Spatial durbin error model for human development index in Province of Central Java.

A.R. Septiawan¹, S.S. Handajani², and T.S. Martini¹

¹Department of Mathematics, Faculty of Mathematics and Sciences, Sebelas Maret University Jl Ir Sutami 36A Kentingan Jebres Surakarta 57126, INDONESIA
²Department of Statistics, Faculty of Mathematics and Sciences, Sebelas Maret University Jl Ir Sutami 36A Kentingan Jebres Surakarta 57126, INDONESIA

E-mail: bertho@unpad.ac.id

Abstract. The Human Development Index (HDI) is an indicator used to measure success in building the quality of human life, explaining how people access development outcomes when earning income, health and education. Every year HDI in Central Java has improved to a better direction. In 2016, HDI in Central Java was 69.98 %, an increase of 0.49 % over the previous year. The objective of this study was to apply the spatial Durbin error model using angle weights queen contiguity to measure HDI in Central Java Province. Spatial Durbin error model is used because the model overcomes the spatial effect of errors and the effects of spatial dependency on the independent variable. Factors there use is life expectancy, mean years of schooling, expected years of schooling, and purchasing power parity. Based on the result of research, we get spatial Durbin error model for HDI in Central Java with influencing factors are life expectancy, mean years of schooling, expected years of schooling, and purchasing power parity.

Keywords: HDI, spatial Durbin error model, spatial error of effect

1. Introduction

Human development aims to create an environment that allows people to enjoy longevity, health and live a productive life. According to the BPS [4], the success of development performance can be judged by how much of the most fundamental issues can be addressed, such as poverty, education, and food security. The United Nations (UN) sets a standard measure to determine the success rate of development performance, the Human Development Index (HDI).

According to the BPS [4], HDI in Indonesia is built through a basic three-dimensional approach. These dimensions include health, knowledge, and decent living. To measure the health dimension of HDI one of them can be used life expectancy at birth. Furthermore, to measure the dimensions of knowledge used combined between the indicator of the mean years of schooling and expected years of schooling. As for measuring a decent life dimension, purchasing power parity is used.

In this study used data HDI in 35 districts/cities from 2010-2016. Therefore, one of the data types that can be used to measure HDI in Central Java is panel data. Panel data is a unit of data cross section compiled periodically from time to time or in units time series. Geographical and demographic factors play a role in the growth of a region's HDI. This is indicated by the magnitude of HDI in a region can affect the value of HDI in adjacent areas. In this regard, the first law of space concerning the effect of the proximity Tobler has ever posited suggests that all things interconnected with one another but
something close to it have greater influence than something far away (Anselin [2]). The law is the basis of a problem that contains spatial effects.

In overcoming the existence of spatial effects on the panel data required statistical methods that can overcome the spatial effect on panel data i.e. spatial panel model. The spatial panel model is a method for obtaining observational information that is influenced by the effect of the space or unit of location present in the panel data. The spatial panel model that can be used to obtain information in an observation is the spatial Durbin error model. According to LeSage [7], the spatial Durbin error model is chosen because it can overcome spatial autocorrelation relationships on independent variables and can overcome spatial errors between regions. The spatial effect of errors results from errors obtained from an area dependent on errors in adjacent areas.

Previous research, including research on HDI modeling in 24 districts/cities in South Sulawesi Province in 2010-2014 conducted by Hamdani [6] with spatial Durbin fixed effect model. The results of this study obtained spatial Durbin model for HDI, but the effect of inter-regional interaction that occurred in the HDI data is not resolved in the spatial Durbin model. So in this study applied spatial Durbin error model for HDI in Central Java Province by adding interaction effect between regions and spatial effects on independent variables.

2. Theoretical Basis

2.1. Spatial Durbin Error Model

In some instances the panel data is composed of locations or areas adjacent to each other, so it is possible that the value of variables between regions have interrelated values. The value observed in a location or area \(i\), depending on the observation value related to the nearest location within the period of \(t\), said that this panel data has the effect of spatial dependencies.

Making the assumption that adjacent observation units, example the \(i\) and \(j\) regions indicate the corresponding error values for weighing units, it can be said that the data have spatial heterogeneity effects. The effects of spatial heterogeneity occur as a result of the effect of spatial dependencies of the errors in the model, or called the spatial effects of errors. Panel data have spatial dependency effects and spatial error effects, then the data has a spatial effect. In addressing the spatial effects in panel data, according to Elhorst [5] it takes a special model to overcome the spatial effect in the data, that is the spatial panel model. The spatial panel model used is the spatial Durbin error model.

LeSage [7] introduces the spatial Durbin error model, with the addition of spatial lag effects on independent variables and spatial error effects. Spatial Durbin error model is expressed in the following equation

\[
Y_{it} = \alpha + X_{it} \beta + W^* X_{it} \theta + u_{it}, \\
\quad u_{it} = \lambda W^* u_{it} + \epsilon_{it}.
\]

\(\theta\) is the spatial autocorrelation coefficient, \(Y_{it}\) is the dependent variable of the data in the observation unit to \(i\) and on time to \(t\), \(X_{it}\) is the independent variable of the data in the observation unit to \(i\) and on time to \(t\), \(W^*\) is standardized spatial weighted row matrix, \(\beta\) is coefficient independent variables, \(\alpha\) is intercept, \(\lambda\) is coefficient error spatial, \(u_{it}\) is a spatial error in the region to \(j\) time to \(t\), and \(\epsilon_{it}\) is the model error on observation to \(i\) and time to \(t\).

3. Research Methodology

3.1 Research Data

This study took data of HDI in Central Java Province obtained from BPS of Central Java Province 2010-2016. The variables used in this study consisted of the dependent variable HDI (Y) and the independent variable is life expectancy as the variable \(X_1\), mean years of schooling as a variable \(X_2\), purchasing power parity as a variable \(X_3\), and expected years of schooling as a variable \(X_4\). Data taken from 35 districts or cities in Central Java province as unit cross section, while the 2010-2016 period is an element of time series in the data. Thus the number of independent variables (\(k\)) is 4, cross section (\(N\)) is 35, and time series (\(T\)) is 7.
3.2 Step Research

This study uses software R to estimate parameters of spatial Durbin error model. The first step taken to achieve the objective of this study is to estimate the parameters of the random effects model. After that tested the classical assumption of the best model for HDI data, tested the effect of spatial interaction using Lagrange error multiplier test, tested autocorrelation spatial dependence between locations with Moran $I$ index.

After that, assign the spatial weighted matrix ($W$) with the method queen contiguity and standardize, and estimate the parameter of the spatial Durbin error model with the panel regression model. Having obtained the model, calculating the value of $R^2$ to find out the spatial Durbin error model is well used for HDI.

4. Model Application

4.1 Regression Panel Model

According Baltagi [3], one of the panel regression models that can be used to estimate parameters is a random effect model. The random effect model is a model that uses the approach of generalized least square (GLS) to estimate its parameters.

Based on the calculation result, we get the model of random effect which is expressed as

$$Y_{it} = 3.465 + 0.43801X_{1it} + 1.342X_{2it} + 0.001006X_{3it} + 1.1331X_{4it}$$

Furthermore, the normality, non-autocorrelation, non-multicolinearity, and homocesdasticity assumptions were used to determine random effects model to meet regression assumptions. In non-autocorrelation test, it is found that there is autocorrelation. Based on the tests on each assumption, it can be deduced that the assumptions of normality, non-multicolinearity, and homocesdasticity are met while non-autocorrelation assumptions are not met. Furthermore, the Moran $I$ index test to determine the existence of spatial autocorrelation in each variable model of random effects.

4.2 Moran $I$ Index.

The error obtained from the analysis is done by Moran $I$ index to determine if there is any diversity effect between locations or there is spatial effect in the data. With Moran $I$ index formulated as follows.

$$IM = \frac{n\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})^2}$$

with the value $i, j = 1, ..., 35$. Here is the calculation of Moran $I$ index on each independent variable ($X_k$) presented in Table 1.

| $X_k$  | $IM$    | Identification | Pattern Autocorrelation | Type Autocorrelation |
|--------|---------|----------------|-------------------------|----------------------|
| $X_1$  | 0.006203379 | $IM_1 > I_0$ | Flocking                | Positive             |
| $X_2$  | -0.09346353  | $IM_2 < I_0$ | Spread                  | Negative             |
| $X_3$  | -0.11941419  | $IM_3 < I_0$ | Spread                  | Negative             |
| $X_4$  | -0.10390050  | $IM_4 < I_0$ | Spread                  | Negative             |

The expected value of the Moran $I$ index is expressed as

$$E(IM) = I_0 = \frac{-1}{n - 1} = -0.0294118.$$ Based on Table 1, the value of $IM$ on each variable indicates that there is an autocorrelation on each independent variable so it can be concluded that there is spatial correlation of each independent variable.
4.3. Test Spatial Effect.
According to Anselin [1], the test model procedure for determining the spatial model is derived from the conclusion of the Lagrange multiplier test as its principal reference. In this research, the Lagrange multiplier test used as the reference is the Lagrange error multiplier test. The Lagrange error multiplier test is used to determine the existence of spatial dependencies in the false values of the model or so-called spatial error effects. The hypothesis test for this case is \( H_0: \lambda = 0 \) (no spatial effect in the model error) and \( H_1: \lambda \neq 0 \) (there is a spatial effect in the model error). The critical area for this test is \( \{LM_\lambda | LM_\lambda > \chi^2_{(a,1)}\} \) with \( \chi^2_{(0.05,1)} = 3.841 \). \( H_0 \) is denied if \( LM_\lambda \) critical area. \( LM_\lambda \) is obtained

\[
LM_\lambda = \frac{(\frac{\lambda^T(I - \Phi W)\lambda}{\sigma^2})^2}{\frac{1}{T}XTw} = 27.331
\]

Because of the \( LM_\lambda \) in critical area so \( H_0 \) is rejected which means there is spatial dependency error \( (\lambda \neq 0) \) in the model. It is concluded that spatial Durbin error model can be tested for HDI in Central Java Province.

4.4. Spatial Durbin Error Model.
4.4.1 F-test and T-test For Spatial Durbin Error Model. The spatial Durbin error model is applied by adding spatial lag to predictor variables, life expectancy, expected years of schooling, purchasing power parity, expected years of schooling, and spatial effects of errors. Spatial Durbin error model uses an area approach to estimate its parameters. Therefore, the spatial weighing matrix used is the matrix of queen contiguity which is based on the sideways or corners between locations (LeSage [7]). To determine the effect of all independent variables, spatial lag and spatial effects simultaneously to dependent variables can be done by simultaneous test. Simultaneous test is used \( F \) test, with the hypothesis \( H_0: \) all variables have no significant effect on the dependent variable, and \( H_1: \) at least one independent variables that have a significant effect on the dependent variable. The critical area for this test is \( \{F_{\text{count}}|F_{\text{count}} > F_{(0.05,38,206)}\} = 1,40702 \) with \( H_0 \) refused if value from test \( F_{\text{count}} \) in critical area. Test statistics can be written as

\[
F_{\text{count}} = \frac{R^2/(N+k-1)}{(1-R^2)/(NT-N-k)} = 42,23961
\]

with \( R^2 \) is the coefficient of determination of the spatial Durbin error model.

Value \( F_{\text{count}} \) in critical area, so \( H_0 \) is rejected. This means there is at least one independent variable that significantly affects the dependent variable, so it is necessary to do partial test for each model parameter with \( t \) test. With the hypothesis that \( H_0: \) the coefficient of the parameter has no significant effect on the model, and \( H_1: \) the parameter coefficient significantly affects the model. The critical area for this test is \( \{t_{\text{count}}|t_{\text{count}} < -t(\frac{\sigma}{\sqrt{n}})\} \) or \( \{t_{\text{count}}|t_{\text{count}} > t(\frac{\sigma}{\sqrt{n}})\} \) with \( t(0.05,206) = 1,982264 \) and \( t(0.05,206) = 1,658785 \). \( H_0 \) is rejected if the value of test \( t_{\text{count}} \) in critical area. Partial test values for \( \lambda, \beta, \theta \) and constants are shown in Table 2.

The partial test value for the \( \lambda, \beta, \theta \) is \( t_{\text{count}} \) in critical area so \( H_0 \) is rejected which means the parameter coefficient \( \beta \) and \( \theta \) has a significant effect on the model. Partial test value for the constant parameter is \( t_{\text{count}} \) in critical area so \( H_0 \) is not rejected which means the coefficient of constant parameter has no significant effect to the model. According to LeSage [7], the insignificant variables in the spatial model are referred to as omitted variables, which arise in the spatial model due to unobserved factors in such variables such as location facilities, road accessibility and so forth that can provide Influence on the dependent variable. This variable still has an effect in the resulting model so it is not removed.
Tabel 2. Estimated parameter values $\beta$, $\theta$, constant, and $t_{count}$ spatial Durbin error model.

| Variable | Estimate Value | $t_{count}$ | $t(g_{i(26)})$ | Conclusion |
|----------|----------------|-------------|----------------|------------|
| $constant$ | 3.8664 | 0.7519 | 1.982264 | $H_0$ is not rejected |
| $\lambda$ | -0.33669 | -1.8432 | 1.658785 | $H_0$ is rejected |
| $\beta_2$ | 1.5161 | 27.0285 | 1.982264 | $H_0$ is rejected |
| $\beta_3$ | 0.000861 | 20.0433 | 1.982264 | $H_0$ is rejected |
| $\beta_4$ | 1.1909 | 28.3017 | 1.982264 | $H_0$ is rejected |
| $\theta_1$ | 0.30135 | 4.5702 | 1.982264 | $H_0$ is rejected |
| $\theta_2$ | 0.40747 | 3.8823 | 1.982264 | $H_0$ is rejected |
| $\theta_3$ | 0.00037496 | 4.8507 | 1.982264 | $H_0$ is rejected |
| $\theta_4$ | 0.3994 | 6.5173 | 1.982264 | $H_0$ is rejected |

4.4.2 The Result Spatial Durbin Error Model. From F-test and t-test, all variables have significant effect to dependent variables. The spatial Durbin error model can be expressed as

$$\tilde{Y}_{it} = 3.8664 - 0.2824X_{1it} + 1.5161X_{2it} + 0.000861X_{3it} + 1.1909X_{4it} + 0.30135W^{*}X_{j1it} + 0.40747W^{*}X_{j2it} + 0.00037496W^{*}X_{j3it} + 0.3994W^{*}X_{j4it} + u_{it},$$

(4.1)

$$u_{it} = -0.33669W^{*}u_{it},$$

(4.2)

With value of $R^2$ is 0.8862573. Value $R^2$ approaching value 1, it can be concluded that variation of dependent variable is HDI well explained by model so that regression model is good use. It can be interpreted that 88.62573% of HDI in Central Java in 2010 through 2016 can be explained by all variables.

Based on the spatial Durbin error model for HDI at (4.1) and (4.2), the $\lambda$ coefficient can be interpreted if a region is surrounded by some other region, The surround can be measured at -0.33669 times the average of the surrounding spatial errors. Thus increasing the value of HDI resulting from the influence of errors around the area.

The spatial variable lag on the predictor variable consists of the coefficient of life expectancy ($X_1$), mean years of schooling ($X_2$), purchasing power parity ($X_3$), and expected years of schooling ($X_4$) with weighted spatial weights are positive. The coefficient $\theta_1$, $\theta_2$, $\theta_3$, and $\theta_4$ is interpreted if an area is surrounded by some other territory. Each surround can be measured as 0.30135 times the average of life expectancy in the surrounding area, 0.40747 times the average of mean years of schooling in the surrounding area, 0.00037496 times the average of purchasing power parity in the surrounding area, and 0.3994 times the average of expected years of schooling in the surrounding area.

The variable coefficient of life expectancy is negative. This shows for every one year increase in life expectancy ($X_1$) will decrease HDI $Y$ by 0.2824%. The average coefficient of mean years of schooling, purchasing power parity and expected years of schooling is positive. This shows for every one year mean years of schooling ($X_2$), one unit (one thousand rupiah) purchasing power parity ($X_3$), and one year expected years of schooling ($X_4$), will increase HDI ($Y$) 1.5161%, 0.000861% and 1.1909% respectively.

5. Conclusion
Based on the results and discussion, spatial Durbin error model for HDI in 35 regencies / cities in Central Java Province is stated as follows
Based on the estimation of spatial Durbin error model and partial test, it can be concluded that the significant factors affecting the HDI in Central Java Province are life expectancy, mean years of schooling, purchasing power parity, and expected years of schooling. Calculation result $R^2$ can be interpreted that 88.62573% HDI in Central Java 2010-2016 can be explained by all variables.

Acknowledgments
The authors would like to thank to Universitas Sebelas Maret for providing financial support through Grant of Fundamental Research 2017.

References
[1] Anselin, L., *Spatial Econometrics*, Bruton Center, University of Texas at Dallas, 1999.
[2] Anselin, L., *Spatial Multipliers, and Spatial Econometrics*, International Regional Science Review, University of Illinois, 2003.
[3] Baltagi, B.H., *Ecnometric Analysis of Panel Data, 3rd* ed., John Wiley and Son, Ltd., England, 2005.
[4] BPS, *Human Development Index 2006-2007*, BPS, Jakarta, 2008.
[5] Elhorst, J. P., *Specification and Estimation of Spatial Panel Data Models*, International Regional Science Review, Netherland, 2003.
[6] Hamdani, K., *Analysis Spasial Durbin Fixed Effect Model Panel Data*, Hasanuddin University, 2015.
[7] LeSage, J. P., *Intoduction to Spatial Econometrics*, CRC Press Taylor and Francis Group, Florida, 2009.