JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 1, January 2024, pp. 1-14

Panel Data Spatial Regression Modeling with a Rook Contiguity Weighting Function on the Human Development Index in West Sumatera Province

Prizka Rismawati Arum[1], Lisa Anggraini[2], Indah Manfaati Nur[3], Eko Andy Purnomo[4]
[1,2,3,4]Departement of Statistics Faculty of Mathematics and Natural Science, Universitas Muhammadiyah Semarang, Indonesia
prizka.rismawatiarum@unimus.ac.id[1]

Article History:
Received : 20-07-2023
Revised : 07-11-2023
Accepted : 27-11-2023
Online : 19-01-2024

Keywords:
HDI;
Rook Contiguity;
Random Effect Spatial Autoregrresive.

ABSTRACT

The achievement of the level of welfare of a region or country can be seen from the level of human development as measured by the Human Development Index (HDI). West Sumatra is one of the provinces with HDI achievements above the national average. However, there are still regencies/cities in West Sumatra Province that have HDI achievements below the national average and HDI achievements in West Sumatra Province Regencies/Cities have changed in 2017-2021. Therefore, in this study, spatial analysis of panel data was used. The aim of this research is to find out the general description of the HDI of West Sumatra Province, obtain a panel data spatial regression model and obtain variables that significantly influence on HDI in West Sumatra Province 2017—2021. The model formed from this analysis using the rook contigutiy weighting function is Random Effect Spatial Autoregressive because the spatial interactions formed in human development index data in West Sumatra Province are real at lag. This model is a suitable model based on panel spatial model selection and has an R^2 value of 92.94%. Analysis of human development index data in regencies/cities in West Sumatra Province using spatial regression panel data obtained results that expectations of school length (X_1), average length of schooling (X_2), and population density (X_3) significantly directly influenced the human development index in regencies/cities in West Sumatra Province.

https://doi.org/10.31764/jtam.v8i1.16675

———————————— ◆ ————————————

A. INTRODUCTION

Success in national development is determined in a variety of factors. One of the benchmarks for the success of national development is the achievement of the quality of human life as measured through the Human Development Index (HDI) (Masruroh & Subekti, 2016). HDI describes how people can gain access to development outcomes such as health, education, income and other aspects of life. Capital to carry out the development process, the government as the organizer certainly requires quality human resources (Rosyadah, 2021). According to the Central Statistics Agency (2021), this potential has not been fully able to contribute to human development. Uneven development and the vast territory of Indonesia are part of the factors that have an impact on inequalities both between provinces and between districts/cities (Siswoyo, 2014).

HDI in Indonesia in 2021 experienced growth of 0.49 percent, higher than the previous year which only grew 0.03 percent. HDI in Indonesia in 2021, with the highest HDI being DKI Jakarta Province at 81.11 and the lowest HDI being Papua Province at 60.62. There are 10 provinces that achieve HDI above the national average, one of which is West Sumatra province and is in third position out of ten provinces on the island of Sumatra with HDI achievements of 72.65 in 2021. Although the HDI achievement of West Sumatra Province is more than the national average, there are still regencies/cities in West Sumatra Province with HDI that is less than the national average. In 2017 West Sumatra's HDI of 71.24 increased to 71.73 in 2018, and 72.39 in 2019. In 2020 it was 72.38 due to the Covid-19 pandemic situation. Then with the recovery, the HDI 2021 showed an increase compared to the previous year. In addition to paying attention to time on the development of human development, the next thing to pay attention to is to look at the state of the region or region.

The achievement of HDI in the City area is different from the Regency area. In 2021, the highest achievement of the human development index by Padang City was 82.90 and the lowest achievement of human development by the Mentawai Islands was 61.35. According to the Central Statistics Agency (2022: 51) the gap occurs due to several factors such as uneven distribution of natural resources, human resources, history, and geographical location. Thus, HDI data from District/City in West Sumatra Province is thought to have spatial effects.

Factors that have a significant influence on the human development index of West Sumatra Province can be known based on a certain method. Therefore, data in the study contains location information and also observed in several periods (years) and this data known as panel data (Yulianti et al., 2021). The panel data spatial regression analysis method is one approach that can be used to examine data involving districts/cities in a certain period of time by involving spatial effects. Classical regression analysis will indicate inaccuracies when combined data that have spatial effects are used, so using the panel data spatial regression method is very suitable to be applied in this study (Anselin, 2010).

The spatial regression model of panel data has a spatial lag effect on the dependent variable and has a spatial error effect. These spatial models are known as the Spatial Autoregressive Model (SAR) model for lag and the Spatial Error Model (SEM) model for error (Elhorst, 2014). In the case of SAR, changes in the explaining variable of a region were indicated to affect not only that region through the direct effect, but that they can also affect the dependent variable in all regions through indirect effects Ayuwida et al. (2021) In SEM models, the spatial dependence is manifested in the error term, highlighting that the errors associated to any observation are an average of the errors in the neighboring regions, added to a random component (Raiher et al., 2017). Panel data has three methods for modeling data, namely the Common Effect Model (CEM) which is the most basic model in panel data regression and is often referred to as the combined influence model, Fixed Effect Model (FEM) or fixed influence model and Random Effect Model (REM) or random influence model.

Modeling on HDI was conducted by Ukra et al. (2022) using panel spatial regression for HDI data in Kalimantan for 2017-2021. The results of this research concluded that the Fixed Effect Spatial Autoregressive (FE-SAR) model is an appropriate model for HDI in Regencies/Cities in Kalimantan in 2017-2021, and the influence of spatial proximity shows that the HDI in Regencies/Cities throughout Kalimantan in 2017-2021 is influenced by the average

HDI in other adjacent districts/cities at a certain time. In addition, from this study, the average length of schooling affects HDI. Another study was conducted by Saputri & Suryowati (2018) Analysis of factors affecting the gini ratio in Papua Province with a spatial model of panel data, spatial autoregressive random effect model as the best model obtained with variables that have a significant effect on the gini ratio are the Human Development Index and crop harvest area.

Based on the problems that have been described that the Human Development Index in the Districts/Cities of West Sumatera Province is indicated to have an influence on the interrelation of neighboring regions and involves the influence of time seen from the difference in HDI development each year. So modeling is needed that considers both effects in order to get good results. The aim of this research is to find out the general description of the HDI of West Sumatra Province, obtain a panel data spatial regression model and obtain variables that significantly influence on HDI in West Sumatra Province 2017-2021. Therefore, the right analysis to examine the Human Development Index in Districts/Municipalities in West Sumatra Province along with HDI factors that influence is used a panel spatial model involving spatial effects with the rook contiguity weighting function for the time period, namely 2017-2021.

B. METHODS

1. Data And Methods

The type of data used in this research is secondary data sourced from publications by the official government agency BPS West Sumatra Province. The data used is Regency/City Human Development Index data in West Sumatra Province for the period 2017-2021 with four Human Development Index indicators, namely morbidity rate, expected length of schooling, average length of schooling, and population density. The method used in this research is the panel data spatial analysis method with the first analysis stage being to describe the data, then carry out panel data analysis, then carry out spatial analysis of panel data with a spatial weighting matrix followed by measuring the goodness of the panel data spatial model based on model criteria, and The final step is to analyze the best model that has been obtained and write conclusions.

2. Panel Data Regression Analysis

Panel data is combined data between time series and cross sections. Data in the form of daily, monthly, yearly and so on includes time series units while data on individuals, households, companies, regions, countries and others include cross section units (Alfiani et al., 2022). The following panel data model is measured from a cross section and time series (Y. Wang et al., 2023).

$$Y_{it} = \alpha + \beta x'_{it} + u_{it} \quad (1)$$
$$i = 1,2,3,\ldots,N \; ; \; t = 1,2,3,\ldots,T$$

where i for show individual; t for shows the dimensions of the time series; α for Intercept coefficient; β for Slope coefficient with dimensions K x 1, where K is the number of free modifiers; Y_{it} for Dependent variable vectors i individual unit and t time unit; x'_{it} is independent variable vectors for i individual units and t time units; u_{it} is error in i observation unit and t time. Panel data regression models have three approaches to estimating models, which are as follows:

a. Common Effect Model (CEM)
   Common Effect Model combining time series and cross section data that assumes that the behavior of the data is constant over time is then regressed by the Ordinary Least Square method resulting in the same intercept and independent variable coefficients for each unit. In general, the equation is as follows (Kim, 2021).

$$\mathbf{Y}_{\boldsymbol{it}} = \alpha + \boldsymbol{\beta x'}_{\boldsymbol{it}} + \boldsymbol{u}_{\boldsymbol{it}} \tag{2}$$
$$i = 1,2, \dots, N \text{ dan } t = 1,2,\dots, \mathrm{T}$$

   Estimation with the OLS method so that the equation obtained is as follows.

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{3}$$

b. Fixed Effect Model (FEM)
   Fixed Effect Model Assumes that the slope coefficients are the same but between individuals there are differences in intercepts. This method is often called the Least Square Dummy Variable model because when guessing Fixed Effect Model parameters using the technique of adding dummy variables. In general, according to (N. Wang et al., 2023) the equation is written as follows:

$$\mathbf{Y}_{\boldsymbol{it}} = \boldsymbol{\alpha}_{\boldsymbol{i}}\boldsymbol{D}_{\boldsymbol{ij}} + \boldsymbol{\beta x'}_{\boldsymbol{it}} + \boldsymbol{u}_{\boldsymbol{it}} \tag{4}$$
$$D_{ij} = 1, i = j \,;\, D_{ij} = 0,\ i \neq j$$

   using the steps of the OLS estimation method, the parameters can be obtained LSDV estimation as follows:

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i}^{N} \boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{Q}\boldsymbol{X}_{\boldsymbol{i}}\right]^{-\mathbf{1}}.\left[\sum_{i}^{N} \boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{Q}\boldsymbol{y}_{\boldsymbol{i}}\right] \tag{5}$$

c. Random Effect Model (REM)
   Random Effect Model used to estimate panel data when residual variables are suspected to be interconnected between time and individuals. The equation is generally as follows (Ukra et al., 2022):

$$\mathbf{Y}_{it} = \alpha_0 + \boldsymbol{\beta x'}_{it} + \cdots + \boldsymbol{v}_{\boldsymbol{it}} \tag{6}$$
$$\boldsymbol{v}_{\boldsymbol{it}} = \boldsymbol{\varepsilon}_{\boldsymbol{i}} + \boldsymbol{\mu}_{\boldsymbol{it}}$$

   using the GLS procedure will be obtained the following estimator.

$$\hat{\beta} = [X'\Omega^{-1}X]^{-1}[X'\Omega^{-1}Y] \tag{7}$$

3. Matrix Spatial Weighting

According to Tobler's First Law of Geography, everything is related to everything else, although close objects are more related to one another than far objects (Tobler, 2010). A spatial weight matrix in spatial models expresses the structure of geographical locations and lends more weight to regions that are closer together (Chi & Zhu, 2020). There are several ways of defining neighborliness according to (Ukra et al., 2022) as follows:

a. Rook Contiguity Rook Contiguity refers to a location that is adjacent to the location of concern.
b. Bishop Contiguity Bishop Contiguity refers to locations where the angular point meets the angle of the location of concern.
c. Queen Contiguity Queen Contiguity refers to a location that is side by side and its corner point meets the location of concern.

4. Spatial Regression Analysis

The general model of spatial regression according to(Lesage & Pace, 2009) is as follows.

$$\mathbf{y} = \gamma\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\phi} \qquad (8)$$
$$\boldsymbol{\phi} = \lambda\mathbf{W}\boldsymbol{\phi} + \boldsymbol{\varepsilon}$$

where y for vector dependent variable of size N x 1; X for matrix of independent variables of size N x (K1); β for vector coefficient of regression parameter size (K 1) x 1; γ for spatial parameter lag coefficient in spatial regression; λ for coefficient of spatial parameter error in spatial regression; φ for equation error vector; ε for equation error vector; W for standardized spatial weighting matrix of size N x N; N for number of cross section units. The general spatial regression model can be used to create a variety of models, including the following.

a. Spatial Autoregressive Model (SAR)

The SAR model to be obtained is as follows (Qi et al., 2021).

$$\mathbf{y} = \gamma\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (9)$$

Parameter estimation for the Spatial Autoregressive Model using the maximum likelihood approach (Lesage & Pace, 2009):

$$\boldsymbol{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \gamma\mathbf{W})\boldsymbol{y} \qquad (10)$$

b. Spatial Error Model (SEM)

The SEM model to be obtained is as follows (Qi et al., 2021):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\phi} \qquad (11)$$
$$\boldsymbol{\phi} = \lambda\mathbf{W}\boldsymbol{\phi} + \boldsymbol{\varepsilon}$$

Estimation of Spatial Error Model parameters obtained using the Maximum likelihood method (Lesage & Pace, 2009):

$$\boldsymbol{\beta} = [(\mathbf{X} - \lambda\mathbf{WX})^{\mathrm{T}}]^{-1}(\mathbf{X} - \lambda\mathbf{WX})^{\mathrm{T}}(\mathbf{I} - \lambda\mathbf{W})\boldsymbol{y} \tag{12}$$

c. Spatial Autoregressive Moving Average (SARMA)
   The general model of SARMA is as follows:

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon} \tag{13}$$

   Estimation of β parameters for the SARMA model using the maximum likelihood method (Widyastuti et al., 2019):

$$\boldsymbol{\beta} = [(\mathbf{X} - \lambda\mathbf{WX})^{\boldsymbol{T}}(\mathbf{X} - \lambda\mathbf{WX})]^{-1}(\mathbf{X} - \lambda\mathbf{WX})^{\boldsymbol{T}}(\mathbf{I} - \lambda\mathbf{W} - \rho\mathbf{W})\boldsymbol{y} \tag{14}$$

5. Spatial Autocolleration

The correlation between a variable and itself based on location is measured as spatial autocorrelation (N. Wang et al., 2023). Moran's index is done using the following formula (Majeed & Mazhar, 2021).

$$\mathrm{I} = \frac{\mathrm{n}\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\mathrm{w_{ij}}(\mathrm{x_i}-\bar{\mathrm{x}})(\mathrm{x_j}-\bar{\mathrm{x}})}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\mathrm{w_{ij}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i}-\bar{\mathrm{x}})^2} \tag{15}$$

where I stands for the global Moran index; x_i and x_j for the level of land use dependency in cities i and j; n for the number of cities; $\bar{x}$ for the average level of land use dependence; and W_{ij} for the spatial weight matrix (N. Wang et al., 2023). The presence or absence of autocorrelation in the data can be determined by comparing the value of Moran's I (I) with the expected value of Moran's I (I_0) in the following hypothesis. H_0 : I = I_0 (no spatial autocorrelation between locations); H_1 : I ≠ I_0 (There is spatial autocorrelation between locations). The values I > I_0 cause the spatial autocorrelation values to be positive and the data pattern formed is clustered. If I< I_0 then it has the meaning of negative autocorrelation and the pattern of data formed is diffuse. if I = I_0 then it means that there is no spatial autocorrelation.

6. Lagrange Multiplier (LM) Test

The Lagrange Multiplier (LM) test is used to select suitable spatial regression models (Lesage & Pace, 2009). The Lagrange Multiplier test is divided into two parts: LM_{lag} and LM_{error}. (Slučiaková, 2021). If both are significant SARMA (Spatial Autoregressive Moving Average) is a suitable model.

7. Regresi Spatial Panel Models

Create a standard linear panel data model without any consideration for spatial effects. This model can be used as a guide for spatial panel data models' estimation outcomes, as well as to evaluate how reliable they are (Guliyev, 2020).

a. Spatial Lag Panel Data Models
   The spatial model equation lags panel data as follows (Golgher & Voss, 2015; H. Wang et al., 2015).

$$\boldsymbol{y} = \delta \boldsymbol{W}_{NT}\boldsymbol{y} + \boldsymbol{X\beta} + (\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\varepsilon} \tag{16}$$

Where δ for Spatial parameter coefficient lag in spatial model; $\mathbf{W}_{NT}$ for Data lag panel standardized spatial weighting matrix i-th row j-th; $\boldsymbol{y}$ for Column dependent variable vector size NT x 1; $\boldsymbol{X}$ for Independent variable matrix size NT x K. If δ is statistically significant, it shows that the dependent variables have a considerable spatial dependence. In other words, the contiguous regions are required for a confirmed case in a region. The magnitude of the geographical dependence is indicated by the value of δ (Gelfand et al., 2010).

b. Spatial Error Panel Data

The spatial model error panel data is expressed as follows (Elhorst, 2014; H. Wang et al., 2015; Yulianti et al., 2021):

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{X\beta} + (\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + \boldsymbol{\phi} \\ \boldsymbol{\phi} = \rho \boldsymbol{W_{NT}}\, \boldsymbol{\phi} + \boldsymbol{\varepsilon} \end{gathered} \tag{17}$$

Where ρ for Error spatial parameter coefficient in panel data error spatial model; $\boldsymbol{\phi}$ for Equation error vector that is size NT x 1; $\boldsymbol{\varepsilon}$ for Equation error vector that is size NT x 1.

8. Spasial Panel Model Estimation

a. Fixed Effect – Spatial Autoregressive Model

The log-likehood function of the fixed effect lag spatial model is shown in the following equation (Kim, 2021; L. Lee & Yu, 2010).

$$Ln\, L = -\frac{NT}{2}\, Ln(2\pi\sigma^2) + TLn|(\boldsymbol{I}_N - \delta \boldsymbol{W})| - \frac{1}{2\pi\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \delta \sum_{j=1}^{N} w_{ij}y_{jt} - \boldsymbol{x_{it}\beta} - \mu_i\right)^2 \tag{18}$$

Estimation of model parameters for spatial lag fixed effect with Maximum Likelihood as follows. The * indicates the fixed effect model.

$$\widehat{\boldsymbol{\beta}} = ((\boldsymbol{X}^*)'\boldsymbol{X}^*)^{-1}(\boldsymbol{X}^*)'\boldsymbol{y}^* - ((\boldsymbol{X}^*)'\boldsymbol{X}^*)^{-1}(\boldsymbol{X}^*)'\delta(\boldsymbol{\iota}_T \otimes \boldsymbol{W})\boldsymbol{y}^* \tag{19}$$

b. Fixed Effect – Spatial Error Model

The log-likehood function of the fixed effect error spatial model is as follows (H. Wang et al., 2015).

$$Ln\, L = -\frac{NT}{2} Ln(2\pi\sigma^2) + TLn|\boldsymbol{I}_N - \delta \boldsymbol{W}| - \frac{1}{2\pi\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it} - \rho \sum_{j=1}^{N} y_{it} - \sum_{p=1}^{K} \beta_p \left(x_{itp} - \rho \sum_{j=1}^{N} w_{ij}x_{jtp}\right) - \left(\mu_i - \rho \sum_{j=1}^{N} w_{ij}\mu_j\right))^2 \tag{20}$$

The estimation of model parameters for spatial error fixed effect with maximum likelihood is as follows. The * indicates the fixed effect model.

$$\widehat{\boldsymbol{\beta}}=\{[\boldsymbol{X}^* - \rho(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]'[\boldsymbol{X}^* - \rho(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]\}^{-1}[\boldsymbol{X}^* - \rho(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]'[\boldsymbol{y}^* - \rho(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{y}^*] \quad (21)$$

c. Random Effect – Spatial Autoregressive Model
The log-likelihood function of the model is as follows with the transformation of the dependent variable to θ (Elhorst, 2014).

$$Ln\, L = -\frac{NT}{2} ln[e(\theta)'e(\theta)] + \frac{N}{2} ln\theta^2 \quad (22)$$

Several parameter values $\widehat{\boldsymbol{\beta}}, \delta$ and σ^2 are used in the iteration procedure until a convergent estimated value θ is obtained.

d. Random Effect – Spatial Error Model
The log-likelihood function of the model obtained from the transformation results is as follows (Elhorst, 2014).

$$y^{\circ}_{it} = y_{it} - \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{j=1}^{N} \left\{ \left[p_{ij} - \left(1 - \rho w_{ij}\right) \right] \frac{1}{T} \sum_{t=1}^{T} y_{jt} \right\} \quad (23)$$

The estimation of $\widehat{\boldsymbol{\beta}}$ and σ^2 given at ρ and φ can be done by regression of OLS between $\mathbf{Y}$ and $\mathbf{X}$. $\mathbf{e}^{\circ} = \mathbf{Y}^{\circ} - \boldsymbol{\beta}\mathbf{X}^{\circ}$ obtained $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\circ\prime}\mathbf{X}^{\circ})^{-1}\mathbf{X}^{\circ\prime}\mathbf{Y}^{\circ}$ and $\hat{\sigma}^2 = (\mathbf{Y}^{\circ}\mathbf{X}^{\circ}\boldsymbol{\beta})'(\mathbf{Y}^{\circ} - \mathbf{X}^{\circ}\boldsymbol{\beta})/n\mathrm{T}$. Where ° indicates the random effect model.

C. RESULTS

1. Description Analysis

Based on Figure 1, it shows that the map of the distribution of the human development index of all districts/cities in West Sumatra Province in 2021, the more intense the color, the higher the HDI achievement in the area, on the contrary, the brighter the color, the lower the HDI achievement value in the area. Areas with very high HDI achievements are Padang City at 82.90 and Bukittinggi City at 80.70. Areas that achieved high HDI categories were Payakumbuh City, Solok City, Padang Panjang City, Pariaman City, Sawahlunto City, Agam, Tanah Datar, Dharmasraya, Padang Pariaman, South Coast. While the areas included in the medium HDI category are Fifty Cities, Solok, South Solok, West Pasaman, Sijunjung, Pasaman and Mentawai Islands. The lowest HDI achievement value was 61.35 in the Mentawai Islands, as shown in Figure 1.

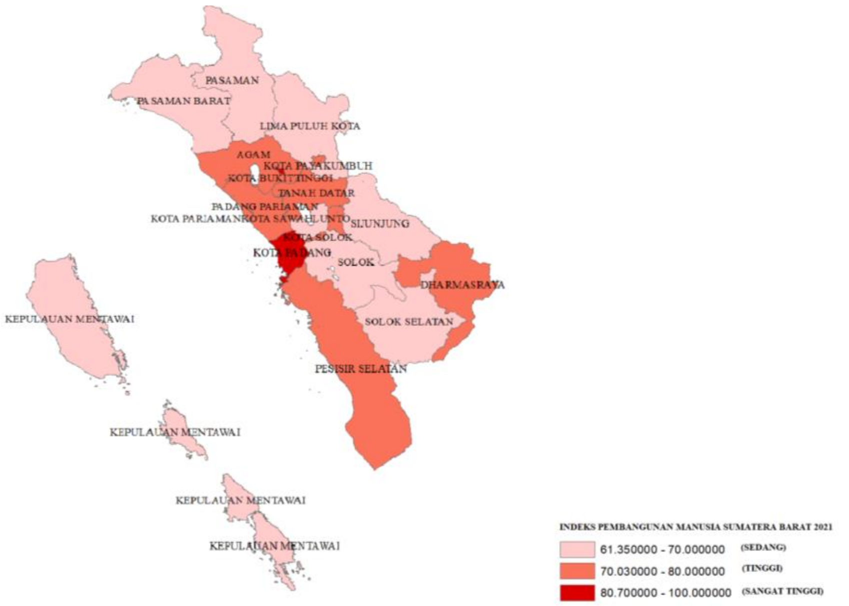

Figure 1. HDI Distribution Map

2. Panel Data Regression Model Selection

a. Chow Test

hypothesis as follows. H_0 is Model Common effect; H_1 is Model Fixed Effect, as shown in Table 1.

Table 1. Chow Test

F Statistic	*P-Value*
132,54	2,2 x 10^{-16}

Based on Table 1 that the P-Value of the Chow test is 2,2 x 10^{-16}. At a significance level of 5%, the P-Value is less than 0.05. So that conclusions can be drawn rejecting H0. That is, the selected model is Fixed Effect.

b. Hausman Test

Hypothesis as follows. H0 is Model Random effect; and H1 is Model Fixed Effect.

Table 2. Hausman Test

Chisq Statistic	*P-Value*
3,3598	0,4995

Table 2 of the Hausman test shows the P-Value is 0.4995. At the significance level of 5%, the P-Value of the hausman test is more than 0.05 and conclusions can be drawn to accept H0. That is, the selected model is the Random Effect model.

3. Spatial Weighted Matrix

The weighting matrix used in this study is arranged based on side intersection. The matrix arranged based on the intersection of the sides is called the Rook Contiguity matrix.

4. Moran's I

Table 3 shows that overall Moran's values $I(I) \neq I_0$ can be concluded that there is a spatial autocorrelation relationship. When $I > I_0$, positive spatial autocorrelation values form a clustered data pattern. When $I < I_0$, negative autocorrelation values form a diffuse data pattern. When $I = I_0$, it means there is no spatial autocorrelation.

Table 3. Moran's I Calculation Results

Year	Variable	*I*	*E(I)*
2017	Y	0.04287	-0.0588
	X_1	0.09224	-0.0588
	X_2	0.18839	-0.0588
	X_3	-0.0789	-0.0588
	X_4	0.19915	-0.0588
2018	Y	0.05221	-0.0588
	X_1	-0.8334	-0.0588
	X_2	0.19897	-0.0588
	X_3	-0.0513	-0.0588
	X_4	0.01981	-0.0588
2019	Y	0.05347	-0.0588
	X_1	0.04278	-0.0588
	X_2	0.21852	-0.0588
	X_3	-0.0562	-0.0588
	X_4	0.2072	-0.0588
2020	Y	0.04532	-0.0588
	X_1	0.17166	-0.0588
	X_2	0.21878	-0.0588
	X_3	-0.0709	-0.0588
	X_4	0.21137	-0.0588
2021	Y	0.04244	-0.0588
	X_1	-0.1955	-0.0588
	X_2	0.21831	-0.0588
	X_3	-0.081	-0.0588
	X_4	0.21085	-0.0588

5. Lagrange Multiplier Test

Table 4 shows the results of the analysis of the multiplier lagrange test for the lag model and model error. The SAR model shows a P-Value of 0.006087 < 5% significance level so that it rejects H_0, meaning that there is a spatial lag dependence. Whereas the P-Value of the SEM model shows 0.9869 > 5% significance level so that it accepts H_0, meaning that there is no spatial error dependence.

Table 4. Lagrange Multiplier Test

Lagrange Multiplier Test	LM Score	*P-Value*
SAR	7,5243	0,006087
SEM	0,00026	0,9869

Based on the output of the LM test, parameter estimation and model building will be continued using the Spatial Autoregressive Model because the spatial variables are real in the lag.

6. Parameter Estimation of Random Effect-SAR

Based on the previous analysis, the parameter model is assumed to be the Random Effect panel model and the spatial model used is the Spatial Autoregressive model. The results of parameter estimation of the Random Effect Spatial Autoregressive Model can be seen in Table 5.

Table 5. Estimation results of the Random Effect-SAR model

Variable	Coefisien	*P-Value*
Intercept	2,9344550	$2{,}2 \times 10^{-16}$
δ	-0,010130	0,01324
X_2	0,2534341	$2{,}025 \times 10^{-9}$
X_3	0,3001283	$2{,}2 \times 10^{-16}$
X_4	0,0108164	0,0002386
	R^2 = 0,9294	

In Table 5 the Random Effect Spatial Autoregressive Model obtained is written as follows.

$$\widehat{HDI}_{it} = 2{,}93446 - 0{,}0101 \sum_{j=1}^{19} W_{ij} HDI_{jt} + 0{,}253 HLS_{it} + 0{,}3 RLS_{it} + 0{,}01 KP_{it} + v_{it}$$

The results of parameter estimation of the Random Effect Spatial Autoregressive model can be explained that the variables of expected length of schooling (X_2), average length of schooling (X_3), and population density (X_4) are significant because the P-Value < 0.05. This means that the expected length of schooling, average length of schooling, and population density in a district/city in West Sumatra affect the HDI in that district/city. In the expected length of school variable (X_2), average length of schooling (X_3), and population density (X_4) the resulting parameter estimates have a positive relationship. That is, the greater the expected length of schooling (X_2), the average length of schooling (X_3), and the population density (X_4) in a district/city, the higher the HDI in that district/city. The HDI spatial lag coefficient (δ) has a significant P-Value, meaning that the HDI in a district/city is affected by the HDI in neighboring districts/cities. The influence of a district/city is equal to the coefficient (δ= -0.0101) multiplied by the average of neighboring districts/cities.

7. Model Interpretation

An example of a model formed in South Solok Regency in 2021 is written as follows.

$$\widehat{HDI}_{\text{Solok Selatan},2021} = 2{,}93446 - 0{,}0101(0.3333\,\widehat{HDI}_{\text{Kab. Solok},t} + 0{,}3333\,HDI_{\text{Pesisir Selatan},t} + 0{,}3333\,\widehat{HDI}_{\text{Dharmasraya},t}) + 0{,}253\,HLS_{it} + 0{,}3\,RLS_{it} + 0{,}01\,KP_{it} - 0.02081$$

$$\widehat{HDI}_{\text{Solok Selatan},2021} = 2{,}91365 - 0{,}0036\,(\widehat{HDI}_{\text{Kab. Solok},t} + \widehat{HDI}_{\text{Pesisir Selatan},t} + \widehat{HDI}_{\text{Dharmasraya},t}) + 0{,}253\,HLS_{it} + 0{,}3\,RLS_{it} + 0{,}01\,KP_{it}$$

The equation for the Random Effect Spatial Autoregressive model in Solok Selatan Regency in 2021 can be interpreted that any increase in the expected number of years of schooling will increase the HDI rate by 0.253. Each increase in the average length of schooling will increase the HDI score by 0.3. Each increase in population density will increase the HDI number by 0.01. The HDI for Solok Selatan Regency in 2021 is influenced by the HDI in Solok Regency, Pesisir Selatan Regency, and Dharmasraya Regency which are neighbors. The influence of the district/city is -0.0036. This means that if the HDI in Solok Regency, Pesisir Selatan Regency, and Dharmasraya Regency increases by one unit, then the HDI in Solok Regency decreases by 0.0036 and other variables are considered constant.

D. CONCLUSION AND SUGGESTIONS

After carrying out the stages of analysis and obtaining the results discussed earlier, it is concluded that in 2021 the area with the highest HDI achievement is Padang City at 82.90 and the lowest is Mentawai Island at 61.35. Areas with very high HDI category achievements are Padang City and Bukittinggi City. Areas that achieved high HDI categories are Payakumbuh City, Solok City, Padang Panjang City, Pariaman City, Sawahlunto City, Agam, Tanah Datar, Dharmasraya, Padang Pariaman, South Coast. While the areas included in the medium HDI category are Fifty Cities, Solok, South Solok, West Pasaman, Sijunjung, Pasaman and Mentawai Islands. The model formed from the analysis of human development index data in West Sumatra Province uses the panel data spatial regression method, namely the Random Effect Spatial Autoregressive Model.

$$\widehat{HDI}_{it} = 2{,}93446 - 0{,}0101 \sum_{j=1}^{19} W_{ij} HDI_{jt} + 0{,}253 HLS_{it} + 0{,}3 RLS_{it} + 0{,}01 KP_{it} + v_{it}$$

This model is used because in the human development index data in West Sumatra Province, the spatial variable is real in lag. The Random Effect Spatial Autoregressive Model has the largest R^2 value with a gain of 0.9242 which means that this shows that the human development index in West Sumatra Province in 2017-2021 can be explained by the factors contained in the model of 92.42%. The variables that significantly affect the human development index in West Sumatra are the expected length of schooling (X_2), the average length of schooling (X_3), and population density (X_4). The author's suggestions for this research is in future studies can use other weights, such as Customize Contiguity which can consider the proximity of neighboring areas in terms of economic, social, transportation and other factors. Then the variables used in this study can be added again for further research so that there are many significant factors in the study so that it can provide better information.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all the individuals that have contributed to the publication of this research paper. We are also grateful to the Department of Statistics at Universitas Muhammadiyah Semarang for providing us with the resources and support we needed to complete this project.

REFERENCES

Alfiani, S., & Arum, P. R. (2022). Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression. In *Universitas Muhammadiyah Semarang Jl. Kedungmundu* (Vol. 15, Issue 2). www.unipasby.ac.iJurnal Ilmiah Teori Dan Aplikasi Statistika (Vol. 15, Issue 2). https://doi.org/10.36456/jstat.vol15.no2.a5506d

Anselin, L. (2010). *Spatial Econometrics Methods and Models* (Vol. 4). Dordrecht, Springer Netherlands.

Ayuwida, C. A., Arum, P. R., & Al Haris, M. (2021). Model Seemingly Unrelated Regression Pada Data Kemiskinan Jawa Timur Menggunakan Matriks Pembobot Queen Contiguity Dan Rook Contiguity. *Jurnal Statistika Universitas Muhammadiyah Semarang*, *9* (1). https://doi.org/doi.org/10.26714/jsunimus.9.1.2021.64-68

Chi, G., & Zhu, J. (2020). *Spatial Regression Models for the Social Sciences*. https://doi.org/10.4135/9781544302096

Elhorst, J. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. https://doi.org/10.1007/978-3-642-40340-8

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of Spatial Statistics* (1st ed., Vol. 1). CRC Press. https://doi.org/10.1201/9781420072884

Golgher, A., & Voss, P. (2015). How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects. *Spatial Demography*, *4*, 1–31. https://doi.org/10.1007/s40980-015-0016-y

Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*, *38*. https://doi.org/10.1016/j.spasta.2020.100443

Kim, I. (2021). Spatial distribution of neighborhood-level housing prices and its association with all-cause mortality in Seoul, Korea (2013–2018): A spatial panel data analysis. *SSM - Population Health*, *16*. https://doi.org/10.1016/j.ssmph.2021.100963

Lee, L., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*(2), 165–185. https://doi.org/10.1016/j.jeconom.2009.08.001

Lesage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics* (1st ed., Vol. 1). Taylor & Francis. https://doi.org/https://doi.org/10.1201/9781420064254

Majeed, M., & Mazhar, M. (2021). An Empirical Analysis of Output Volatility and Environmental Degradation: A Spatial Panel Data Approach. *Environmental and Sustainability Indicators*, *10*, 100104. https://doi.org/10.1016/j.indic.2021.100104

Masruroh, M., & Subekti, R. (2016). Aplikasi Regresi Partial Least Square Untuk Analisis Hubungan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia Di Kota Yogyakarta. *Media Statistika*, *9(2)*, 75–84. https://doi.org/10.14710/medstat.9.2.75-85

Qi, Z., Lin, H., & Junya, B. (2021). Spatial Autoregressive Analysis of Nationwide Street Network Patterns with Global Open Data. *Environment and Planning B Urban Analytics and City Science*, *0*, 1–18. https://doi.org/10.1177/2399808320987846

Raiher, A., Carmo, A., & Stege, A. (2017). The effect of technological intensity of exports on the economic growth of Brazilian microregions: A spatial analysis with panel data. *EconomiA*, *18*. https://doi.org/10.1016/j.econ.2017.03.001

Saputri, W. A. K., & Suryowati, K. (2018). Analisis Faktor-Faktor yang Mempengaruhi Gini Ratio Di Provinsi Papua Dengan Model Spasial Data Panel. *Jurnal Statistika Industri Dan Komputasi*, *03*(2), 1–11. https://doi.org/10.34151/statistika.v3i02.1060

Siswoyo, E. (2014). *Kesenjangan dan Ketimpangan Akibat Pembangunan Tidak Merata*. https://tubasmedia.com/

Slučiaková, S. (2021). Effects of the unit-based pricing of waste in Slovakia: Spatial panel data models and matching approach. *Environmental Challenges*, *2*, 100022. https://doi.org/10.1016/j.envc.2021.100022

Tobler, W. (2010). Geographical Filters and Their Inverses. *Geographical Analysis*, *1*, 234–253. https://doi.org/10.1111/j.1538-4632.1969.tb00621.x

Ukra, M. G., Nusrang, M., & Poerwanto, B. (2022). Regresi Panel Spasial Untuk Pemodelan Indeks Pembangunan Manusia Di Kabupate/Kota Se-Kalimantan. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(2), 70–78. https://doi.org/10.35580/variansiunm34

Wang, H., Du, Z., Wang, X., Liu, Y., Yuan, Z., Liu, Y., & Xue, F. (2015). Detecting the association between meteorological factors and hand, foot, and mouth disease using spatial panel data models. *International Journal of Infectious Diseases*, *34*, 66–70. https://doi.org/10.1016/j.ijid.2015.03.007

Wang, N., Zhang, X., Wang, Z., Chen, Y., & Li, S. (2023). Can financial development improve environmental quality? New findings from spatial measures of Chinese urban panel data. *Heliyon*, *9*(7), e17954. https://doi.org/10.1016/j.heliyon.2023.e17954

Wang, Y., Wang, M., Wu, Y., & Sun, G. (2023). Exploring the effect of ecological land structure on PM2.5: A panel data study based on 277 prefecture-level cities in China. *Environment International*, *174*, 107889. https://doi.org/10.1016/j.envint.2023.107889

Widyastuti, M. N., Srinadi, I. G. A. M., & Susilawati, M. (2019). Pemodelan Jumlah Kasus Pneumonia Balita Di Jawa Timur Menggunakan Regresi Spatial Autoregressive Moving Average. *E-Jurnal Matematika*, *8*(3), 236. https://doi.org/10.24843/mtk.2019.v08.i03.p259

Yulianti, S., Widyanigsih, Y., & Nurrohmah, S. (2021). Spatial panel data model on human development index at Central Java. *Journal of Physics: Conference Series*, *1722*, 012090. https://doi.org/10.1088/1742-6596/1722/1/012090