Science and Technology Indonesia
e-ISSN:2580-4391 p-ISSN:2580-4405
Vol. 10, No. 1, January 2025

Research Paper

Modeling of Human Development Index Using Bayesian Spatial Autoregressive Approach

Ferra Yanuar[1*], Sintya Wulandari[1], Yudiantri Asdi[1], Aidinil Zetra[2], Haripamyu[1]

[1]*Department of Mathematics and Data Science, Andalas University, Padang, 25163, Indonesia*
[2]*Department of Political Sciences, Andalas University, Padang, 25163, Indonesia*
***Corresponding author**: ferrayanuar@sci.unand.ac.id

Abstract
Spatial regression analysis is a technique employed to examine the relationship between independent and dependent variables in datasets that exhibit regional neighborhood influences or spatial effects. When a spatial effect exists for the independent variable, the Spatial Autoregressive (SAR) regression can be utilized. The Maximum Likelihood Estimation (MLE) is a commonly used parameter estimator for SAR. However, due to the limitations of MLE, the Bayesian method provides an alternative approach for parameter estimation. This study compares the results of SAR estimations using both MLE and Bayesian methods to determine the most accurate estimation model. Both methods were implemented in this research to model the factors affecting the Human Development Index (HDI) in East Java Province for the year 2022. The findings indicate that the Bayesian SAR offers a superior proposed model compared to the MLE SAR. The factors influencing the HDI in East Java Province in 2022 include poverty, per capita expenditure, and the presence of an upper middle-class manufacturing industry.

Keywords
HDI, Spatial Autoregressive, Bayesian Spatial Autoregressive, Maximum Likelihood Estimation

Received: 9 February 2024, Accepted: 1 October 2024
https://doi.org/10.26554/sti.2025.10.1.72-79

1. INTRODUCTION

The Human Development Index (HDI) is a method used to measure and evaluate efforts to improve the quality of life of people in an area (Caniago and Wibowo, 2024). One of the provinces whose HDI score has increased from year to year is East Java Province, which when compared to other provinces on the island of Java shows a fairly high increase. In 2022, East Java's HDI reached 72.75, an increase of 0.61 points from the previous year, which was at 72.14 points. However, when compared to other provinces on the island of Java, East Java Province is the province with the lowest HDI score (Badan Pusat Statistik Provinsi Jawa Timur, 2023). The high or low score of the HDI is influenced by various factors that serve as guidelines for improving human development (Ogujiuba et al., 2024; Tyas and Sukartini, 2022). For this reason, modeling is needed to identify the factors that influence and must be improved in efforts to increase the HDI in East Java.

Factors across regions often play a significant role in influencing HDI within a specific area (Dai and Jin, 2021; Yanuar et al., 2023a,b), commonly referred to as spatial effects, including spatial dependence and spatial diversity (Zhang et al., 2021a,b). Spatial dependence arises from the interconnection between regions, whereas spatial diversity stems from the variations between different regions (Yasin et al., 2020, 2022; Yu et al., 2022). In the event of such occurrences, the appropriate data modeling technique to employ is the Spatial Autoregressive (SAR) regression method (Ver Hoef et al., 2018; Yanuar et al., 2023b). The estimation method commonly used in SAR is the Maximum Likelihood Estimation (MLE) method (Anselin, 2009). Because of the MLE method's limitations, such as its inability to overcome heterogeneity issues, its poor performance in estimating models for small data, the difficulty in calculating standard error estimates for hypothesis testing for large sample sizes, and the strict assumption of the need for normality distribution, modeling is performed using the Bayesian method. A Bayesian method provides various advantages. It provides a handy approach for merging prior knowledge with data within the right statistical context. Bayesian inference is conditional on the data and exact without relying on asymptotic approximations, which eliminates concerns about the error term's normal distribution. For practitioners, the results of Bayesian estimate are more comprehensible and understandable.

The development of Bayesian techniques provides a wide range of tools suitable for estimating spatial models. Anselin (1988) investigates the Bayesian approach applied to pure spatial autoregressive and spatial error models, proposing diffuse priors for model parameters and analytically deriving the marginal posterior distributions for these parameters. Hepple (1995) further builds on this by creating Bayesian analyses for important spatial specifications, including the SARAR(1,0) model, the SARAR(0,1) model (commonly referred to as the SEM), and spatial moving average models, which are known as SARMA(0,1). In each case, the joint posterior distributions of the parameters are presented, from which the marginal posterior distributions of the spatial autoregressive parameters are analytically derived.

Recent research has utilized the Markov Chain Monte Carlo (MCMC) method to estimate spatial models, particularly in instances where analytically simplifying the marginal posterior distributions or numerically integrating them poses challenges. The MCMC method has been applied to various types of spatial models in the studies conducted by Lesage (1997) and LeSage and Pace (2009). Kakamu and Wago (2008) compare the finite sample properties of Bayesian estimators derived from the MCMC method with those obtained through Maximum Likelihood Estimation (MLE) in the context of the static panel spatial autoregressive model. In spatial models, the boundaries of the parameter space for spatial autoregressive parameters are clearly defined, facilitating the selection of suitable uninformative priors for these parameters. As a result, many studies assign uniform priors across the parameter space for spatial autoregressive parameters, as evidenced in the research of (Kakamu and Wago, 2008; Lesage, 1997; Parent and Lesage, 2008). Oliveira and Song (2008) investigate two forms of Jeffreys' prior, referred to as independence Jeffreys and Jeffreys-rule priors, specifically for the spatial autoregressive parameter within the framework of the SARAR(0,1) model. Jeffreys' priors are viewed as uninformative improper priors derived from the information matrix, and one of their appealing attributes is their resistance to changes in model reparameterization. Furthermore, Parent and Lesage (2008) discuss the hierarchical Bayesian approach for estimating a conditional autoregressive (CAR) spatial model, operating under the assumption that the spatially structured random component of the model follows an autoregressive process.

This paper aims to contribute to the existing literature by providing a comprehensive review of the Bayesian spatial autoregressive modeling framework, highlighting its theoretical foundations, methodological advancements, and practical applications. The comparison between MLE and Bayesian framework in estimating the spatial autoregressive case is also presented.

2. EXPERIMENTAL SECTION

2.1 Moran's Index

In addition to concerns regarding sample size and normality, spatial dependence presents a significant challenge in spatial econometrics. The spatial weight matrix ($\boldsymbol{W}$) serves as a representation of spatial dependence. There are several methods available for constructing this matrix, including distance, contiguity, and geostatistical approaches (Ver Hoef et al., 2018). Defining the matrix incorrectly can result in misleading outcomes. Currently, there is no universally accepted procedure or criteria for identifying the optimal spatial structure, leaving this as an area for further research. In practice, $\boldsymbol{W}$ is typically defined subjectively. To address this, an optimization procedure utilizing the k-Nearest Neighbors (k-NN) method is applied to create a spatial weight matrix that maximizes the spatial autocorrelation coefficient (Jaya et al., 2018). Spatial autocorrelation is assessed using Moran's index, which ranges from -1 to 1; a Moran's index value close to 1 denotes strong spatial autocorrelation. The formula for Moran's index can be written as (Yanuar et al., 2023a);

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{j=1}^{n} (y_j - \bar{y})^2} \quad (1)$$

with i and j are the number of data (n), for $i \neq j$. While y_i is the response variable for all observations i, $\bar{y}$ is the mean of y, w_{ij} is an element of spatial weight matrix $\boldsymbol{W}$. The presence of spatial dependence on the dependent variable is also examined in the hypothesis model using the Lagrange Multiplier Lag (LM_ρ) test. The general structure of the Lagrange Multiplier Lag is presented as follows (Anselin, 2009);

$$LM_\rho = \frac{(u'WY)^2}{[\frac{u'u}{n}]^2 D}, \quad (2)$$

where $D = \left[\frac{(WX\beta)'(I - X(X'X)^{-1}X')(WX\beta)}{\sigma^2}\right] + \mathrm{tr}(\boldsymbol{W'W} + \boldsymbol{WW})$, $\boldsymbol{X}$ is an independent matrix with size $n \times (m+1)$, β is a factor of the regression coefficient of size $m \times 1$. If $LM_\rho > X^2_{(\alpha)1}$, it indicates the presence of spatial dependence on the dependent variable. Therefore, the modeling of data is carried out using a Spatial Autoregressive model (SAR).

2.2 Spatial Autoregressive (SAR)

The Spatial autoregressive (SAR) model is a linear regression model that incorporates spatial correlation among the dependent variables (Dai et al., 2020; Jin et al., 2016). The SAR model can be presented as (Jaya et al., 2018; Lesage, 1999):

$$y_i = \rho \sum_{j=1}^{n} \omega_{ij} y_j + \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i \quad (3)$$

In this context, y_i denotes the response variable for the i-th data point, ρ represents the autoregressive spatial coefficient, and ω_{ij} is an element of the spatial weight matrix derived from the HDI scores across provinces. The parameter β_0 indicates the coefficient for the intercept, while β_k represents the coefficient of the regression slope for the k-th exogenous variable. Moreover, x_{ik} indicates the value of the k-th exogenous variable at the i-th HDI score, and ϵ_i denotes a random error that is assumed to follow an identically independent normal distribution with a mean of zero and a variance of σ^2, written as $\epsilon_i \sim N(0, \sigma^2)$.

2.3 Bayesian Spatial Autoregressive (Bayesian SAR)

The The Bayesian estimation method is particularly effective in cases where the error term is not normally distributed and does not meet the assumptions of homoscedasticity. In comparison to maximum likelihood estimation (MLE), Bayesian estimation generally produces less favorable results when dealing with small sample sizes. Additionally, Bayesian methods are suitable for situations involving spatial heteroscedasticity, where the variance of the error term in the spatial model is unequal. Overall, the structure of the Bayesian SAR model resembles that of the standard SAR model, with the main difference lying in the parameter estimation approach utilized in the Bayesian method. The general form of the Bayesian SAR model is presented in Equation (3) and can also be represented in matrix notation as follows (Dai et al., 2020; Lesage, 1999):

$$y = \rho W y + X\beta + u, \quad \text{with } u \sim N(0, \sigma^2 I), \tag{4}$$

where y is a $(n \times 1)$ vector of the HDI score, ρ is an autoregressive parameter, W is a $(n \times n)$ spatial weight matrix, and X is a $(n \times p)$ design matrix including the unit vector. The Bayesian method assumes that the parameters β and σ^2 are random variables following the normal and inverse gamma distribution, respectively. While ρ is a random variable following the uniform distribution. The formulation for the joint posterior distribution is as follows:

$$p(\beta, \sigma^2, \rho|y, X) = \frac{L(y, X|\beta, \sigma^2, \rho)\pi(\beta|\sigma^2)\pi(\sigma^2)\pi(\rho)}{p(y, X)}. \tag{5}$$

The posterior distribution in Equation (5) is obtained based on the likelihood function and priors as follows:

1. The likelihood function $L(y, X|\beta, \sigma^2, \rho)$:
   It is assumed that the responses follow a normal distribution, thus the likelihood function can be written as:

$$L(y, X|\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-\frac{n}{2}}|A| \exp\left(-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)\right), \quad \text{with } |A| = (I - \rho W). \tag{6}$$

2. The prior distribution for $\beta|\sigma^2$, or $f(\beta|\sigma^2)$, follows a normal distribution:

$$f(\beta|\sigma^2) = \frac{1}{(2\pi)^{p/2}(\sigma^2)^{p/2}|T|^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - s)'T^{-1}(\beta - s)\right), \quad \beta|\sigma^2 \sim N(s, \sigma^2 T). \tag{7}$$

3. The prior distribution for σ^2, or $f(\sigma^2)$, follows an inverse gamma distribution:

$$f(\sigma^2) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right), \quad \sigma^2 \sim IG(a, b). \tag{8}$$

4. The prior distribution for ρ, or $f(\rho)$, follows a uniform distribution:

$$f(\rho) = \frac{1}{\lambda_{\max}^{-1} - \lambda_{\min}^{-1}}, \quad \rho \sim \text{UNIF}(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}). \tag{9}$$

5. The joint prior distribution for β and σ^2, or $f(\beta, \sigma^2)$, follows a normal-inverse gamma distribution:

$$f(\beta, \sigma^2) = f(\beta|\sigma^2) \times f(\sigma^2), \tag{10}$$

   and LeSage and Pace (2009) proved:

$$\beta, \sigma^2 \sim N(s, \sigma^2 T) \times IG(a, b) = NIG(s, T, a, b).$$

The likelihood function and all prior distributions obtained above are used to construct the joint posterior distribution as follows:

$$p(\beta, \rho, \sigma^2|y, X) = \frac{L(y, X|\beta, \sigma^2, \rho)f(\beta, \sigma^2)f(\rho)}{f(y, X)}. \tag{11}$$

By multiplying the expressions for the likelihood and the prior, we can identify the form of the posterior distribution, omitting a constant term $p(D)$ that does not depend on the model parameters. Therefore, the posterior distribution is proportional to the product of the likelihood function and the prior distribution.

$$\begin{aligned} p(\beta, \rho, \sigma^2|y, X) &\propto L(y, X|\beta, \sigma^2, \rho)f(\beta, \sigma^2)f(\rho) \\ &= L(y, X|\beta, \sigma^2, \rho)f(\beta|\sigma^2)f(\sigma^2)f(\rho) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}|A| \exp\left(-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)\right) \times \\ &\quad \frac{1}{(2\pi)^{p/2}(\sigma^2)^{p/2}|T|^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - s)'T^{-1}(\beta - s)\right) \\ &\quad \times \frac{b^a}{\Gamma(a)}(\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right) \times \frac{1}{\lambda_{\max}^{-1} - \lambda_{\min}^{-1}}. \\ &= (\sigma^2)^{-(a+\frac{n+p}{2}+1)}|A| \exp\Bigg(-\frac{1}{2\sigma^2}\Big[(Ay - X\beta)' \\ &\quad (Ay - X\beta) + (\beta - s)'T^{-1}(\beta - s) + 2b\Big]\Bigg). \end{aligned} \tag{12}$$

The joint posterior distribution above can be written as follows:

$$p(\beta, \rho, \sigma^2|y, X) = (\sigma^2)^{-(a^*+1)}|A| \exp\Bigg(-\frac{1}{2\sigma^2}\Big[2b^*$$

$$+(\beta - s^*)'(T^*)^{-1}(\beta - s^*)\Big]\Big),$$

with

$$T^* = (X'X + T^{-1})^{-1},$$
$$s^* = (X'X + T^{-1})^{-1}(X'Ay + T^{-1}s),$$
$$b^* = b + \frac{(s'T^{-1}s + y'A'Ay) - (s^*)'(T^*)^{-1}s^*}{2},$$
$$a^* = a + \frac{n+p}{2}.$$

The marginal posterior distribution for each estimated parameter, β, ρ and σ^2 are determined as follows:

1. Marginal posterior distribution for $(\beta|\rho, \sigma^2)$:
$$p(\beta|\rho, \sigma^2) \sim N(s^*, \sigma^2 T^*).$$
2. Marginal posterior distribution for $(\sigma^2|\beta, \rho)$:
$$p(\sigma^2|\beta, \rho) \sim IG(a^*, b^*).$$
3. Marginal posterior distribution for $\rho|\beta, \sigma^2$:
$$f(\rho|\beta, \sigma^2) = \frac{f(\rho, \beta, \sigma^2|y, X)}{f(\beta, \sigma^2|y, X)}$$
$$\propto f(\rho, \beta, \sigma^2|y, X)$$
$$\propto |I_n - \rho W| \exp\left(-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)\right). \quad (13)$$

This conditional distribution of $\rho|\beta\sigma^2$ does not follow any form of standard distribution. The exact values of parameter estimates cannot be obtained from this distribution. Therefore, the estimation is done by numerical procedure by using the MCMC approach, which is a combination of the Gibbs sampling and the metropolis methods (Ntzoufras, 2009).

2.4 Convergence Diagnostic Test

The collection of MCMC samples extracted from the true posterior distribution forms the basis for parameter inference in Bayesian methods. To evaluate the reliability of this inference, a convergence diagnostic is employed to ascertain the minimum sample size necessary for accurately approximating the target posterior density. Common graphical techniques used for assessing convergence include trace plots, ergodic mean plots, and autocorrelation plots. The stabilization of these plots after a certain number of iterations indicates that the algorithm has successfully converged.

3. RESULTS AND DISCUSSION

This study uses secondary data on the Human Development Index (HDI) in regencies/cities in East Java in 2022. Data was obtained from the website of the Central Statistics Agency of East Java Province. The scope of the study was limited to 38 regencies/cities located in East Java province. The distribution of HDI score for each city in East Java is shown in Figure 1. The independent variables used in this study are proper sanitation (X_1), Average Length of Schooling (X_2), Poverty (X_3), Open Unemployment Rate (X_4), Life Expectancy (X_5), and Per Capita Expenditure (X_6). In addition, this study also used a dummy variable (D). The dummy variable was selected based on the existence of medium to large-scale industrial areas in regencies/cities in East Java Province. If in a region there is a medium to large-scale industry, a value of 1 is given to the dummy variable ($D = 1$), or $D = 0$, else. Table 1 shows the descriptive analysis of data.

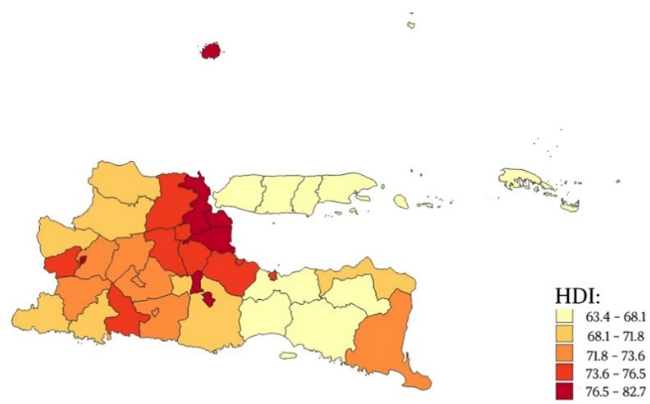

Figure 1. The Distribution of the HDI Score for Each City in East Java

Statistics of the HDI's Score in East Java are presented in Table 1. The mean of HDI's score (Y), proper sanitation (X_1), and life expectancy (X_5) are more than 60%, while average length of schooling (X_2), poverty (X_3), and open unemployment rate (X_4) are less than 11%. There are Five cities that have medium to large-scale manufacturing industry companies in East Java, i.e., Gresik, Mojokerto City, Pasuruan, Sidoarjo, and Surabaya City. Thus, the dummy variable is used to mention the availability of this manufacturing. We coded 1 for each five cities and 0 for the others. In this study, the spatial weight matrix is constructed using the nearest neighbor method, known as the k-Nearest Neighbor (k-NN) spatial weighting matrix. This k-NN spatial weight matrix is created by calculating the Euclidean distance between a designated area and all other areas. The resulting distances are then organized in order of proximity. This measurement process is repeated for other areas as well. Subsequently, we simulate four different values of $k (k = 2, 3, 4, 5)$ for the nearest neighbors and determine the optimal k based on the highest value of Moran's index statistics.

Let choose two ($k = 2$) nearest neighbors to construct the weight matrix C with the value of each element, $c_{ij} = 1$, if i and j are neighbors or $c_{ij} = 0$, if not. Table 2 presents the two selected nearest distances. Then the standardization of the weight matrix is obtained by dividing each element to the number of rows to produce a spatial weight matrix ($\boldsymbol{W}$), with

Table 1. Statistics of the HDI's Score in East Java

Variable	$Q1$	Mean	Median	$Q3$
HDI (Y)	69.440	72.790	72.440	74.880
Proper sanitation (X_1)	78.370	82.160	83.930	90.570
Average Length of Schooling (X_2)	7.915	8.246	7.915	9.545
Poverty (X_3)	6.840	10.170	9.6350	12.450
Open Unemployment Rate (X_4)	4.343	5.229	5.315	6.338
Life Expectancy (X_5)	70.780	72.110	72.860	73.180
Per Capita Expenditure (X_6)	10278	11835	11571	12979

Table 2. The Two Nearest Neighbor (k = 2)

Area	Neighbors Areas
Bangkalan	Surabaya City and Sampang
Banyuwangi	Bondowoso and Jember
Batu City	Malang and Malang City
Blitar	Kediri and Blitar City
Bojonegoro	Nganjuk and Tuban
Bondowoso	Jember and Situbondo
Gresik	Lamongan and Surabaya City
Jember	Bondowoso and Probolinggo
Jombang	Mojokerto and Mojokerto City
Kediri	Kediri City and Nganjuk
Blitar City	Blitar and Kediri City
Kediri City	Kediri and Nganjuk
Madiun City	Magetan and Madiun
Malang City	Batu City and Malang
Mojokerto City	Jobang and Mojokerto
Pasuruan City	Pasuruan and Sidoarjo
Probolinggo City	Pasuruan City and Probolinggo
Lamongan	Gresik and Mojokerto City
Lumajang	Probolinggo City and Probolinggo
Madiun	Madiun City and Magetan
Magetan	Madiun City and Ngawi
Malang	Batu City and Malang City
Mojokerto	Batu City and Mojokerto City
Nganjuk	Kediri and Kediri City
Ngawi	Madiun City and Magetan
Pacitan	Ponorogo and Trenggalek
Pamekasan	Bangkalan and Sampang
Pasuruan	Malang City and Pasuruan City
Ponorogo	Magetan and Trenggalek
Probolinggo	Probolinggo City and Lumajang
Sampang	Bangkalan and Pamekasann
Sidoarjo	Mojokerto and Surabaya City
Situbondo	Banyuwangi and Bondowoso
Sumenep	Pamekasan and Situbondo
Surabaya	Bangkalan and Sidoarjo
Trenggalek	Ponorogo and Tulungagung
Tuban	Bojonegoro and Lamongan
Tulungagung	Blitar City and Trenggalek

$w_{ij} = \frac{c_{ij}}{k}$, or 0 for others. This spatial weight matrix is then used to calculate the Moran's Index statistics. Using Equation (1), we got that the Moran's index is 0.5103 (p-value is 0.0065). The same steps are done for k = 3, 4, and 5 selected nearest neighbors. The optimal value of k is the one that yields the highest Moran's index. The simulation results are illustrated in Figure 2 below, which indicates that the maximum Moran's index value is achieved when k = 2. The analysis did not extend to k = 6, 7, and beyond, as higher values of k typically lead to a decrease in the Moran's index. The peak Moran's index value facilitates the creation of the optimal spatial weight matrix ($\boldsymbol{W}$). Hence, in this study, we select k = 2, as it produces the highest Moran's I value according to the k-NN method for generating the optimal spatial weight matrix ($\boldsymbol{W}$).

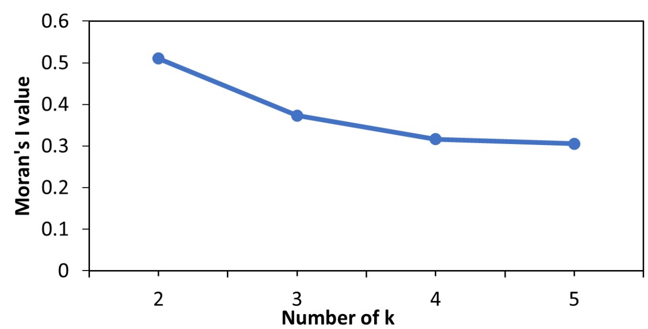

Figure 2. Moran's Index Value for k = 2, 3, 4, 5

Moran's Index in this context is positive, indicating the presence of spatial dependence in the dependent variable. Preliminary studies revealed a spatial effect characterized by this spatial dependence. Consequently, the SAR model was employed to derive an appropriate model. Table 3 displays the results of parameter estimation derived from the SAR model using both Maximum Likelihood Estimation (MLE) and Bayesian estimation methods.

Based on Table 3, it is found that not all independent variables have a significant influence on HDI in East Java. We include the significant independent variables in the model, which are Poverty (X_3), Per Capita Expenditure (X_6), and Dummy (D). The result of the estimation model is presented in Table 4.

This study revealed that all independent variables signifi-

Table 3. Estimation Model Results with the SAR Model

Variable	MLE SAR		Bayesian SAR	
	Estimation Means	Standard Error	Estimation Means	Standard Error
Intercept	25.5175*	12.7570	25.3900*	15.1704
Proper sanitation (X_1)	0.0123	0.0336	0.0123	0.2989
Average Length of Schooling (X_2)	0.0144	0.4069	0.0154	0.4883
Poverty (X_3)	-0.2211*	0.0947	-0.2201*	0.1143
Open Unemployment Rate (X_4)	-0.0558	0.1900	-0.0541	0.2208
Life Expectancy (X_5)	0.2668	0.1875	0.2704	0.2247
Per Capita Expenditure (X_6)	0.0011*	0.0001	0.0011*	0.0002
Dummy (D)	5.8179*	0.7759	2.3870*	0.9339
Coefficient SAR (ρ)	0.2099*	0.0857	0.2077*	0.0912

*Significant at $\alpha = 0, 1, |Z\alpha/2| = 1, 645$

Table 4. Estimation Model Results Part 2

Variable	MLE SAR		Bayesian SAR	
	Estimation Means	Standard Error	Estimation Means	Standard Error
Intercept	42.2773*	5.7660	42.4740*	5.7104
Poverty (X_3)	-0.2765*	0.0788	-0.2768*	0.0879
Per Capita Expenditure (X_6)	0.0011*	0.0001	0.0011*	0.0001
Dummy (D)	2.2792*	0.7951	2.2860*	0.8843
Coefficient SAR (ρ)	0.2567*	0.0748	0.2538*	0.0691
R^2	0.9023		0.9037	
MAE	1.1187		1.1165	
RMSE	1.4543		1.4537	

*Significant at $\alpha = 0.1, |Z_{\alpha/2}| = 1.645$

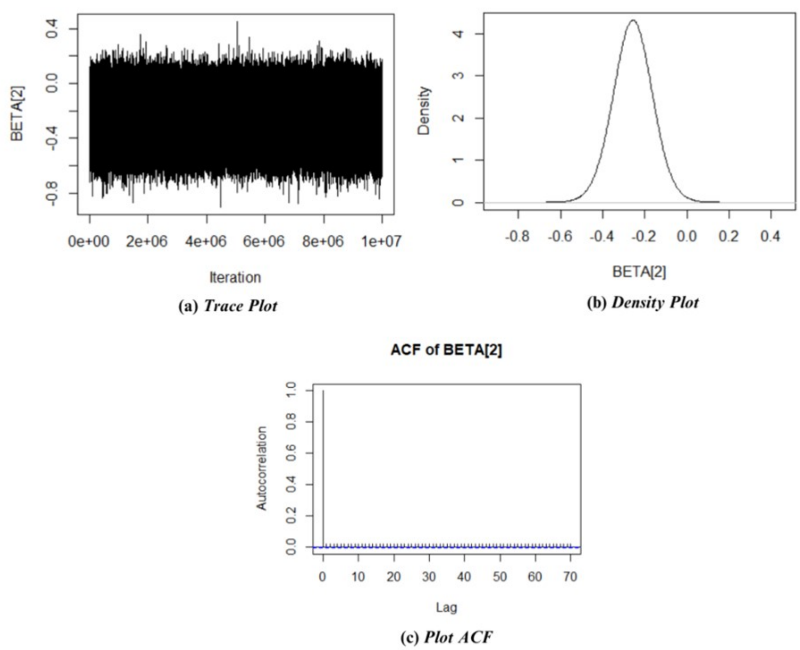

Figure 3. Convergence Test for Coefficient of Poverty Trace Plot. (b). Density plot. (c). Plot ACF

cantly influence the HDI rate $\alpha = 0.1$, with consistent results from both MLE and Bayesian approaches. To determine the superior model, a comparison between the two methods is necessary. The findings indicate that the Bayesian SAR model outperforms the MLE model, as it yields a higher value of R^2 along with lower values of MAE and RMSE. The next step in the analysis involves conducting a convergence test for the estimated parameters derived from the Bayesian SAR model. The convergence test in the Bayesian method is performed by analyzing the trace plot after running the Gibbs sampler for 50,000 iterations, including an initial burn-in period of 5,000 iterations. Figure 2 illustrates the trace plot and the density plot for the selected parameter, "Poverty."

Based on the trace plot in Figure 3(a), it can be seen that the distribution of selected parameter lies within two parallel horizontal lines. This indicated that the parameters has converged. Additionally, Figure 3(b) demonstrates how symmetrical the density plot generated for each parameter estimate is suggesting that the predicted values of the model's parameters follow a normal distribution curve. Regarding Figure 3(c). the autocorrelation value gradually approaches zero as the lag increases as shown by the ACF plot. It indicates that the estimated value is generated in the direction of stability and eventually converges to a value at which it is determined to be acceptable. Therefore, the proposed model for the HDI rate based on the Bayesian SAR model is accepted:

$$\hat{y} = 0.2538\boldsymbol{Wy} + 42.4740 - 0.2768X_3 + 0.0011X_6 + 2.2860D \tag{14}$$

4. CONCLUSIONS

The model optimization performed using the k-NN method identified that the appropriate spatial weight matrix ($\boldsymbol{W}$) occurs when k equals 2. This suggests a strong spatial autocorrelation among the two nearest neighbors, indicating that the Human Development Index (HDI) rates of adjacent areas are quite similar. Due to the violation of the normality assumption in the Maximum Likelihood Estimation (MLE) SAR approach, the estimation and inference of the spatial autoregressive model were conducted using the mean Bayesian SAR method. An empirical analysis of the HDI scores for 38 regencies and cities in East Java revealed that significant factors influencing HDI include poverty, per capita expenditure, and the presence of both large-scale and medium-scale industrial areas in these regions. The parameter estimates obtained from both the MLE and Bayesian approaches yielded similar results. This study demonstrated that the Bayesian SAR method is superior to the MLE SAR method in modeling HDI rates in East Java, as it provides higher R^2 estimates along with lower values for Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in comparison to the MLE SAR method.

5. ACKNOWLEDGMENT

This research was funded by LPPM, Universitas Andalas under Penelitian Skripsi Sarjana (PSS) fund with Contract Number 156/UN16.19/PT.01.03/PSS/2024.

REFERENCES

Anselin, L. (1988). The Scope of Spatial Econometrics. In L. Anselin, editor, *Spatial Econometrics: Methods and Models*, volume 4. Springer Netherlands, pages 7–15

Anselin, L. (2009). Spatial Regression. In A. S. Fotheringham and P. A. Regerson, editors, *The Sage Handbook of Spatial Analysis*. SAGE Publications Ltd, pages 225–276

Badan Pusat Statistik Provinsi Jawa Timur (2023). Indeks Pembangunan Manusia Provinsi Jawa Timur 2022. Badan Pusat Statistik Provinsi Jawa Timur

Caniago, M. A. I. and M. G. Wibowo (2024). Determinants of Human Development Index in Indonesia with Maqashid Sharia Approach. *Jurnal Ilmiah Ekonomi Islam*, **10**(1); 200

Dai, X. and L. Jin (2021). Minimum Distance Quantile Regression for Spatial Autoregressive Panel Data Models with Fixed Effects. *PLoS One*, **16**(12); e0261144

Dai, X., Z. Yan, M. Tian, and M. Tang (2020). Quantile Regression for General Spatial Panel Data Models with Fixed Effects. *Journal of Applied Statistics*, **47**(1); 45–60

Hepple, L. W. (1995). Bayesian Techniques in Spatial and Network Econometrics: 2. Computational Methods and Algorithms. *Environment and Planning A: Economy and Space*, **27**(4); 615–644

Jaya, I. G. N. M., T. Toharudin, and A. S. Abdullah (2018). A Bayesian Spatial Autoregressive Model with k-NN Optimization for Modeling the Learning Outcome of the Junior High Schools in West Java. *Model Assisted Statistics and Applications*, **13**(3); 207–219

Jin, L., X. Dai, A. Shi, and L. Shi (2016). Detection of Outliers in Mixed Regressive-Spatial Autoregressive Models. *Communications in Statistics - Theory and Methods*, **45**(17); 5179–5192

Kakamu, K. and H. Wago (2008). Small-Sample Properties of Panel Spatial Autoregressive Models: Comparison of the Bayesian and Maximum Likelihood Methods. *Spatial Economic Analysis*, **3**(3); 305–319

Lesage, J. P. (1997). Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review*, **20**(1-2); 113–129

Lesage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*. Department of Economics University of Toledo

LeSage, J. P. and R. K. Pace (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. John Wiley and Sons

Ogujiuba, K., L. Maponya, and N. Stiegler (2024). Determinants of Human Development Index in South Africa: A Comparative Analysis of Different Time Periods. *World*, **5**(3); 527–550

Oliveira, V. D. and J. J. Song (2008). Bayesian Analysis of Simultaneous Autoregressive Models. *Sankhya: The Indian Journal of Statistics, Series B (2008-)*, **70**(2); 323–350

Parent, O. and J. P. Lesage (2008). Using the Variance Structure of the Conditional Autoregressive Spatial Specification to Model Knowledge Spillovers. *Journal of Applied Econometrics*, **23**(2); 235–256

Tyas, D. P. P. and N. M. Sukartini (2022). Determinants of the Human Development Index (HDI) in Indonesia, 2014–2021. *Media Trend*, **17**(2); 481–495

Ver Hoef, J. M., E. E. Peterson, M. B. Hooten, E. M. Hanks, and M.-J. Fortin (2018). Spatial Autoregressive Models for Statistical Inference from Ecological Data. *Ecological Monographs*, **88**(1); 36–59

Yanuar, F., T. Abrari, and I. R. Hg (2023a). The Construction of Unemployment Rate Model Using SAR, Quantile Regression, and SARQR Model. *Pakistan Journal of Statistics and Operation Research*, **19**(3); 447–458

Yanuar, F., T. Abrari, I. Rahmi HG, and A. Zetra (2023b). Spatial Autoregressive Quantile Regression with Application on Open Unemployment Data. *Science and Technology Indonesia*, **8**(2); 321–329

Yasin, H., A. R. Hakim, and B. Warsito (2020). Development Life Expectancy Model in Central Java Using Robust Spatial Regression with M-Estimators. *Communications in Mathematical Biology and Neuroscience*, **69**; 1–16

Yasin, H., B. Warsito, A. R. Hakim, and R. N. Azizah (2022). Life Expectancy Modeling Using Modified Spatial Autoregressive Model. *Media Statistika*, **15**(1); 72–82

Yu, T., F. Gao, X. Liu, and J. Tang (2022). A Spatial Autoregressive Quantile Regression to Examine Quantile Effects of Regional Factors on Crash Rates. *Sensors*, **22**(5); 1–15

Zhang, J., Q. Lu, L. Guan, and X. Wang (2021a). Analysis of Factors Influencing Energy Efficiency Based on Spatial Quantile Autoregression: Evidence from the Panel Data in China. *Energies*, **14**(2); 504

Zhang, Y., J. Jiang, and Y. Feng (2021b). Penalized Quantile Regression for Spatial Panel Data with Fixed Effects. *Communications in Statistics - Theory and Methods*; 1–13