Spatial Analysis of The Influence of Education And Income on Indonesia's Democracy Index In 2015 And 2019

Fauzan Zahid Abiduloh¹
Master Program in Population and Employment Economics
University of Indonesia, Indonesia
Fauzanzahid95@gmail.com

Chotib Hasan²
Urban Development Study Program
University of Indonesia, Indonesia
Chotib@hotmail.com

Abstract:
This research discusses the spatial effects of education and income influences on the 2015 and 2019 Democracy Index of Indonesia. The income variable used is Gross Regional Domestic Product (GRDP) per capita, and the education variable used is the average year of schooling. All of the data are collected from the website of the Indonesian Central Bureau of Statistics, namely www.bps.go.id. Using spatial econometrics, researchers found that the distribution of the democratic index value in each province forms a group spatial systemic pattern. Provinces with high democracy index scores tend to be surrounded by provinces that have high democracy index scores, while provinces with low democracy index scores tend to be surrounded by provinces that have low democracy index scores. Researchers also found a spatial dependence on the influence of education and income on the index of democracy in neighboring provinces. Thus, it can be concluded that the quality of democracy in a province is not only caused by the level of education and income in the province, but also by its neighboring provinces.

Keywords: indonesia’s democracy index, education, income, spatial econometrics.

Abstrak:
Penelitian ini membahas pengaruh spasial variabel pendidikan dan pendapatan terhadap Indeks Demokrasi Indonesia 2015 dan 2019. Variabel pendapatan yang digunakan adalah Produk Domestik Regional Bruto (PDRB) perkapita, dan variabel
pendidikan yang digunakan adalah rata-rata lama sekolah. Semua data dikumpulkan dari website Badan Pusat Statistik Indonesia, yaitu www.bps.go.id. Dengan menggunakan ekonometrik spasial, peneliti menemukan bahwa sebaran nilai indeks demokrasi di setiap provinsi membentuk pola sistemik spasial berkelompok. Provinsi dengan skor indeks demokrasi yang tinggi cenderung dikelilingi oleh provinsi yang memiliki skor indeks demokrasi yang tinggi pula, sedangkan provinsi dengan skor indeks demokrasi yang rendah cenderung dikelilingi oleh provinsi yang memiliki skor indeks demokrasi rendah. Peneliti juga menemukan adanya efek spasial dalam pengaruh pendidikan dan pendapatan terhadap indeks demokrasi di provinsi tetangga. Dengan demikian, dapat disimpulkan bahwa kualitas demokrasi di suatu provinsi tidak hanya disebabkan oleh tingkat pendidikan dan pendapatan di provinsi tersebut, tetapi juga oleh provinsi tetangganya.

Kata kunci: indeks demokrasi Indonesia, pendidikan, pendapatan, ekonometrik spasial

Introduction
The writer finds it difficult to find a spatial analysis of democratization. This can be seen from the limited number of studies that use spatial analysis in democratization studies, both at the international level and in Indonesia. This trend is certainly not without reason. There are two reasons why democratization is often regarded as a phenomenon whose forming factors are merely domestic, namely: pedagogical reasons where the study of democratization develops separately from studies of international relations, and another reason is the difficulty of experts to formulate external factors on democratization in a region. (Slender, 2017).

In fact, political behavior, political processes, and events are always in a certain geographical location (Darmofal, 2008). Also, instead of being individual and micro, data on democratization studies as in other political science studies tend to have a collective character. The reasons above then show why many researchers find that various political phenomena, ranging from political communication, mobilization, conflict, and democratization tend to show spatial symptoms. Thus, neglecting space can result in all processes and mechanisms being postulated as if they were occurring in a 'spaceless' realm (Azis, 1994). As a result, it is not uncommon for the results of policy implementation to be decided from the results of the study to deviate too much from the expected hypothesis (Chotib, 2019).

In Indonesia itself, the importance of democratization studies based on spatial or territorial analysis cannot be separated from the large variety of ethnicities, races, cultures, and spiritual beliefs of its inhabitants. Based on population census data conducted by the Indonesian Central Bureau of Statistics in 2010, Indonesia has 1,340 ethnic groups. Moreover, since the reform era, the Government of Indonesia has issued a Regional Autonomy Law which has an impact on changing the bureaucratic system in Indonesia from centralized to decentralized (Haris, 2007). This gives local governments the flexibility to produce more autonomous policies to deal with the unique characteristics of their regions (Haris, 2007), which in turn can affect the quality and condition of democracy in each of these regions. Thus, there is a strong indication that the phenomenon of democratization between regions in Indonesia has a certain pattern that shows its political locality.
At the same time, the quality of democracy in various regions in Indonesia may also be influenced by the socio-economic conditions of the population. Various studies have found how the level of income and education of the population in a region can affect the quality of democracy in that region. Research conducted by Acemoglu et al (2008), Fayad et al (2011), and Cervellati et al (2014) found that the level of income of the population, measured by Gross Domestic Product (GDP) per capita, has a positive and significant effect on the quality of democracy in a country. The same effect was found on the education variable. Research conducted by Kurzman and Lehaer (2004), Glaeser (2007), and Aleman and Kim (2015) found that education, measured by average length of schooling, has a positive and significant effect on the quality of democracy in a country.

According to The Economist Intelligence Unit's report, which examines the level of freedom in 167 countries, the ranking of democracy in Indonesia has dropped 20 places from 48 to 68 (IDI, 2018). After independence, identity politics based on, ethnicity, religion and territory, which Geertz called primordial sentiment, still exists (Nur, 2019). With intolerance and identity politics getting stronger, the future of Indonesian democracy will face serious vulnerabilities and challenges. Identity politics is not only a threat to diversity and democracy, but also a serious threat to a diversity-based Indonesia. Therefore, the pluralistic values, Bhineka Tungga Ika and Pancasila need to be echoed again.

The researcher is interested in conducting a spatial analysis of the quality of democracy in Indonesia because one of these identity politics is based on location, region, and territory. In addition to examine whether the quality of democracy in various regions in Indonesia forms certain spatial patterns, the researcher will also examine whether the quality of democracy in a region is not only influenced by the level of education and income in that area, but also by the level of education and income in other areas adjacent to the region.

**Theoretical Framework**

**Modernization Theory**

The conception that states the relationship between education and income on the quality of democracy in a region is based on modernization theory. Modernization theory is a theory that states the systematic relationship between social, economic and political dimensions (Huber et al, 1993). Each dimension is always in a state of influence, so that a change in direction in one dimension corrects the direction or even the structure of another dimension. In line with the modernization theory, Lipset (1959) argues that high levels of income always have an impact on the desire of the population to protect or even increase the income they have. Residents will always look for ways to increase their income. The desire of the population to protect and increase income will eventually form the collective efforts that underlie the birth of various community organizations. In the end, these community organizations are capable of repelling all state intervention and control that threaten the fulfillment of the utility of society (Lipset, 1959) so as to create a democratic social condition.

Not only the economic aspect in the form of income, the socio-demographic aspect also has an influence on the political aspect. Education as one of the socio-demographic aspects is considered to have an influence on the development of democracy in a region. This is because education has provided individuals with a set of civic skills. Civil skills (civic skills) are intellectual skills characterized by individuals who are broad-minded, effective and responsible, as well as the presence of critical thinking skills (Yuniarto,
2018). Critical thinking skills include the ability to identify, describe, explain, analyze, evaluate, determine and defend opinions with regard to public affairs (Yuniarto, 2018). Thus, an increase in the education level of the population can make the population more active in voicing their political rights and aspirations so as to create a more democratic social condition.

However, the conceptions of the effects of education and income in modernization theory do not say anything about the role of distance and space between regions. In other words, modernization theory has not been able to explain that the socio-demographic, economic and political aspects of one region can influence the socio-demographic, economic and political aspects of another. It was only a few years later that the concept of the spatial effect on democracy began to be explained by adapting Waldo R. Tobler's first law of geography.

Spatial Econometrics in Theoretical Review

Waldo R. Tobler’s (1970) first law of geography states that "everything is related to one another, but something that is close has more influence than something far away" (p. 376). On the basis of this first law of geography, a methodological discipline known as spatial econometrics was born. Spatial econometrics is a part of econometrics that deals with spatial interactions and spatial structures in regression modeling of inter-spatial data (cross-section) (Anselin, 1995). In spatial econometrics, the assumption of defining proximity between regions, or also known as neighboring areas, is usually based on two things, namely vertices or distances (LeSage, 1999). The regional neighbor specification resulting from spatial weighting is very important to base the identification of spatial dependencies on observed data in an area with similar data in nearby areas.

One of the spatial dependencies that can be identified is spatial autocorrelation. Basically, spatial autocorrelation shows a certain systemic pattern in the values observed in adjacent areas (Achjar, 2019). This can be caused by two things. First, there is measurement error when collecting data associated with spatial units such as zip-code, region, country, census guide, etc. (LeSage, 1999). This can happen when the administrative boundaries of the data collection process do not properly reflect the precedents for the characteristics of the data. Second, the spatial dimensions of socio-demographic, economic and other regional characteristics play an important role in the modeling process (LeSage, 1999). Referring to these two reasons, the occurrence of spatial systemic patterns in data in an area with similar data in the nearest area indicates a dependency between those regions.

The influence of neighboring areas on the area of concern really depends on the type of spatial interaction that occurs between variables. If the errors between the dependent variables in the neighboring region are correlated, then the type of spatial dependence on the variables in the neighboring region is called the error spatial dependency (Figure 1). Meanwhile, if the correlation between these errors is added by the interaction between the dependent variable in the observation area with the independent variable in the neighboring area, then the type of spatial dependency is the lag spatial dependency (Figure 1).
Research Methodology

The data in this study are secondary in nature, taken from the website of the Central Bureau of Statistics of the Republic of Indonesia (www.bps.go.id). From this website, various data are obtained including: Indonesian Democracy Index (IDI) data which will be used as the dependent variable of the quality of democracy, data on gross regional domestic product (GRDP) per capita as independent variables of income, and data on average year of schooling as an independent variable of education.

The unit of analysis in this study are all provinces in Indonesia in 2015, amounting to 34 provinces, and all provinces in Indonesia in 2019 which are also 34 provinces. The selection of the unit of analysis at the provincial level is because the IDI data only reaches the provincial level. Meanwhile, the 2015 and 2019 as years of observation were based on the actuality of the data and the similarity of the unit of analysis. This is because before 2015, the number of provinces in Indonesia had not yet been 34 provinces. The results of the analysis from 2015 and 2019 in this study will also be compared to see if there are differences in the spatial effect of the influence of income and education on the quality of democracy in each of the two years.

The research method in this research is quantitative, with descriptive analysis techniques, and inferential. With descriptive analysis, a description of each variable will be displayed, and a summary of its statistics, univariately in the form of tables and graphs. The researcher will visualize the IDI, GRDP per capita, and the average length of schooling data into a thematic map. Researchers will first divide each province into 3 categories based on regions, namely: western, central, and eastern regions. The division of provinces based on the location of each province in the Indonesian time zone. The western of Indonesia, covering every province on the Islands of Sumatra, Java, Province West Kalimantan, and Central Kalimantan. The central region includes the provinces of Bali, West Nusa Tenggara, East Nusa Tenggara, North Kalimantan, East Kalimantan, South Kalimantan, and all provinces in Sulawesi Island. The Eastern Region includes the Provinces of West Papua, Papua, North Maluku and Maluku.
With inferential analysis including: the Moran's Index; Local Indicator of Spatial Association (LISA); and spatial regression calculated using either the spatial lag regression model or the error spatial regression model, Researchers will look at the spatial relationship between the democratic index values in each of the nearby provinces, and see whether education and income in a region have an effect on the quality of democracy in neighboring areas.

**Spatial Weight Matrix**

Any spatial analysis that considers the effect of neighboring areas requires a spatial weight to define the concept of neighborliness in the units of observation. Spatial weighting is generally expressed by binary code in a matrix to express the relationship between one region and another (Anselin, 2014). The general form of the spatial weighted matrix (W) is:

\[
W = \begin{bmatrix}
w_{11} & \cdots & w_{1n} \\
\vdots & \ddots & \vdots \\
w_{n1} & \cdots & w_{nn}
\end{bmatrix}
\]  

(1)

Referring to the matrix above, the spatial weighting matrix is symbolized by W of size n x n where the element \(w_{ij}\) is the spatial weight in the matrix. The weight of \(w_{ij}\) will have a value or 'non-zero' when i and j are neighbors. Meanwhile, \(w_{ij}\) is zero when i and j are not neighbors. The diagonal weighting of the matrix (describing self-neighbors) will be zero or \(w_{ii} = 0\) (Anselin, 2014). This is because a unit of observation i cannot be neighbors with itself.

The definition of proximity is based on two things, namely based on regional boundaries (vertex) or based on distance. In this study, where the observation units are the provinces in Indonesia, the researchers think that the spatial weighting type of distance is more appropriate. This is due to the geographical nature of Indonesia which is an archipelago, where there are provinces that are isolated by the strait or sea from other provinces such as Bangka Belitung, Bali, West Nusa Tenggara, East Nusa Tenggara, Maluku and North Maluku. Therefore, the spatial weighting in this study uses the distance type spatial weighting with a threshold of 434,3789 miles which is obtained from the measurement of the minimum distance between two adjacent provinces.

**Moran’s Index**

Basically, the calculation of the Moran’s Index value is used to see the global spatial autocorrelation of the values to be observed in various regions (Achjar, 2019). Moran's index value can be calculated using the formula (Achjar, 2019):

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

(2)

Notes:

| I    | Moran's Index |
| n    | Number of spatial units |
| xi   | The value of the spatial unit variable i |
| xj   | The value of the spatial unit variable j |
| wij  | Spatial weighted matrix element |

Using the formula above, if \(I > 0\) then the autocorrelation value is positive, which indicates that the data forms groups (Achjar, 2019). If \(I = 0\) then there is no spatial
autocorrelation or the relationship between observation values in adjacent areas is random (Achjar, 2019). Meanwhile, if I < 0, the autocorrelation value is negative, which indicates that the pattern of observation values in the adjacent area is dispersed (Achjar, 2019).

To find out the significance of the Moran's Index value, it is necessary to test the significance by calculating the Z statistical value of the Moran's Index value and comparing it to the specified p-value. The hypothesis in testing the significance of the Moran's Index value is as follows:

H₀: There is no global spatial autocorrelation
H₁: There is global spatial autocorrelation

The test statistic of the significance of the Moran's Index value is (Hamdan, 2019):

\[ Z_{\text{count}} = \frac{1 - E(I)}{\sqrt{\text{Var}(I)}} \] (4)

\[ E(I) = \frac{1}{n(n - 1)} \sum_{i \neq j} W_{ij} z_i z_j \] (5)

\[ \text{Var}(I) = E[I^2] - E[I]^2 \] (6)

Notes:
E (I) : Expectation value of Moran's Index
Var (I) : Variance of Moran's Index
n : Number of spatial units

With a significance level of a = 5%, the decision criteria for the hypothesis are "Reject H₀ if | Zcount | > Z1-a".

**Local Indicator of Spatial Association (LISA) Index**

The Moran index calculation can be supplemented with the LISA index. The LISA index is basically used to identify the presence of local spatial patterns. In other words, the LISA index can identify the presence of spatial clusters of values observed in adjacent areas. The LISA index is expressed in the following equation (Lee and Wong, 2001):

\[ I_i = z_i \sum_{j=1}^{n} W_{ij} z_j \] (7)

where \[ z_i = \frac{(x_i - \overline{x})}{\sigma_x} \] (8)

\[ z_j = \frac{(x_j - \overline{x})}{\sigma_x} \] (9)

Testing the LISA Index value can be done with the following hypothesis:

H₀: There is no local spatial autocorrelation
H₁: There is local spatial autocorrelation

With the test statistic:

\[ Z_{\text{count}} = \frac{1 - E(I)}{\sqrt{\text{Var}(I)}} \] (10)

With a significance level of a = 5%, the decision criteria for the hypothesis are "Reject H₀ if | Zcount | > Z1-a".

By performing a statistical test on the LISA index value, it can be seen that there are four clusters of the observed value in adjacent areas. The four clusters are none other
than (Hamdan, 2019): hot spots or high-high spatial relationship (HH), which is a spatial relationship where an area has a high value and is surrounded by other areas that have a high value as well; outlier 1 or Low-High spatial relationship (LH), which is a spatial relationship where an area that has a low value is surrounded by other areas that have a high value; cold spot or Low-Low spatial relationship (LL), which is a spatial relationship where an area with a low value is surrounded by other areas with a low value; outlier 2 or high-low spatial relationship (HL), which is a spatial relationship where an area that has a high value is surrounded by other areas that have a low value.

**Lagrange Multiplier (LM) Test**

The Moran and LISA index only show a spatial systemic pattern in the values observed in an area. To test and study the spatial effect of independent variables on the dependent variable, it is necessary to do a spatial dependency test, namely the Lagrange Multiplier (LM) test. In a model that is considered to have spatial dependencies, there are two possible types of spatial dependencies, namely: spatial error and spatial lag (Chotib, 2019). A model is said to have a spatial dependency type of spatial error when there is a linear relationship between errors or no homoscedasticity. Meanwhile, a model is said to have a spatial dependency type of spatial lag when there is a linear relationship between errors, and the dependence between observed values is added in an area with the same value in the nearest area. This is because there is mutual influence between the independent variables and the dependent variable in adjacent areas.

Considering the above ideas, the LM test is very important to identify what kind of spatial dependencies exist on the model. The hypothesis testing for LM-Lag are (Hamdan, 2019):

\[ H_0 : \rho = 0 \) (there is no interaction and error correlation between regions)
\[ H_1 : \rho \neq 0 \) (there are interactions and correlation errors between regions)

The test statistic for the spatial lag model is as follows (Hamdan, 2019):

\[ LM_{\text{Lag}} = \left[ \frac{N e^2 W y}{e^2} \right] \left[ \frac{N (X'X) + M (X'X)}{e^2} \right] + \text{tr} \left[ (W + W^\prime) W \right]^{-1} \]

Keterangan:
N : Number of observation
W_y : Spatial lag for weighted matrix W (N x N)
M : \( I - X' (X'X)^{-1} X' \)
\( \beta \) : OLS estimation parameter

Meanwhile, for testing the spatial error model, the hypothesis are (Hamdan, 2019):

\[ H_0 : \lambda = 0 \) (there is no error correlation between regions)
\[ H_1 : \lambda \neq 0 \) (there are correlation errors between regions)

The test statistic for the spatial error model is (Hamdan, 2019):

\[ LM_{\text{Error}} = \left[ \frac{N e^2 W e}{e^2} \right]^2 \left[ \text{tr} \left[ (W + W^\prime) W \right] \right]^{-1} \]

Where \( W_e \) is weighting matrix for spatial errors. Both the LM-Lag and LM-Error tests follow the Chi-square distribution with a degree of freedom of one. \( H_0 \) is rejected if the LM statistical value is \( \geq \) the critical value of the chi-square statistic.
If each of the LM-Lag and LM-Error shows significant results, it is necessary to conduct the Robust LM test to determine which of the spatial lag models or spatial error models will be used. The formulation of the Robust LM-Lag test is as follows (Hamdan, 2019):

$$LM_{Lag} = \frac{(\frac{\sum W_y - \sum W e}{\sum e^2})^2}{(NJ_{\rho\beta})^{-1}} - t$$  \hspace{1cm} (13)

where,

$$S^2 = \frac{\sum e^2}{N}$$  \hspace{1cm} (14)

$$(NJ_{\rho\beta})^{-1} = \left[ t + \left( \frac{WX\beta}{S^2} \right) M(WX\beta) \right]^{-1}$$  \hspace{1cm} (15)

$$M = I - X' (X'X)^{-1} X'$$  \hspace{1cm} (16)

$$t = tr \left[ (W + W') W \right]$$  \hspace{1cm} (17)

The formulation of Robust LM-Error test is as follow (Hamdan, 2019):

$$LM_{Error} = \frac{\left[ \frac{\sum W e}{S} - t \right] - (NJ_{\rho\beta})^{-1}}{t - t^2(NJ_{\rho\beta})^{-1}}$$  \hspace{1cm} (18)

Like the LM-Lag and LM-Error tests, the Robust LM-Lag and Robust LM-Error tests also follow a chi-square distribution with degrees of freedom of one. Thus, the null hypothesis is rejected when the Robust LM statistical result is $\geq$ the critical value of the chi-square statistic.

**General Model of Lag and Error Spatial Regression Model**

The results of the LM and Robust LM tests actually show what the most appropriate spatial regression model to be used. This is because the spatial regression model is divided into two types, namely the spatial lag regression model and the error spatial regression model. The spatial lag regression model is appropriate to regress models that have a lag type spatial dependency. Meanwhile, the error spatial regression model is appropriate to regress models that have a spatial dependency of the error type. The general model of spatial lag regression model is (Hamdan, 2019):

$$y = a + \rho W y + X_i \beta + u$$  \hspace{1cm} (19)

Notes:

- $a$: Constant
- $y$: Dependent variable
- $X_i$: Independent variable $i$
- $\beta$: Coefficient vector of independent variable $(k \times 1)$
- $\rho$: Coefficient parameter of spatial lag dependent variable
- $u$: Error term

Spatial error model, or also often called the Spatial Moving Average, is a model that considers the interaction of correlation errors between regions. The Spatial Error Model can be formulated as follows (Hamdan, 2019):

$$y = a + X_i \beta + \epsilon$$  \hspace{1cm} (20)

where,

$$\epsilon = \lambda W \epsilon + u$$  \hspace{1cm} (21)
Notes:
\( a \) : Constant
\( y \) : Dependent variable
\( X_i \) : Independent variable \( i \)
\( \beta \) : Coefficient vector of independent variable (\( k \times 1 \))
\( \lambda \) : Coefficient of moving average which accommodate spatial error
\( \varepsilon \) : Vector of error term which is not explained in the model
\( \varepsilon \sim N(0, \sigma^2I) \)
\( u \) : Other error term

The Diagnostic Framework of Spatial Regression Model

The sequencing of the formulas above actually forms a diagnostic framework for choosing the spatial model. The specification of determining the model can be described in a graph as follows (Anselin dan Rey, 2014):

![Figure 2. The Diagnostic Framework of Spatial Regression Model](image)

The step-by-step of the spatial regression model specification framework above, can be sorted as follows (Hamdan, 2019):
1. Estimating the initial model using the Ordinary Least Square (OLS).
2. Perform a significance test of the Lagrange Multiplier (LM) Error and Lag. If the p-value of LM-Error is smaller than alpha, the spatial error model is more appropriate. If the p-value of LM-Lag is smaller than alpha, then the Lag spatial model is more appropriate. If the two p-values are not significant, the OLS model is more appropriate. However, if both p-values are significant, then do a further test using the Robust LM test.
3. If the p-value of Robust LM Error is significant, then the spatial error model is more appropriate. If the p-value of Robust LM (Lag) is significant, then the spatial lag model is more appropriate. If both are equally significant, then choose the most significant value.

However, the diagnostic framework above does not explain the steps to be taken if each robust test is not significant. To determine the steps if this happens, it is necessary to review the theoretical basis used. If there is no theoretical explanation that can be considered, then a decision can be made through the most significant value of the robust LM test, or by considering the degree of the goodness of fit of each model.

**Research Model**

Referring to Figure 2, every spatial model always begins with the formation of an OLS model first. The OLS model in this study is as follows:

\[
IDI = \theta_0 + \theta_1 \text{LnGRDPC} + \theta_2 \text{AYS} + \varepsilon
\]  

(22)

**Notes:**
- \(IDI\) : Indonesian Democracy Index
- \(\text{LnGRDPC}\) : Log natural of Gross Regional Domestic Product per Capita
- \(\text{AYS}\) : Average Years of Schooling
- \(\theta\) : Constant
- \(\varepsilon\) : Vector of error term which is not explained in the model \(\varepsilon \sim N (0, \sigma^2I)\)

The OLS model will later be tested by the LM test to see if the model has any spatial dependencies. If only the LM lag test is significant, then a spatial regression is performed using the spatial lag model. The spatial lag regression model in this study can be formulated as follows:

\[
IDI = \beta_0 + \rho \text{WIDI} + \beta_1 \text{LnGRDPC} + \beta_2 \text{AYS} + u
\]  

(23)

**Notes:**
- \(\beta_0\) : Constant
- \(\rho\) : Coefficient parameter if IDI spatial lag
- \(\text{W}\) : Spatial weight matrix
- \(u\) : Error term

If only the LM test error is significant, then a spatial regression is performed using the spatial error model. The spatial error regression model in this study can be formulated as follows:

\[
IDI = a_0 + \beta_1 \text{LnGRDPC} + \beta_2 \text{AYS} + \varepsilon
\]  

where,

\[
\varepsilon = \lambda \text{W} \varepsilon + u
\]  

(24)

(25)

**Notes:**
- \(a_0\) : Constant
- \(\lambda\) : The moving average coefficient which accommodate spatial error

If the LM lag test and LM error test are both significant, then the Robust LM lag test and the Robust LM error test are performed. If the Robust LM lag test value is more significant, then perform spatial regression using the spatial lag model. If the Robust LM
error test value is more significant then perform spatial regression using the spatial error model. If both are still equally significant or insignificant, then a spatial regression model can be used according to the Robust LM test with the smallest p-value.

**Measures of Fit in Spatial Model**

In order to measure the fit or the goodness of fit of the spatial model, there are four indicators that can be used: the pseudo $R^2$, the maximized log-likelihood, the Akaike information criterion (AIC), and the Schwartz Criterion (Putra et al, 2020). The pseudo $R^2$ here is different from $R^2$ in the OLS model, which is calculated based on the decomposition of the total sum of squares into residual sums of squares (Putra et al, 2020). The pseudo $R^2$ is the ratio of the variance of the predicted values to the variance of the observed values of the dependent variable (Putra et al, 2020). The higher the value of the pseudo $R^2$, the better the model. The pseudo $R^2$ can be formulated as follows (Itaoka, 2012):

$$R^2 = 1 - \frac{SSR}{SST}$$

where,

$$SSR = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

and,

$$SST = \sum (y_i - \bar{y})^2$$

Notes:

- $SSR$ : Residual sum of squares
- $SST$ : Total sum of squares

For the reason above, the comparison between OLS model and spatial regression model here should not be taken seriously. Indeed, the researcher will still mention the $R^2$ value of the OLS regression model and the spatial regression model, and add AIC. The lower the AIC of the model, the better the model. AIC can be formulated as follows (Snipes and Taylor, 2014):

$$AIC = 2K - 2 \log (L (\hat{\theta} | y))$$

Notes:

- $K$ : Number of estimable parameters
- $\log (L (\hat{\theta} | y))$ : Log-likelihood at its maximum point of the model estimated

**Result and Discussion**

**Descriptive Analysis of IDI, Income, and Education in 2015 and 2019**

Based on the results of IDI data collected from Central Bureau of Statistics, the researchers found that the quality of democracy in Indonesia has increased from 72.82 in 2015 to 74.92 in 2019 nationally. The increase in the value of IDI nationally was in line with the drastic increase in aspects of democratic institutions, from 66.87 in 2015 to 78.73 in 2019. The value of Political Rights also increased from 70.63 in 2015 to 70.71 in 2019. However, the value of the aspects of Civil Liberties decreased from 80.30 in 2015 to 77.20.

**Figure 3.**

IDI Value and Its Aspects Nationally in 2015 and 2019
Regionally, the quality of democracy in the western and central regions of Indonesia is relatively much better than the quality of democracy in the eastern region. The average IDI score for provinces in the western region of Indonesia was 72.73 in 2015 and increased to 76.19 in 2019. The average IDI score for provinces in the central region of Indonesia was 74.83 in 2015 and is increasing to 76.96 in 2019. This also shows that the central region of Indonesia is the most democratic region compared to the other two regions. Meanwhile, the average IDI score for provinces in eastern Indonesia was only 61.24 in 2015 and increased to only 65.25 in 2019. This shows that the people in eastern Indonesia is the most undemocratic population in Indonesia.

**Figure 4.**
Province IDI 2015 and 2019 by Region

Although the population in the Central region of Indonesia is the most democratic population, relative to the population in the West and East regions, the average years of schooling for residents in the Central region of Indonesia is not the highest. Based on Figure 5, it is known that as you go east, the average length of schooling for the population is decreasing, both in 2015 and 2019.
Figure 5.
Average Years of Schooling in 2015 and 2019 By Region

The same trend is also shown by the average GRDP per capita in each province in each region. Based on Figure 6, it is known that as we go further to the east, the average GRDP per capita has decreased, both in 2015 and in 2019. This shows that the average achievement of the highest IDI score in the Central region of Indonesia may not be caused by these two averages, average years of schooling nor GRDP per capita. In other words, there are other factors that are not included in this research model.

Figure 6.
GRDP per Capita in 2015 and 2019 By Region (in thousand rupiah)

Based on Table 1, it is known that the population (N) of this study amounted to 34. The minimum, maximum, and average values of each variable from 2015 to 2019 also increased. According to the standard deviation of each variable, it is known that the dependent variable IDI2015 and IDI2019 have the highest standard deviation, while the independent variables only have a standard deviation < 1. This shows that the distribution of IDI values in each observation unit tends to be varied, while the distribution the LnGRDPC and AYS values tend to be concentrated in their average values.
Table 1.
Statistical Summary of Dependent and Independent Variables

| Variable     | N  | Minimum | Maximum | Mean  | Standard Deviation |
|--------------|----|---------|---------|-------|--------------------|
| IDI2015      | 34 | 57,55   | 85,32   | 72,12 | 6,753              |
| IDI2019      | 34 | 57,62   | 88,29   | 75,17 | 6,31               |
| LnGRDPC2015  | 34 | 9,31    | 11,87   | 10,36 | 0,57               |
| LnGRDPC2019  | 34 | 9,45    | 12,07   | 10,49 | 0,54               |
| AYS2015      | 34 | 6,59    | 11,06   | 8,02  | 0,96               |
| AYS2019      | 34 | 6,65    | 11,06   | 8,54  | 0,93               |

Source: Researchers, 2019

Global Spatial Autocorrelation of IDI Year 2015 and 2019

Based on Figure 4, there is an indication that the IDI values that are scattered in each province tend to form a spatial group pattern. Provinces in the West and Central regions tend to have high IDI scores, while the IDI scores in the East tend to be low. To prove this indication, it is necessary to do a calculation of the Moran’s Index.

Using the formula for calculating the Moran's Index and it’s testing hypothesis in the methodology sub-chapter, it is known that the calculated Moran’s Index values of IDI are 0.3065 in 2015, and 0.3586 in 2019 which are both significant at $\alpha = 5\%$. This shows that there is a group spatial pattern in the achievement of IDI scores between provinces in Indonesia, both in 2015 and 2019. The finding of the spatial group pattern shows that provinces with good democratic quality tend to be close to provinces that have good democratic qualities as well, and vice versa.

In 2015, provinces with relatively high Democracy Index scores tended to be concentrated in the central region of Indonesia, covering almost all of Java and Kalimantan (Figure 7). Meanwhile, provinces in eastern Indonesia tend to have relatively low Democracy Index scores, covering all of Papua and Maluku (Figure 7).

Figure 7.
The Thematic Map of the Democracy Index Value of 34 Provinces of Indonesia in 2015
Something similar happened in 2019 (Figure 2). This is in line with the value of the Moran IDI Index in 2019 which, although increasing, does not show a significant difference from the value of the Moran IDI Index in 2015. Provinces with relatively high Democracy Index scores are still concentrated in the central region of Indonesia, as are the provinces with a relatively low value for the Democracy Index, it is still concentrated in the East.

Figure 8.
Thematic Map of the Democracy Index Value of 34 Provinces of Indonesia in 2019

Local Spatial Autocorrelation of IDI Year 2015 and 2019

The results of the calculation of the Moran Index value only show the spatial pattern of the Democracy Index for each province in Indonesia as a whole. Therefore, it is necessary to calculate the LISA Index value to determine the existence of spatial clusters in the achievement of the Democracy Index value for each province in 2015 and 2019.

After calculating the LISA Index value and its testing hypothesis, which the formula has been mentioned in the methodology sub-chapter, it is known that in 2015 there were only five provinces that had a significant LISA Index value with a p-value < 0.05, namely East Java Province, Central Kalimantan Province, West Papua Province, Papua Province, and Maluku Province. From the results of calculating the LISA Index value, a thematic map can be visualized to show the local spatial patterns as follows:
Based on Figure 9, it is known that there are High-High spatial clusters in the neighboring areas of Central Kalimantan and East Java Provinces. Meanwhile, it is known that there are Low-Low spatial clusters in neighboring areas of Papua Province, West Papua Province, and Maluku Province. The finding of High-High and Low-Low local spatial clusters are in line with what Figure 3 shows, where provinces with high Democracy Index scores tend to be concentrated in the central region of Indonesia, and provinces with low Democracy Index scores are concentrated in the eastern region of Indonesia.

Figure 10.
LISA Cluster Map of the Democracy Index Value of 34 Provinces of Indonesia in 2019
Something similar happened in 2019 (Figure 10). There are local spatial clusters of High-High Democracy Index in the neighboring regions of Central Kalimantan, West Kalimantan, South Kalimantan, and East Java Provinces. There are also local spatial clusters of Low-Low Democracy Index in the provinces of Papua, West Papua, Maluku, and North Maluku. However, there is a dispersing pattern of Low-High in the neighboring areas of West Java Province. This shows that West Java Province which has a relatively low quality of democracy tends to be close to provinces that have a relatively high quality of democracy.

The existence of a dispersed local spatial pattern in the neighboring areas of West Java Province can be caused by the value of the Civil Freedom Aspect, as one of the three aspects of the IDI composite, dropping drastically from 79.10 in 2015 to 65.16 in 2019. Based on data collected and processed by the Setara Institute in 2015, West Java Province is considered the most intolerant province, especially in terms of inter-religious harmony. Of the 10 most intolerant cities in Indonesia in 2015, 6 of them are located in West Java, such as Bogor, Bekasi, Depok, Bandung, Sukabumi, and Tasikmalaya. This explains why West Java Province actually is in inverse with relation to its neighboring provinces.

Spatial Dependence in OLS Regression Model of IDI 2015 and 2019
The results of the Moran and LISA calculation indexes only have a function to identify any spatial systemic patterns in the distribution of IDI values in each province. In other words, the results of these calculations do not show any information regarding the existence of spatial dependence on the quality of democracy in Indonesia. To test for this spatial dependence, it is necessary to carry out the LM test which formula has been applied in the sub-chapter of the methodology.

| Spatial Dependence Test       | 2015       | 2019       |
|-------------------------------|------------|------------|
| Moran’s I                     | 4.0092***  | 4.2219***  |
| LM Lag                        | 9.3747***  | 11.2709*** |
| LM Error                      | 9.9474***  | 11.3130*** |
| Robust LM Lag                 | 0.0002     | 0.0601     |
| Robust LM Error               | 0.5729     | 0.1022     |

*p-value<0,10; **p-value<0,05; ***p-value<0,01

Based on Table 2, it is known that in the OLS model, which variables have been mentioned in the methodology section, there is a spatial dependency. This is indicated by the significant value of Moran’s I in both 2015 and 2019. Based on the results of the LM test, it is known that the results of the LM Lag and LM Error tests are also significant, both in 2015 and 2019, so it is necessary to carry out the Robust LM Lag test and Error. However, both the Robust LM Lag and Error test results do not show significant results. Referring to the spatial model diagnostic framework (Figure 2), the model selection in this study is based on the most significant Robust LM test results, namely the Robust LM
Error test results. Thus, the most appropriate spatial regression model used in this study is the error spatial regression model.

Variable Coefficients of OLS, Spatial Lag, and Spatial Error Regression Model

Although the spatial error regression model is the most appropriate spatial regression model to be used in this study, researchers will also provide the result of lag spatial regression model as a comparison. The coefficient of each variable from the results of the OLS regression model, spatial regression, and spatial error regression in this study are as follow:

Table 3. Variable Coefficients According to the Regression Models Used in 2015 and 2019

| Variable | OLS  | Spatial Lag | Spatial Error |
|----------|------|-------------|---------------|
|          | 2015 | 2019        | 2015          | 2019          | 2015          | 2019          |
| WIDI     | -    | -           | 0.528***      | 0.487**       | -             | -             |
| Constant | 45,735** | 52,277**    | 10,775        | 23,107        | 51,716***     | 67,579***     |
| LnGRDPC  | 0.644 | 0.778       | 0.391         | 0.071         | -0.023        | -1,076        |
| AYS      | 2,458* | 1,724       | 2,404**       | 1,749         | 2,592**       | 2,259**       |
| LAMBDA   | -    | -           | -             | -             | 0.544***      | 0.546***      |

*p-value<0.10; **p-value<0.05; ***p-value<0.01

Based on Table 3, it is known that the independent variable average years of schooling has a positive and significant influence in the OLS model, and the 2015 spatial lag regression model. That variable also has a positive and significant influence in the spatial error regression model in 2015 and 2019. Meanwhile, the independent variable income in the form of LnGRDPC is not significant in each model.

WIDI in Table 3 is a spatially independent variable in the spatial lag regression model, which is the average effect of neighboring provinces captured in the model. The spatial independent variable in the form of WIDI in the spatial lag regression model has a positive and very significant effect with a coefficient value of 0.528 in 2015 and 0.487 in 2019.

LAMBDA in Table 3 shows the average effect of neighboring units on the residuals of the model. In other words, LAMBDA shows the influence of neighboring areas that have not been captured in the model so that when LAMBDA is displayed in the spatial regression model error, it will correct the coefficients of each independent variable in the model.

LAMBDA in the error spatial regression model has corrected the value of the coefficient of the independent variable on the average years of schooling from only 2.458 in the OLS model to 2.592 in the error spatial regression model in 2015, and from 1.724 which is not even significant in the OLS model to be significant with the coefficient value 2.259 in the spatial error regression model in 2019. Thus, it can be seen that the influence of the average years of schooling on IDI in nearby provinces, in the residual model of spatial error regression model, has increased the effect of the average years of schooling on IDI in provinces that are in attention.
The Goodness of Fit of OLS, Spatial Lag, and Spatial Error Regression Model

Table 4. The Goodness of Fit of OLS, Spatial Lag, and Spatial Error Regression Model

| Variable | OLS  | Spatial Lag | Spatial Error |
|----------|------|-------------|--------------|
|          | 2015 | 2019        | 2015         | 2019         |
| R-Squared| 0.142| 0.085       | 0.343        | 0.289        |
| AIC      | 226,126| 223,670       | 221,083      | 218,745      |

The spatial error regression model is more efficient in explaining the dependent variable IDI2015 and IDI2019. This is known based on Table 4, which shows that the R-Squared value of the spatial error regression model is greater than the other two models. In addition, the AIC values of the error spatial regression model are also smaller than the other two models, which also shows that the spatial error regression model can better explain the dependent variables IDI2015 and IDI2019 than the OLS model and the spatial lag regression model.

The efficiency of the spatial error regression model can be shown by predicting the IDI value based on the coefficient value of each variable in each model. For example, it can be done predicting the achievement of the IDI value of Central Sulawesi Province in 2015 using the spatial error regression model as follows:

\[51,716 + 10,27(-0.023) + 7.97(2.592) = 72.14\] (26)

The prediction of IDI value in South Sulawesi Province in 2015 using the spatial lag regression model is as follows:

\[10,775 + 76.67(0.528) + 10.27(0.391) + 7.97(2.404) = 71.93\] (27)

Meanwhile, the prediction of IDI value in South Sulawesi Province in 2015 using the OLS model is as follows:

\[45,735 + 10.27(0.644) + 7.97(2.458) = 71.76\] (28)

Referring to the real IDI value of the 2015 of South Sulawesi, which was 76.67, it can be seen that the predicted value from the spatial error regression model is more accurate than the spatial lag regression model and the OLS model.

Conclusion

Based on the results of the analysis in this research, it can be concluded that the quality of democracy in various provinces in Indonesia in 2015 and 2019 forms a certain spatial systemic pattern, which indicates clusters of democratization in adjacent areas. In addition, as shown by the LM-test, the quality of democracy in a province in Indonesia is also influenced by the quality of democracy in its nearby provinces. Thus, every effort to develop democracy in regional level, would always in need of consolidation and coordination between regional governments, especially in neighboring regions.

Suggestion

The number of observation units is relatively small because the number of
observation units in this study was 34 due to IDI data only reached the provincial level. The impact is, the socio-economic-political level of the district/city is not captured. For further research, it is expected to be able to explain the Indonesian Democracy Index at district / city level in order to see more credible analysis results. This study only uses modernization theory which states that there is a relationship between education and income on the quality of democracy in a region. For further research, it is expected to be able to use other relevant theories. There are more influential factors outside the model used because in the results of this study the LM-error is more significant.

About Authors

Fauzan Zahid Abiduloh is a final semester student at the Master in Economics of Population and Labor (MEKK), Universitas Indonesia. He is also actively working as a State Civil Servant Government Employee at the Directorate General of Population and Civil Registration of the Ministry of Home Affair.

Chotib Hasan is the Head of the Urban Studies Program at the School of Strategic and Global Studies, Universitas Indonesia. As an alumni of the Faculty of Geography at Universitas Indonesia, he is also active in teaching in various faculties and programs at the University of Indonesia, and has expertise in spatial economics.

Acknowledgements

The writings and research that underlie it cannot be possible without the enormous help of colleagues, partners, and also families who always motivating us to the completion of this research.

References

Acemoglu, D., et al. (2008). Income and Democracy. American Economic Review 2008, 98:3, pp. 808-842.

Achjar, Nuzul. (2019). Analisis Pola dan Konsentrasi Spasial Sektor Industri Manufaktur di Provinsi Banten 2010-2018. Dalam “Rahatiningtyas, S.D & Hasilah, F, editor (2019). Aplikasi Spasial untuk Pembangunan. Departemen Geografi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia”.

Aleman, E., & Kim, Y. (2015). The Democratizing Effect of Education. Research and Politics, Oktober-Desember 2015: 1-7., DOI: 10.1177/2053168015613360. SAGE.

Anselin, L. (1995). Local Indicator of Spatial Association-LISA. Geographical Analysis 27, issue. 2. Pp. 93-184.

Anselin, L., & Rey, S.J. (2014). Modern Spatial Econometrics in Practice: A Guide to GeoDaSpace and PySAL. GeoDa Press LLC.

Azis, I. (1994). Ilmu Ekonomi Regional dan Beberapa Aplikasinya di Indonesia. Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia.

Cervellati, M., et al. (2014). Income and Democracy: Comment. The American Economic Review.

Chotib, H. (2019). Spatial Lag dan Spatial Error: Solusi Masalah Heteroskedastisitas dalam Pemodelan Spasial. Dalam Rahatiningtyas, S.D & Hasilah, F, editor (2019).
Aplikasi Spasial untuk Pembangunan. Departemen Geografi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

Darmofal, D. (2008). Bayesian Spatial Survival Models for Political Event Processes. American Journal of Political Science., Vol. 53., Issue 1., pp: 241-257. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2008.00368.x

Fayad, G., et al. (2011). Income and Democracy: Lipset’s Law Inverted. Department of Economics OxCarre (Oxford Centre for The Analysis of Resource Rich Economies): University of Oxford.

Glaeser, L.E., et al. (2007). Why Does Democracy Need Education? J Econ Growth, Vol. 12., pp. 77-99., DOI 10.1007/s10887-007-9015-1. Springer.

Hamdan, O.F. (2019). Rasio Tenaga Pendidik, Rasio Tenaga Kesehatan, dan Capaian Pembangunan Manusia dalam Analisis Spasial. Depok: Tesis Program Magister Ekonomi dan Kependudukan Fakultas Ekonomi dan Bisnis Universitas Indonesia.

Haris, S. (2007). Desentralisasi dan Otonomi Daerah. Jakarta: LIPPI Press.

Huber, E., Rueschemeyer, D., and Stephens, J. (1993). The Impact of Economic Development on Democracy. Journal of Economic Perspectives 7, no. 3, pp. 71-85.

Itaoka, K. (2012). Regression and Interpretation Low R-Squared!. Social Research Network 3nd Meeting Noosa., April 12-13, 2012.

Kurzman, C., & Leaher, E. (2004). Intellectuals and Democratization. The American Journal of Sociology., Vol. 10., No. 9., pp. 937-986.

Lee, J., & Wong, D.W.S. (2001). Statistical Analysis with Arcview GIS. New York: John Wiley and Sons.

LeSage, J.P. (1999). The Theory and Practice of Spatial Econometrics. ResearchGate.

Lipset, M.S. (1959). Some Social Requisites of Democracy: Economic Development and Political Legitimacy. American Political Science.

Nur, Syurya Muhammad. (2019). Demokrasidan Tantangannya dalam Bingkai Pluralisme di Indonesia. Jurnal Ilmiah Mimbar Demokrasi Volume 19 No. 1.,

Putra, S.A., Tong, G., & Pribadi, O.D. (2020). Spatial Analysis of Socio-Economic Driving Factors of Food Expenditure Variation between Provinces in Indonesia. Sustainability, 2020: 12, 1638.

Slender, D. (2017). Building Democracy: National and International Factors. IntechOpen., pp. 175-194., http://dx.doi.org/10.5772/intechopen.71984

Snipes, M., & Taylor, D.C. (2014). Model Selection and Akaike Information Criteria: An Example from Wine Ratings and Prices. Wine Economics and Policy.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46, pp. 234-240.

Yunarto, B. (2018). Pendidikan Demokrasi dan Budaya Demokrasi Konstitusional. Sleman: Deepbulish.