domestic product, high school participation rates, population density, labor force participation rates, percentage of poor population, and percentage of households with proper sanitation. Based on this study the observed data contains location information, so these observations could be spatially correlated. Method that could handle the issue and also observed in three time periods (years) is spatial panel data model. The spatial panel data model in this study is divided into two, known as spatial lag panel data model and spatial error panel data model. Based on the comparison of two models, it is obtained that the best model could explain the Human Development Index at Central Java from 2017 to 2019 is spatial lag panel data model with all the explanatory variables significant.

1. Introduction

Indonesia is the fourth most populous country in the world with population of 274 million. This could be Indonesia's advantage or weakness if the government does not take the appropriate actions to cope with population growth. The weakness if the government does not cope with the population growth is unemployment which could cause poverty so the population could not fulfill adequate nutrition and place because of their limitation. These limitations also could cause disease and short-lived the population. Therefore, the primary actions that could be taken by the government to cope with population growth is to empower people to be able to move the wheels of the economy by creating capable human resources, and this action is known as human development. Human development is a measure of overall development performance and is formed with three basic components, namely: long and healthy life, have knowledge, and have decent standard of living [1]. These three components summarize one composite index, known as the Human Development Index (HDI). The Human Development Index was first introduced by the United Nations Development Program (UNDP) in 1990.

Human development in Indonesia could vary because it depends on the policies of each region, but the focus of attention in this study is Central Java because HDI in this region has always increased from 2017 to 2019. Therefore, this study discusses Human Development Index (HDI) in regencies or cities at Central Java from 2017 to 2019. The observed data in the study contains location information and also observed in several periods (years) and this data known as panel data. The dependent variable on this study is Human Development Index, with explanatory variables are gross regional domestic product, high school participation rates, population density, labor force participation rates, percentage of poor population, and percentage of households with proper sanitation. The advantages of using panel data are could control regional heterogeneity, more informative, better able to study dynamic changes because the data is observed in several periods so this occasion could describe the explanatory variables that affect dependent variable from three times periods [2].

Another issue that could arise when the observed data contains location information is the observations could be spatially correlated (have spatial autocorrelation) and also have spatial heterogeneity. Spatial autocorrelation is condition where the value of observations at a location is influenced by the observed value of its closest neighbor [3]. Spatial heterogeneity is condition of the diversity of functional forms and parameters at each location so that it could have different models for each location [4]. The multiple regression linear model could not describe these issues. Therefore, if spatial autocorrelation and spatial heterogeneity are indicated but it is not included in the model as in multiple linear regression, the resulting estimates will be biased and inconsistent [4]. The model capable of involving spatial autocorrelation and spatial heterogeneity that observed in several periods is spatial panel data model. This study uses a spatial panel data model with parameter estimation is using the Maximum Likelihood method. The spatial panel data model used in this study is divided into two that are the spatial lag panel data model (SAR) and spatial error panel data model (SEM).

2. Method

This study discusses the Human Development Index in thirty-five regencies or cities Central Java Province which is observed from 2017 to 2019, and total observation are 105 units. The reason for choosing Central Java as a province in the study because Central Java succeeded in increasing the HDI and changed the province category from the “medium” category in 2016 to category "high" in 2017 [5]. Based on data from BPS, the Human Development Index in Central Java has always increased from 2017 to 2019, that is 70.52; 71.12; and 71.73, so the further will be inspected more.

2.1 Multiple Linear Regression

Multiple linear regression model is a model to analyze more than one explanatory variable that affect dependent variable. This model assumes single relationship for all data samples (regardless of location and time of data). The following is a general equation for linear regression [4]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \quad (1)$$

where y is dependent variable, β_p is regression parameter on p, x_p is explanatory variables on p, ε is error term, with $p = 1,2, \dots k$, k is number of explanatory variables.

2.2 Spatial Analysis

At the time when the observed data contains location and time information, so it tends to be spatially correlated (spatial autocorrelation). Spatial autocorrelation is condition where the value of observations at a location or region is influenced by the observed value of its closest neighbor. Spatial autocorrelation could be detected with Moran’s I test [6]. At the time when spatial autocorrelation is indicated, it is needed a model that includes spatial autocorrelation that is spatial panel data model. Spatial panel data is a compound of observations that contain location information (postal code, region, state, country, etc) and observed in several periods. Spatial panel data model is model that describes the interaction of location in several periods. The spatial panel data model controls spatial interactions at location and time [7]. Spatial panel data model is divided into two, namely spatial lag panel data model and spatial error panel data model.

1. Spatial Lag Panel Data Model

Spatial lag panel data model or spatial autoregressive model (SAR) is model that the dependent variable depends on the dependent variable observed in neighboring units (observed local characteristics) or the dependent variables correlate between locations [8]. The following is the equation for the spatial lag panel data model [9]:

$$y_{it} = \delta \sum_{j=1}^{N} w_{ij}\, y_{jt} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \mu_i + \varepsilon_{it};\ i = 1,2,..N, t = 1,2 \dots, T \tag{2}$$

where i is index for location or spatial unit with $i = 1,2, \dots, N$, t is index for time periods with $t = 1,2, \dots, T$. y_{it} is observation on dependent (response) variable at i and t, $\boldsymbol{x}_{it}$ is (1 x K) vector of explanatory variable at i and t, δ is spatial autoregressive coefficient, w_{ij} is element of spatial weight matrix $\boldsymbol{W}$, μ_i is spatial specific effect at i , ε_{it} is error term with $\varepsilon_{it} \sim NIID(0, \sigma^2)$.

The estimation of parameters in the spatial lag panel data model is using the Maximum Likelihood (ML) method. Log-likelihood function in the spatial lag panel data model with spatial specific effects [8]:

$$Ln\, L = -\frac{NT}{2} Ln(2\pi\sigma^2) + T Ln|(\boldsymbol{I}_N - \delta\boldsymbol{W})| - \frac{1}{2\sigma^2}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left(y_{it} - \delta\sum\nolimits_{j=1}^{N} w_{ij}y_{jt} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \mu_i\right)^2 \tag{3}$$

The parameter estimates in the error spatial panel data model are:

$$\hat{\mu}_i = \frac{1}{T}\sum\nolimits_{t=1}^{T}\left(y_{it} - \delta\sum\nolimits_{j=1}^{N} w_{ij}y_{jt} - \boldsymbol{x}_{it}\boldsymbol{\beta}\right) \tag{4}$$

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y}^* - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{y}^* - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W}_N)\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta})}{NT} \tag{5}$$

$$\widehat{\boldsymbol{\beta}} = ((\boldsymbol{X}^*)'\boldsymbol{X}^*)^{-\mathbf{1}}(\boldsymbol{X}^*)'\boldsymbol{Y}^* - ((\boldsymbol{X}^*)'\boldsymbol{X}^*)^{-\mathbf{1}}(\boldsymbol{X}^*)'\boldsymbol{\delta}(\boldsymbol{I_T}\otimes\boldsymbol{W_N})\boldsymbol{y}^*) \tag{6}$$

with $\boldsymbol{y}^* = \boldsymbol{Q}\boldsymbol{y}$ $\quad \boldsymbol{X}^* = \boldsymbol{Q}\boldsymbol{X}$ $\quad \boldsymbol{Q} = \boldsymbol{I}_{NT} - \frac{1}{T}\boldsymbol{\iota}_T(\boldsymbol{\iota}_T)'\otimes\boldsymbol{I}_N$

Estimate of δ is obtained by substituting $\hat{\mu}_i$ and $\widehat{\boldsymbol{\beta}}$ into the log-likelihood function, then a numerical iteration procedure is added to obtain an estimate of δ.

2. Spatial Error Panel Data Model

Spatial error panel data model (SEM) is model that has a spatial dependence on correlated errors between locations [8]. The spatial dependence in the panel data spatial model error is in the error or nuisance dependence [6]. The following is the equation for the spatial panel data model error [9]:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \mu_i + \phi_{it}, \text{ where } \phi_{it} = \rho \textstyle\sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} w_{ij}\, \phi_{it} + \varepsilon \tag{7}$$

with ϕ_{it} is spatially autocorrelated error term, and ρ is spatial autocorrelation coefficient.

The estimation of parameters in the spatial error panel data model is using the Maximum Likelihood (ML) method. The log-likelihood function in the spatial error panel data model with spatial specific effects [8]:

$$Ln\ L = -\frac{NT}{2}Ln(2\pi\sigma^2) + TLn|(\boldsymbol{I}_N - \rho\boldsymbol{W}_N)| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \rho\sum_{j=1}^{N}w_{ij}y_{jt} - \sum_{k=1}^{N}\beta_k(x_{itk} - \rho\sum_{j=1}^{N}w_{ij}x_{jtk}) - (\mu_i - \rho\sum_{j=1}^{N}w_{ij}\mu_j)\right)^2 \tag{8}$$

The parameter estimates in the error spatial panel data model are:

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta}) \tag{9}$$

$$\hat{\sigma}^2 = \frac{[(\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \rho\boldsymbol{W}_N))(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta})]'(\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \rho\boldsymbol{W}_N))(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta})}{NT} \tag{10}$$

$$\hat{\boldsymbol{\beta}} = [(\boldsymbol{X}^* - (\boldsymbol{I}_T\otimes(\boldsymbol{I}_N - \rho\boldsymbol{W}_N)\boldsymbol{X}^*)'(\boldsymbol{X}^* - (\boldsymbol{I}_T\otimes(\boldsymbol{I}_N - \rho\boldsymbol{W}_N)\boldsymbol{X}^*)]^{-1} \times (\boldsymbol{X}^* - (\boldsymbol{I}_T\otimes(\boldsymbol{I}_N - \rho\boldsymbol{W}_N)\boldsymbol{X}^*)'(\boldsymbol{Y}^* - (\boldsymbol{I}_T\otimes(\boldsymbol{I}_N - \rho\boldsymbol{W}_N)\boldsymbol{y}^*) \tag{11}$$

with $\boldsymbol{y}^* = \boldsymbol{Q}\boldsymbol{y}$ $\quad \boldsymbol{X}^* = \boldsymbol{Q}\boldsymbol{X}$ $\quad \boldsymbol{Q} = \boldsymbol{I}_{NT} - \frac{1}{T}\boldsymbol{\iota}_T(\boldsymbol{\iota}_T)'\otimes\boldsymbol{I}_N$

Estimate of ρ is obtained by substituting $\hat{\mu}_i$ and $\hat{\boldsymbol{\beta}}$ into the log-likelihood function, then numerical iteration procedure is added to obtain an estimate of ρ.

3. Result and Discussion

In this section, to determine the explanatory variables that affect Human Development Index (HDI or IPM) or the dependent variable in Central Java with multiple linear regression model. Multiple linear regression model is used whether the model is good to predict the dependent variable (HDI) in Central Java from 2017 to 2019, regardless of location and time information. Dependent variable is used in this study is: Human Development Index (HDI), and the explanatory variables are labor force participation rates (x_1), high school participation rates (x_2), percentage of households with proper sanitation (x_3), gross regional domestic product (x_4), percentage of poor population (x_5), population density (x_6), t2018 is a dummy variable with entry 1 for the year of 2018, 0 for another, and t2019 is a dummy variable with entry 1 for the year of 2019, 0 for another.

Table 1. Estimate value on multiple linier regression model and residual analysis

	Estimate	**Standard Error**	**t-value**	**p-value**	**VIF**
Intercept	50,284	6,5191	7,7134*	1,143e^{-11}	
x_1	0,10113	0,078428	1,2895	0,2003291	1,256876
x_2	0,18613	0,026524	7,0175*	3,195e^{-10}	1,495187
x_3	0,025616	0,013611	1,8820	0,0628694	1,656936
x_4	0,00003718	0,0000088614	4,1957*	6,070e^{-5}	1,201517
x_5	-0,27536	0,072884	-3,7781*	0,0002742	1,755575
x_6	0,0007761	0,00010521	7,3764*	5,789e^{-11}	1,402485
t2018	-0,19819	0,50145	-0,3952	0,6935448	
t2019	0,056906	0,50839	0,1119	0,9111106	
F test	48,59			$2{,}2e^{-16}$	
R-squared	0,80193				
Lilliefors test	0,053568			0,6482	
Breusch-Pagan test	6,626			0,5775	
Durbin-Watson test	0,97248			1,418e^{-8}	

*significant

The F test is used to see whether the model is useful for predicting the value of the Human Development Index (HDI) or dependent variable. As seen in Table 1, the F test show that the model is statistically useful for predicting the value of the HDI (*p-value* = $2{,}2e^{-16}$ < 0,05). The explanatory variables in the model that affect the dependent variable is marked by star symbol (*), that are the high school participation rate, gross regional domestic product, percentage of poor people, and population density. The dummy variable in 2018 (t2018), and dummy variable in 2019 (t2019) are not significantly affected by HDI.

The multiple linear regression model requires the classic assumptions that must be fulfilled. The classic assumptions are the assumption of normality, homoscedasticity, no autocorrelation, and no multicollinearity [10]. The formal test used for the normality, homoscedasticity, and autocorrelation are the Liliefors test, Breusch-Pagan test, and Durbin Watson test. The result of the normality test is concluded that the residuals are generated normally distributed model (p-value = 0.6482 > 0.05). The result of the homoscedasticity test is no heteroscedasticity symptoms (p-value = 0.5775> 0.05). The result of the autocorrelation test is indicated there is autocorrelation in the residuals in the model (p-value = $1{,}418e^{-8}$ < 0.05). The classic assumption for no multicollinearity is using the VIF value. The result is no multicollinearity in all explanatory variables in the observation (VIF value < 10).

Based on the tests, there is an unfulfilled model assumption (there is autocorrelation), so the use of multiple linear regression models could not be used because it causes the parameter estimates in the model to be biased and inconsistent [2]. Violation of this autocorrelation assumption could result from model that is indicated to have spatial correlation [3].

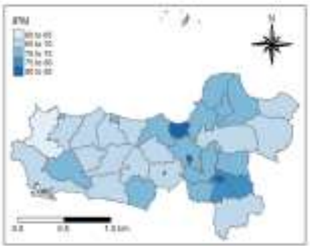

Figure 1. Map of human development index distribution on 2017

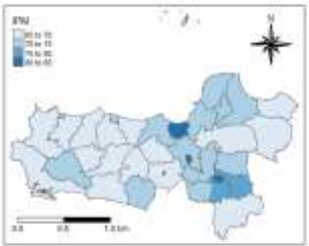

Figure 2. Map of human development index distribution on 2018

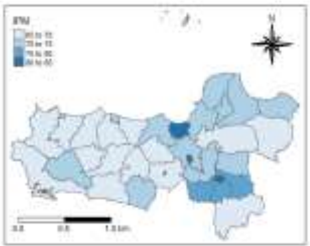

Figure 3. Map of human development index distribution on 2019

Figures 1 to 3 show that there is a tendency for a similar spatial color pattern to gather in a region from 2017 to 2019 on Human Development Index. These figures show that high values (dark color) tend to cluster in a region and low values (light color) tend to another region, so it could be possible for each variable there is a tendency too. Based on the figures it could occur the possibility of spatial autocorrelation. At the time when there is the possibility of spatial autocorrelation, parameter estimates in multiple linear regression could generate biased and inconsistent [2]. Spatial autocorrelation among regencies or cities in Central Java could be detected with Moran's I Test [6]. Before use the formal test for detecting the spatial autocorrelation, we need to determine a spatial weight matrix. This study uses rook contiguity as a spatial weight matrix.

Table 2. Moran's I statistical value test of each explanatory variable and its p-value

Year	HDI	x_1	x_2	x_3	x_4	x_5	x_6
2017	2,37824*	1,3218	2,9016*	3.186*	2,7331*	-0,0628	0,67966
	(0,0086)	(0,09311)	(0,0018)	(0,00072)	(0,003)	(0,525)	(0,2484)
2018	2,3751*	1,86*	2,2699*	3,116*	2,2699*	-0,0019	0,69804
	(0,0087)	(0,03144)	(0,01161)	(0,00092)	(0,012)	(0,5008)	(0,2426)
2019	2,3751*	1,0644	2,2929*	3,1843*	2,0358*	-0,1179	0,71929
	(0,0087)	(0,1436)	(0,01093)	(0,00073)	(0,021)	(0,547)	(0,236)

*significant

Table 2 show that the statistical value of Moran's I test for all the variables for three years tends to have a spatial correlation or indication of spatial autocorrelation (p-value < 0.05). Therefore, model is needed to be able to control the spatial autocorrelation, so that a better estimate could be generated.

The model could include spatial autocorrelation observed in the three years is spatial panel data model. As seen in Table 2, some variables in the study do not indicate spatial autocorrelation but if an explanatory variable is omitted from the model equation when it should exist, the resulting estimates on the explanatory variable coefficients will be biased and inconsistent [8]. Therefore, all variables in the study were included in the spatial panel data model. The spatial panel data model in this study is divided into two, so to see whether dependence that occur is spatial lag or spatial error, the Lagrange Multiplier test is performed [8]. The Lagrange Multiplier test resulted in spatial error dependence (SEM) (p-value = $1,555e^{-15}$ < 0,05) and spatial lag dependence (SAR) (p-value = $7,885e^{-5}$ < 0,05) are significant, so that further will be testing estimates for two spatial panel data models to get the best model in this study.

Table 3. Estimate value on spatial panel data model and the test's

	SAR	SEM
x_1	0,13455*	0,14766*
x_2	0,099682*	0,10539*
x_3	0,041372*	0,037733*
x_4	0,000040713*	0,0000045515*
x_5	-0,34884*	-0,3734*
x_6	0,00083779*	0,000088966*
δ	0,15522*	-
ρ	-	0,20976
Likelihood-Ratio test	69,3204*	-126,131
R-squared	0,8982391	0,8899207
Mean Square Error	1,97809337	2,13978987

*significant

Table 3 show that the results of the spatial panel data model estimation. The likelihood ratio test is used for seeing whether the spatial panel data model is useful model for predicting the Human Development Index (HDI) in Central Java. The model that useful for predicting HDI is the spatial lag panel data model (LR = 69,3204 > 48,6024). Besides, the Lagrange Multiplier value for the spatial lag panel data model is also significant, which means that there is spatial interaction on the dependent variable between neighboring regencies or cities in Central Java. Goodness of fit model could measure with coefficient determination (R^2) model [9]. Model has the largest R^2 value with the smallest MSE is spatial lag panel data model. Therefore, the best model in this study to predict the Human Development Index at Central Java from 2017 to 2019 is spatial lag panel data model. As seen in Table 3, explanatory variables that affect Human Development Index in Central Java are the labor force participation rate, high school participation rate, gross regional domestic product, percentage of households with proper sanitation, percentage of poor people, and population density, and the spatial autoregressive coefficient.

The following is an estimate of the spatial lag panel data model for HDI data in Central Java from 2017 to 2019 whose parameter values as seen in Table 3:

$$\widehat{HDI}_{it} = 0{,}15522\sum_{j=1}^{35} w_{ij}HDI_{jt} + 0{,}13455{x_1}_{it} + 0{,}099682{x_2}_{it} + 0{,}041372{x_3}_{it} + 0{,}000040713{x_4}_{it} - 0{,}34884{x_5}_{it} + 0{,}00083779{x_6}_{it} + \hat{\mu}_i \quad (12)$$

with $\hat{\mu}_i$ is spatial specific effect as seen in Table 4, $i, j = 1,2, \dots ,35; i \neq j;\ t = 2017, 2018, 2019.$

The model interpretation is as follows:

1) A one percent increase in the labor force participation rate will increase the Human Development Index in Central Java by 0.13455 if the value of other variables is constant.

2) A one percent increase in the high school participation rate will increase the Human Development Index in Central Java by 0.099682 if the value of other variables is constant.
3) A one percent increase in the percentage of households with proper sanitation will increase the Human Development Index in Central Java by 0.041372 if the value of other variables is constant.
4) A one billion rupiahs increase in gross regional domestic product will increase the Human Development Index in Central Java by 0.000040713 if the values of other variables are constant.
5) A one percent increase in the percentage of poor people will reduce the Human Development Index in Central Java by 0.34884 if the value of other variables is constant.
6) A ten population/km^2 increase in the population density of an area will increase the Human Development Index in Central Java by 0.00083779 if the values of other variables are constant.

Table 4. Spatial specific effect estimation

ID	Province	Estimation	ID	Province	Estimation	ID	Province	Estimation
1	Cilacap	0,274933	14	Sragen	0,756489	27	Pemalang	-0,189168
2	Banyumas	0,805108	15	Grobogan	0,948189	28	Tegal	-0,375406
3	Purbalingga	1,812407	16	Blora	-0,0243409	29	Brebes	0,320111
4	Banjarnegara	2,580704	17	Rembang	0,447447	30	Magelang (City)	-0,539377
5	Kebumen	3,760421	18	Pati	1,465786	31	Surakarta (City)	-0,050457
6	Purworejo	1,380999	19	Kudus	-1,396716	32	Salatiga (City)	0,489700
7	Wonosobo	0,224708	20	Jepara	-1,224885	33	Semarang (City)	2,323550
8	Magelang	-1,124639	21	Demak	-3,146674	34	Pekalongan (City)	1,632060
9	Boyolali	-1,603206	22	Semarang	-1,787953	35	Tegal (City)	1,659600
10	Klaten	-0,832565	23	Temanggung	-2,322521			
11	Sukoharjo	-1,367476	24	Kendal	-1,311905			
12	Wonogiri	-0,907723	25	Batang	-1,825055			
13	Karanganyar	0,740665	26	Pekalongan	-1,373738			

To make complete model of spatial panel data model could be written by looking at concept of neighborliness (spatial weight matrix). Table 4 show that spatial specific estimation ($\hat{\mu}_i$) for spatial lag panel data model. Example model is shown for Cilacap Regency.

Figure 4. Map of neighborliness of spatial weight matrix

Figure 4 show that Cilacap Regency (i = 1) is close to Banyumas Regency (i = 2), Kebumen Regency (i = 5), and Brebes Regency (i = 29) so that the spatial lag panel data model in Cilacap Regency is:

$$\widehat{HDI}_{1t} = 0{,}15522\left(\frac{1}{3}HDI_{2t} + \frac{1}{3}HDI_{5t} + \frac{1}{3}HDI_{29t}\right) + 42{,}35 + 0{,}13455x_{1_{1t}} + 0{,}099682x_{2_{1t}} + 0{,}041372x_{3_{1t}} + 0{,}000040713x_{4_{1t}} - 0{,}34884x_{5_{1t}} + 0{,}00083779x_{6_{1t}} + 0{,}274933 \quad (13)$$

$$\widehat{HDI}_{1t} = 42{,}624933 + 0{,}05174(HDI_{2t} + HDI_{5t} + HDI_{29t}) + 0{,}13455x_{1_{1t}} + 0{,}099682x_{2_{1t}} + 0{,}041372x_{3_{1t}} + 0{,}000040713x_{4_{it}} - 0{,}34884x_{5_{1t}} + 0{,}00083779x_{6_{1t}} \quad (14)$$

4. Conclusion

Based on the discussion, this study indicates a spatial autocorrelation. To test whether the spatial panel data model can be used with the Lagrange Multiplier test, and the observed data have dependency on lag and error, so the spatial panel data model is used. The selected spatial panel data model in this study is the spatial lag panel data model based on the tests to compare models with likelihood ratio test, coefficient determination, and mean square error. Explanatory variables affect Human Development Index in Central Java are labor force participation rate (x_1), high school participation rate (x_2), percentage of households with proper sanitation (x_3), gross regional domestic product (x_4), percentage of poor people (x_5), and population density (x_6), or all the explanatory variables affect HDI in Central Java. It can be concluded that the spatial lag panel data model is the best model in this study to analyze the explanatory variables that affect the Human Development Index in thirty-five regencies or cities in Central Java from 2017 to 2019.

Acknowledgments

This research was funded by Directorate of Research and Development of Universitas Indonesia (DRPM UI) as a grant of Publikasi Terindeks Internasional (PUTI) Prosiding 2020 No. NKB-1028/UN2.RST/HKP.05.00/2020.

References

[1] Badan Pusat Statistik 2015 *Indeks Pembangunan Manusia 2014* (Jakarta: Badan Pusat Statistik)

[2] Baltagi B H 2005 *Econometrics Analysis of Panel Data* (Chichester: John Wiley & Sons LTD)

[3] LeSage J P 1999 The Theory and Practice of Econometrics *Economica* **48** 429

[4] Ervina M E, and Jaya I G N M 2018 Analyzing Factors Affecting GRDP in Indonesia Using Spatial Panel Data Model *Modern Applied Science* **13** 148

[5] Badan Pusat Statistik 2019 *Indeks Pembangunan Manusia 2018* (Jakarta: Badan Pusat Statistik)

[6] Anselin L 2001 *A Companion to Theoretical Econometrics Spatial Econometrics* (Oxford: Blackweel Publising Ltd)

[7] Millo G, and Miras P 2012 splm: Spatial Panel Data Models in R *Journal of Statistical Software* **47** 1

[8] Elhorst J P 2010 *Handbook of Applied Spatial Analysis* (Berlin: Springer)

[9] Elhorst J P 2014 *Spatial Econometrics From Cross-Sectional Data to Spatial Panels* (Dordrecht: Springer)

[10] Gujarati D N 2004 *Basic Econometrics* (New York: McGraw-Hill/Irwin)