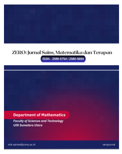
Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.30392
Page: 780-792



Spatial Determinants of Stunting in East Java: A Comparative Assessment between OLS and Spatial Models Approach

[1]Mieke Nurmalasari
Health Information Management Department, Universitas Esa Unggul, Jakarta, Indonesia

[2]Dede Yoga Paramartha
Directorate of Statistical Analysis and Satellite Accounts, BPS-Statistics, Jakarta, Indonesia

[3]Nandya Rezky Utami
Directorate of Statistical Methodology and Data Science, BPS-Statistics, Jakarta, Indonesia

[4]Dinda Fahrani
Politeknik Statistika STIS, Jakarta, Indonesia

[5]Andrey Sinlay
Health Information Management Department, Universitas Esa Unggul, Jakarta, Indonesia

[6]Tria Saras Pertiwi
Health Information Management Department, Universitas Esa Unggul, Jakarta, Indonesia

[7]Setia Pramana
Politeknik Statistika STIS, Jakarta, Indonesia

Article Info

Article history:

Accepted: 03 August 2026

Keywords:

OLS;
SAR;
SEM;
Spatial Analysis;
Stunting.

ABSTRACT

Stunting remains a critical challenge in Indonesia, aligning with the 2030 SDGs. This study examines the spatial patterns of stunting prevalence across 29 districts and 9 cities in East Java Province using 2022 data. Classic Ordinary Least Squares (OLS), Spatial Autoregressive (SAR), and Spatial Error Models (SEM) were deployed. Baseline OLS shows that Gender Development Index, Poverty, GRDP per Capita, Access to Adequate Sanitation, and nurse density simultaneously exert a significant joint effect on stunting. Although Moran's I indicated marginal evidence of positive spatial autocorrelation in the OLS residuals ($p = 0.0556$), the SAR and SEM specifications yielded non-significant spatial parameters (ρ and λ) and no meaningful AIC improvement over OLS. Local Indicators of Spatial Association (LISA) further show that, unlike sanitation access, stunting itself does not form a statistically significant High-High cluster, suggesting that the observed residual spatial dependence may be partly accounted for by the included structural covariates, although the study's small sample size ($n = 38$) may also limit the power to detect spatial effects directly. Consequently, the classical OLS specification was retained as the preferred model for inference, with SAR and SEM results reported as spatial diagnostics.

Corresponding Author:

Mieke Nurmalasari,
Health Information Management Department, Faculty of Health Sciences,
Universitas Esa Unggul, Jakarta, Indonesia
Email: mieke@esaunggul.ac.id

1. INTRODUCTION

Stunting continues to be a significant nutritional issue globally, particularly in developing nations such as Indonesia. It refers to a condition wherein children under the age of five experience impaired growth, resulting in shorter stature and height compared to their age-appropriate standards [1]. This measurement uses the child growth standards established by WHO, which involves interpreting stunting if it falls more than two standard deviations below the median [2]. Stunted toddlers have problems related to nutrition caused by factors such as the mother's nutritional status during pregnancy, the social and economic conditions of the toddler's family, the baby's health condition, and the toddler's minimal dietary intake [3]. A comprehensive review by [4] on the World Health Organization conceptual framework, identified consistent evidence that child stunting in Indonesia is driven by a combination of individual, household, and community-level factors, including non-exclusive breastfeeding, low socioeconomic status, poor maternal education, and inadequate water and sanitation, as well as limited access to health care.

In addition, the problem of stunting is also influenced by environmental factors and geographical conditions such as population density, climate and inadequate sanitation [5]. Stunting is the focus of attention because it affects children's growth, especially the risk of impaired physical and cognitive development. The short-term effects of stunting are increased morbidity and mortality rates and insufficient cognitive, motoric and verbal development. While the long-term effects of stunting include: toddlers' body posture that is not optimal, triggering the risk of obesity and non-communicable diseases, learning capacity is not optimal, productivity and work capacity are not optimal and inhibited economic growth [6].

The Sustainable Development Goals (SDGs) are a world sustainable development program initiated by WHO to prosper the world's people and preserve nature through 17 main goals with 169 achievement targets by 2030. Stunting in children is part of the Sustainable Development Goals (SDGs) framework. 2nd to end hunger, achieve food security and better nutrition and support sustainable agriculture by 2025 [7]. The target set by WHO is to reduce stunting rates by 40% in children under five years [8]. Various efforts to control stunting to reduce the stunting rate worldwide significantly have been carried out through activities to increase per capita GDP growth, increase food availability, and improve health, education, and empower women [9]

Based on the 2022 Indonesian Nutritional Status Survey (SSGI) results, Indonesia's stunting rate decreased from 24.4% in 2021 to 21.6% in 2022. The Province of East Nusa Tenggara has the highest prevalence of stunting at 35.3%, while Bali occupies the province with the lowest prevalence of stunting, namely 8.0% [10]. However, the reduction in the stunting rate is still far from the Indonesian government's target, namely a stunting prevalence of 14% in 2024. Based on the Presidential Regulation of the Republic of Indonesia number 72 of 2021 concerning accelerating the reduction of stunting, the Indonesian government is targeting the achievement of stunting prevalence to decrease to 14% in 2024. This acceleration is being pursued through comprehensive intervention efforts, including specific and sensitive interventions, implemented collaboratively, holistically, integrated, and in a high-quality manner with multiple stakeholders' involvement [11]. The government has established a national strategy to accelerate the reduction of stunting to achieve the target of sustainable development goals in 2030. A decrease of 3.8% per year is required to achieve the target of 14% in 2024 [10].

East Java Province is one of the provinces with a prevalence of stunting that tends to decrease yearly. In 2022, East Java will experience a reduction in % the stunting rate of 23.5% in 2021 to 19.2% in 2022 [10]. However, since 2018 the province of East Java has become one of the priority areas for handling stunting in Indonesia [1]. It is related to the problem of extreme poverty in the province of East Java. The poverty that occurs is reciprocal, where poverty can cause malnutrition, and malnourished toddlers encourage the process of poverty due to loss of work productivity caused by decreased cognitive function [12]. Lack of nutrition contributes to poverty indicators for poor households in East Java where a lack of protein intake in the long term can exacerbate the incidence of stunting in toddlers [13]. In addition, the problem of stunting in East Java is also related to the high incidence of stunting inequality in certain districts/cities, causing spatial effects due to one area affecting other regions due to geographical factors [1].

The proximity of districts/cities in the province of East Java provides opportunities for one region to another to have the same problem. This is also related to the inequality of stunting events, where the spatial effect of an area that has a high percentage of stunting, the surrounding area will experience the same thing [14]. Spatial diversity can occur due to the influence of differences in regional characteristics and geographical location in each observation area. The factors that cause stunting in each region are different due to differences in the characteristics of an area and the relationship between the distances between regions [1].

Numerous spatial studies on stunting have heavily relied on specific, isolated dimensions. For instance, initial diagnostic approaches often evaluate crude spatial dependency through the global Moran's I index without adjusting for neighboring exogenous covariates, which frequently yields statistically insignificant results [15]. Furthermore, existing literature has predominantly focused on individual-level health data [16], [17], [18] or as opposed to isolated socioeconomic metrics [19]. This study addresses these empirical gaps by comprehensively integrating multi-dimensional determinants, specifically bridging socioeconomic vulnerabilities, gender development, and localized healthcare infrastructure into a unified spatial econometric framework. By doing so, this research offers a novel contribution by evaluating how these macro-level structural factors simultaneously interact to drive the geographical prevalence of stunting across all regencies and cities in East Java Province. Grounded in this empirical backdrop, the primary objective of this study is to examine the spatial effects of stunting prevalence in East Java by evaluating spatial econometric frameworks (including their local spatial clustering structure) against a baseline classical regression model.

2. RESEARCH METHOD

2.1 Data

This research focuses on spatial analysis of the prevalence of stunting toddlers in East Java in 2022. The research analysis unit consisted of 29 districts and 9 cities in East Java Province. The data used in this research secondary data come from the Indonesian Nutrition Status Study (SSGI), which was obtained from the Ministry of Health of the Republic of Indonesia, the Central Statistics Agency and data collected from the results of Regency/Municipality Potential Data Collection (PODES). The research variables consist of the dependent variable, stunting prevalence, and five independent variables to explain the occurrence of stunting as shown in Table 1.

Table 1. Variables

Variables	Data Source
Y = Stunting Prevalence	Indonesian Nutrition Status Survey (SSGI) Ministry of Health
X_1 = Gender Development Index	BPS-Statistics Indonesia
X_2 = Poverty rate	BPS-Statistics Indonesia
X_3 = GRDP per Capita	BPS-Statistics Indonesia
X_4 = Access to Adequate Sanitation	PODES BPS-Statistics Indonesia
X_5 = Number of Nurse per 10,000 Population	PODES BPS-Statistics Indonesia

2.2 Data Analysis

Descriptive Statistics

The descriptive analysis began by classifying all utilized variables, allowing for a subsequent examination of their spatial patterns. The classification was determined using the Natural Breaks method. The Natural Breaks method is a common technique in spatial data analysis. It groups data by identifying "natural breakpoints" in the distribution of values [20]. In other words, this method minimizes variance within each class (within class variance) while maximizing the differences between classes (between class variance).

Multiple Linear Regression Model

The regression analysis begins with the construction of a global regression using multiple linear regression. Following the formation of the global regression model, assumptions are tested, including the examination of spatial effects.

The regression model is a formal representation that explains the relationship between two or more numerical variables, enabling the prediction of the dependent variable from the independent variables [21]. The equation for the linear regression model is expressed as Equation (1).

$$y_i = \beta_0 + \sum_{k=1}^{n} \beta_k x_{ik} + \varepsilon_i \tag{1}$$

where y_i is the dependent variable at the i^{th} observation, β_0 is the intercept, β_k is the coefficient of the k^{th} independent variable, x_{ik} is the value of the k^{th} independent variable at the i^{th} observation, p is the number of independent variables, and ε_i is the random error of the i^{th} observation. In order to provide parameter estimates that are BLUE (Best Linear Unbiased Estimation), the regression modeling must satisfy certain assumptions derived in the Gauss-Markov theorem [21]. The null hypothesis (H_0) stated that all regression coefficients of the independent variables are equal to zero, indicating no significant effect on stunting prevalence. The alternative hypothesis (H_1) stated that at least one coefficient is statistically different from zero. Hypothesis testing was conducted using t-tests for individual regression coefficients and an F-test for overall model significance. The decision rule was to reject H_0 when the p-value was less than α=0.10.

Spatial Analysis

Spatial Effects

In a regression context, spatial effects pertain to two categories of specifications. One deals with spatial dependence, or its weaker expression, spatial autocorrelation, and the other with spatial heterogeneity [22].

Moran's I

The most widely used test for spatial autocorrelation is Moran's I, introduced by Moran (1948) as a two-dimensional version of a time-series correlation test. In matrix form, Moran's I is [22]:

$$I = (e'We/e'e) \tag{2}$$

where *e* represents OLS residuals matrix and *W* represents weight matrix. The spatial weighting matrix W reflects the relationship between areas based on distance or adjacency information (contiguity) and expressed as follow [23]:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & w_{ij} & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \tag{3}$$

The dimension of this matrix is n × n, where n represents the number of locations, zones, or cross-object units being analyzed. This study utilizes a queen contiguity matrix to define spatial neighborhoods based on shared edges and corners. A spatial weight of $w_{ij} = 1$ indicates that locations i and j share a common boundary and/or vertex, whereas $w_{ij} = 0$ signifies that no such spatial contact exists between them [23].

The Moran's I test is similar to the Durbin-Watson statistic and has shown superior power in simulations [24]. Moran's, I range from -1 to 1, with exact bounds determined by the weight matrix. A positive value indicates clustering of similar high or low values, while a negative value signals contrasting values in neighboring areas. Under the null hypothesis (H_0: no spatial autocorrelation, I=0). Rejecting H_0 implies spatial autocorrelation [24].

If the result of the Moran's I test is not significant, the analysis proceeds with either the global regression model or multiple linear regression. However, if the result is significant, the analysis continues with a spatial model. Subsequently, spatial heterogeneity is examined using the Breusch-Pagan test. If there is no spatial heterogeneity then the Spatial Autoregressive (SAR) model and the Spatial Error Model (SEM) can be applied.

Spatial Autoregressive (SAR)

Given the presence of significant spatial dependence, the analysis proceeded with spatial regression models, namely the Spatial Autoregressive (SAR) and Spatial Error Model (SEM). For the SAR model, the null hypothesis (H_0) stated that the spatial autoregressive coefficient (ρ) equals zero, implying that stunting prevalence in a district is not influenced by neighboring districts. The alternative hypothesis (H_1) stated that $\rho \neq 0$, indicating the existence of spatial spillover effects. The significance of ρ was tested using Wald statistics or likelihood-based inference, and H_0 was rejected when the p-value was below 0.05. The SAR model is appropriate when spatial interaction occurs directly through the dependent variable [22], [25].

Formally, a spatial lag model, or a mixed regressive, spatial autoregressive model is expressed as [22] :

$$Y = \beta_0 + X\beta + \rho WY + \varepsilon \tag{4}$$

where ρ is the spatial autoregressive coefficient and ε is the vector of error terms; Y represents the matrix of dependent variable, X is the matrix of independent variables, W denotes the spatial weight matrix, and ρ serves as the spatial lag coefficient in the model.

Spatial Error Model (SEM)

Other specifications that rely on the spatial autoregressive process can be used to produce spatial regression models that exhibit: spatial dependence in the disturbance process ε which leads to a spatial error model (SEM) [25] :

$$Y = X\beta + u \tag{5}$$

$$u = \lambda Wu + \varepsilon \tag{6}$$

where Y represents the matrix of dependent variable, X is the matrix of independent variables, W denotes the spatial weight matrix, λ is the spatial lag coefficient parameter for the error term, u and ε are error vectors of dimension $n \times 1$, and β is the regression coefficient vector.

Model Comparison

Model estimation of OLS, SEM, and SAR are compared to identify the most appropriate spatial specification. The comparison was made by evaluating the Akaike Information Criterion (AIC) values for both models, where the model with the lower AIC is considered to have a better fit.

Note that the analysis included 38 spatial units, consisting of 29 districts and 9 cities in East Java Province. The relatively small number of spatial units may limit the statistical power to detect spatial dependence.

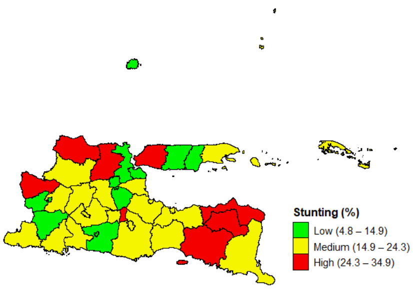
3. RESULT AND DISCUSSION

3.1 Overview of the Prevalence of Stunting and Variables in Toddlers Allegedly Influenced

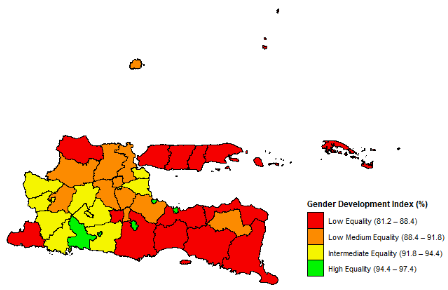
Collectively, the spatial visualizations in Figure 1 reveal a visually clustered pattern across the 29 districts and 9 cities in East Java Province, with higher stunting prevalence appearing concentrated in a number of geographically contiguous areas. At this descriptive stage, such a pattern is consistent with either a genuine spatial process, for example geographic proximity or regional spillover effects, or with spatial clustering of the underlying structural covariates themselves, namely poverty, the Gender Development Index and sanitation access. The formal tests reported in Sections 3.3 to 3.6 are used to distinguish between these two explanations and, as shown below, the evidence favors the latter, that is, spatial clustering of the covariates rather than an independent spatial process in stunting itself. The descriptive account that follows should therefore be read as an exploratory description of co-located patterns rather than as evidence of causal or spillover mechanisms. It should further be noted that, given the study's limited sample size (n = 38), the tests for spatial autocorrelation and for the spatial model parameters reported below carry limited statistical power, so that non-significant findings may reflect either the genuine absence of spatial dependence or insufficient power to detect it at this sample size, and should be interpreted with corresponding caution.

The spatial distribution of stunting prevalence (Figure 1a) exhibits a solid, high-prevalence cluster along the northern coast and the eastern peninsula (the Tapal Kuda region). Crucially, this stunting hotspot directly mirrors the spatial concentration of poverty (Figure 1c), particularly on Madura Island and its peripheral archipelagos. This visual overlap confirms that poverty acts as a primary structural driver of stunting, where household financial distress in geographically isolated areas directly translates into food insecurity, poor maternal nutrition, and an inability to afford nutrient-dense food for child development.

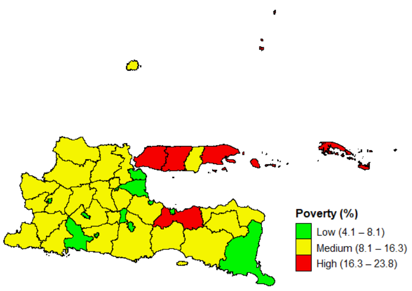
Meanwhile, Gender inequality, measured by the Gender Development Index (GDI) in Figure 1b, is pervasive across nearly all rural and coastal districts that suffer from high stunting rates. This widespread disparity indicates that limited women's empowerment—particularly in health-related decision-making and financial autonomy directly compromises child caregiving practices. This vulnerability is further exacerbated by the distribution of GRDP per capita (Figure 1d), which reveals economic stagnation across the same high stunting zones. While the metropolitan core of Surabaya and its industrial satellites (Sidoarjo and Gresik) leverage high economic output to maintain low stunting rates, the rural periphery is trapped in a low-income cycle that severely restricts child nutrition.



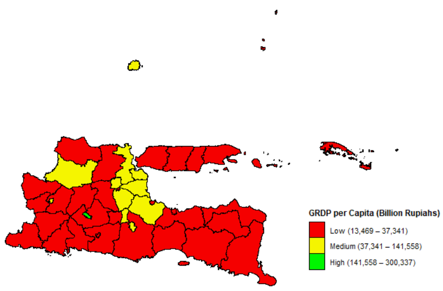
(a) Stunting



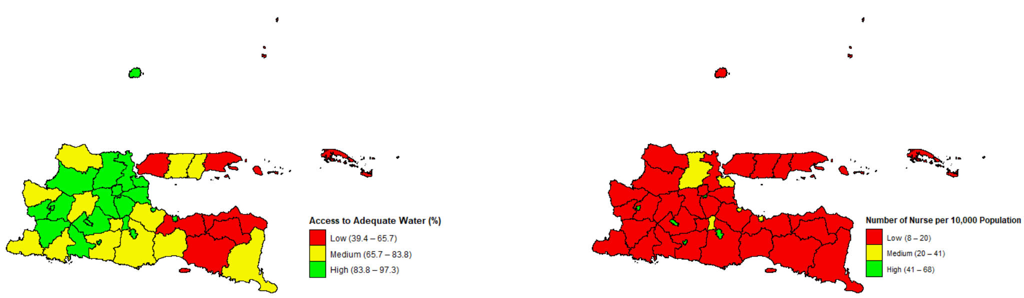
(b) Gender Development Index



(c) Poverty



(d) GRDP per Capita



(e) Access to Adequate Sanitation **(f) Number of Nurse per 10,000 population**

Figure 1. Mapping of dependent and independent variables

The sharp East-West spatial divide in clean sanitation access (Figure 1e) serves as an environmental catalyst for stunting. The severe sanitation deficits in eastern and island districts correspond directly with elevated stunting levels, as poor sanitation triggers recurrent waterborne diseases and chronic diarrhea in children, preventing proper nutrient absorption. This environmental risk is heavily compounded by the critical maldistribution of healthcare professionals. The distribution of nurses per 10,000 population (Figure 1f) is overwhelmingly depleted (red zones) in the exact peripheral areas where stunting peaks.

In summary, Figure 1 provides visual confirmation that stunting in East Java does not occur in a vacuum; its distribution is bound to spatial effect in poverty, gender inequality, water deficits, and healthcare scarcity. The synchronized alignment of these spatial patterns provides an empirical foundation for this study, and, as shown in Section 3.6, is further examined using a formal significance test (LISA) rather than visual class breaks alone.

3.2 Multiple Linear Regression

Firstly, a multiple linear regression model was developed to test the model without considering spatial effects. The results of multiple linear regression modeling can be seen in Table 2.

Table 2. Estimate Parameter of Multiple Linear Regression Model

Variable	Estimate	Std. Error	p-value
Intercept	63.16888	47.64818	0.1943
Gender Development Index (X_1)	-0.14691	0.42741	0.7333
Poverty rate (X_2)	-0.59344	0.37157	0.1201
log(GRDP per Capita) (X_3)	0.15409	2.12804	0.9427
Access to Adequate Sanitation (X_4)	-0.27711	0.12251	0.0306*
Number of Nurse per 10,000 population (X_5)	-0.12256	0.08937	0.1798

The resulting Ordinary Least Squares (OLS) model is formulated as follows:

$$y_i = 63.1689 - 0.1469X_1 - 0.5934X_2 + 0.1541X_3 - 0.2771X_4 - 0.12256X_5 + \varepsilon_i \qquad (7)$$

It can be observed that Access to Adequate Sanitation (X_4) is the sole statistically significant variable ($\beta = -0.2771$; $p = 0.0306$). The OLS model indicates that a 1 percent increase in access to adequate sanitation is associated with a 0.2771 percentage point decrease in stunting prevalence within the region, aligning with theoretical expectations and government targets [11]. The remaining variables do not exhibit statistical significance at the designated threshold.

Furthermore, the positive coefficient sign for log GRDP per capita appears counter-intuitive, since economic growth is ordinarily associated with lower rather than higher stunting rates. One explanation commonly offered for such a sign reversal is that the effect of the economic variable is confounded by other covariates within the model, an aggregation or suppression effect that can invert the direction of an observed association [26]. To establish whether this explanation applies to the present specification, Variance Inflation Factors were computed for all five predictors, as reported in Table 3.

Table 3. Variance Inflation Factors of the Independent Variables in the OLS Model

Variable	VIF
Gender Development Index (X_1)	2.299
Poverty rate (X_2)	2.695
log(GRDP per Capita) (X_3)	1.851
Access to Adequate Sanitation (X_4)	2.322
Number of Nurse per 10,000 population (X_5)	1.837

All Variance Inflation Factors fall within a narrow range, from 1.837 for nurse density to 2.695 for the poverty rate, and every value lies well below the conventional thresholds of 5 and 10. Multicollinearity is therefore not a substantial concern in this specification, and the counter-intuitive sign of the log GRDP per capita coefficient cannot be attributed to collinearity among the regressors. The aggregation or suppression interpretation is accordingly not supported by the diagnostic evidence. Given its very high p-value (p = 0.9427), the coefficient for log GRDP per capita is statistically indistinguishable from zero, and its sign should not be interpreted substantively. It is best treated as sampling noise rather than as evidence of a genuine relationship between regional economic output and stunting prevalence.

The F-statistic with a p-value lower than the significance level $\alpha = 0.10$ indicates that, simultaneously, the Gender Development Index, Poverty, GRDP per capita, Access to Adequate Sanitation, and the number of nurses per 10,000 population exert a significant joint effect on stunting rates at the α level. However, with an AIC of 257.5454, the OLS model rests on the assumption that observations across regions are independent. This assumption is weak for inter-regency stunting data, given that geographically adjacent regions share public health spillovers and regional socioeconomic patterns. This limitation motivates the testing of spatial dependence utilizing Spatial Autoregressive (SAR) and Spatial Error Models (SEM).

3.3 Test of Classical Assumptions and Spatial Effects

a. Normality Test

The normality assumption was evaluated using the Shapiro-Wilk test. The test produced a Shapiro-Wilk statistic of 0.9894 with a p-value of 0.9711, indicating that the residuals are normally distributed and the assumption of normality is satisfied.

b. Spatial Autocorrelation

Moran's I test was used to examine spatial independence among the residuals of the Ordinary Least Squares (OLS) model. The first step in this analysis is to detect the presence of spatial effects within the model's error term. The Moran's I test on the OLS residuals yielded a value of 0.1790 with a p-value of 0.0556. Because Moran's I is known to have reduced statistical power when the number of areal units is small, and this study observes only 38 regencies and cities, the analysis follows common practice in small-sample spatial econometric research by adopting a relaxed significance threshold of 10% rather than the conventional 5%, consistent with evidence that Moran's I retains comparatively better power than alternative tests when the number of spatial units is small [22], [27]. Under this threshold the p-value narrowly falls below 0.10, which provides marginal or borderline evidence against the null hypothesis of spatial randomness and is indicative of weak rather than strong positive spatial autocorrelation in the residuals. Since the same result would not reach significance at the conventional 5% level, it is treated here as motivating further investigation through explicit spatial models, reported in Sections 3.4 and 3.5, rather than as conclusive evidence of spatial dependence. On this basis a spatial analysis approach is examined as a diagnostic step in modeling the prevalence of stunting among toddlers in East Java Province.

c. Spatial Heterogeneity

Spatial heterogeneity was assessed using the Breusch-Pagan (BP) test, conducted on the residuals of the OLS model. The test yielded a BP statistic of 2.9538 with a p-value of 0.7071. Since the p-value exceeds 0.10, the null hypothesis of homoscedasticity cannot be rejected, indicating no evidence of spatial heterogeneity in the data.

3.4 Spatial Model

a. Spatial Autoregressive Model

The Spatial Autoregressive (SAR) model examines whether stunting in a specific region is influenced by the weighted average of stunting in neighboring regions (representing a spatial lag on the dependent variable, Y). The parameter estimation results for the SAR model are presented in Table 4.

Table 4. Estimate Parameter of SAR

Variable	Estimate	Std. Error	p-value
Intercept	53.7993	45.4818	0.2369
Gender Development Index (X_1)	-0.0954	0.3957	0.8096
Poverty rate (X_2)	-0.5240	0.3476	0.1317
log(GRDP per Capita) (X_3)	0.3345	1.9501	0.8634
Access to Adequate Sanitation (X_4)	-0.2846	0.1117	0.0108*
Number of Nurses per 10,000 population (X_5)	-0.0845	0.0930	0.3638
Rho (ρ)	0.1080	0.1539	0.5101

The SAR equation formed is as follows:

$$y_i = 53.7993 - 0.0954X_1 - 0.5240X_2 + 0.3345X_3 - 0.2846X_4 - 0.0845X_5 + 0.1080 \sum_{i=1, i \neq j}^{38} w_{ij} y_j + \varepsilon_i \quad (8)$$

Examining the independent variables, the coefficient for Access to Adequate Sanitation remains statistically significant ($\beta = -0.2846$; $p = 0.0108$). Meanwhile, consistent with the OLS model, the GRDP per capita variable is not statistically significant ($p = 0.8634$), with its coefficient sign remaining directionally inconsistent with theoretical expectations; as shown in Section 3.2, the Variance Inflation Factors rule out multicollinearity as the explanation, so this sign reversal is better attributed to sampling noise, given the coefficient's high p-value, than to a genuine Simpson's paradox effect. Regarding the spatial effect, the autoregressive parameter $\rho = 0.1080$ with a p-value $= 0.5101$ indicates that the spatial lag effect in the SAR model is not statistically significant at the α level (p-value $> \alpha$).

For model evaluation, the SAR lag model yielded an AIC value of 259.1116. Furthermore, the p-value from the Lagrange Multiplier (LM) test for residual autocorrelation is 0.16566 (> 0.10). This indicates that there is no spatial autocorrelation remaining in the residuals of the stunting model across regencies and cities in East Java.

b. Spatial Error Model

The Spatial Error Model (SEM) assumes that spatial dependence resides within the error term (capturing spatially correlated omitted variables) rather than the dependent variable (Y) itself. The parameter estimation results for the SEM are presented in Table 5.

Table 5. Estimate Parameter of SEM

Variable	Estimate	Std. Error	p-value
Intercept	62.2761	44.8382	0.1649
Gender Development Index (X_1)	-0.1108	0.3951	0.7793
Poverty rate (X_2)	-0.5355	0.3544	0.1307
log(GRDP per Capita) (X_3)	0.0884	1.990	0.9646
Access to Adequate Sanitation (X_4)	-0.3084	0.1097	0.0049*
Number of Nurse per 10,000 population (X_5)	-0.0971	0.0833	0.2441
Lambda (λ)	0.2795	0.1895	0.1570

The SEM equation formed is as follows:

$$y_i = 62.2761 - 0.1108X_1 - 0.5355X_2 + 0.0884X_3 - 0.3084X_4 - 0.0971X_5 + 0.2795 \sum_{i=1, i \neq j}^{38} w_{ij} u_j + \varepsilon_i \quad (9)$$

Consistent with both the OLS and SAR models, Access to Adequate Sanitation remains statistically significant ($p = 0.0049$). The finding that access to adequate sanitation is the only statistically significant predictor is consistent with previous evidence identifying sanitation as one of the strongest determinants of childhood stunting. Beal et al. (2018) concluded that poor sanitation contributes to repeated enteric infections and environmental enteric dysfunction, thereby reducing nutrient absorption and increasing the risk of chronic growth failure [4].

However, the spatial autoregressive error parameter, lambda ($\lambda = 0.2795$; $p = 0.1570$), is not statistically significant at the $\alpha = 10\%$ level. This indicates that spatial shocks from unobserved factors external to the model, which might affect neighboring regions, are not empirically supported. Given this combination of variables, if a particular region experiences a high error value such as from an economic shock or a natural disaster not captured by the independent variables its neighboring regions do not necessarily experience a corresponding high error. As a point of comparison with the alternative specifications, the SEM yields an AIC value of 257.5429.

3.5 Model Comparison

The OLS model demonstrates that the Gender Development Index, Poverty, log(GRDP per Capita), Access to Adequate Sanitation, and the number of nurses per 10,000 population simultaneously exert a significant effect on stunting rates at the α significance level. However, the Moran's I test provides marginal evidence of spatial autocorrelation in the OLS residuals (p = 0.0556), as discussed in Section 3.3. This borderline result is suggestive of weak spatial clustering in stunting prevalence, rather than conclusive evidence of a strong, independent spatial process, and is broadly consistent in direction with Trimono et al. (2025), who identified positive spatial autocorrelation of stunting prevalence across districts in East Java, suggesting that neighboring regions tend to exhibit similar stunting conditions [15]. Comparable findings were also reported by Iriany et al. (2023) who demonstrated that geographic proximity contributes substantially to similarities in stunting prevalence and recommended the application of spatial regression approaches instead of conventional regression models [1]. These studies reinforce the importance of considering spatial dependence during the initial stage of epidemiological modelling. This initially suggested that a spatial analysis approach would be appropriate for modeling the prevalence of toddler stunting across the regencies and cities of East Java Province.

Table 6. Model Comparison

Method	AIC
OLS	257.5454
SAR	259.1116
SEM	257.5429

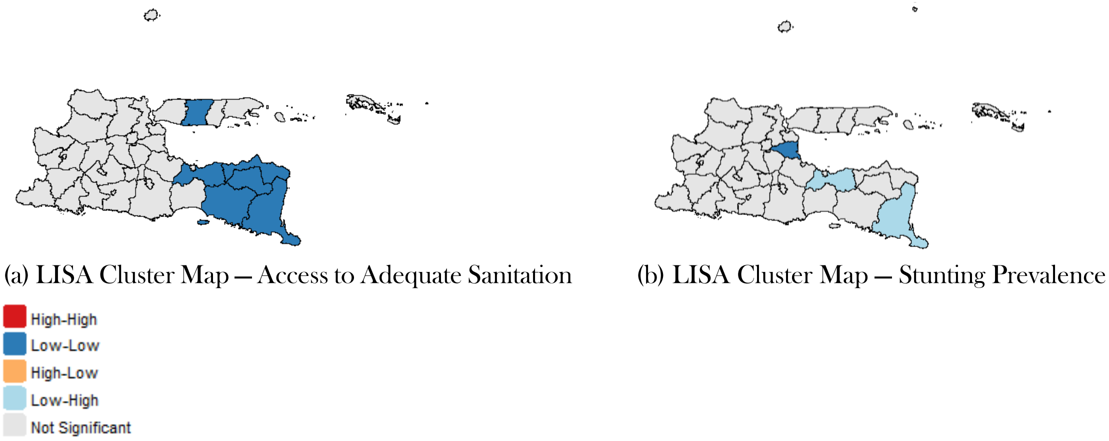
Furthermore, the Moran's I statistic and the SAR/SEM spatial parameters (ρ and λ) address related but distinct questions. Moran's I on the OLS residuals is an unconditional diagnostic: it tests whether residuals from a model with no explicit spatial structure are spatially clustered. Its significance therefore establishes that spatial dependence is present in the data-generating process, but it does not by itself indicate how much of that dependence would remain once a spatial lag or spatial error term is explicitly estimated alongside the covariates.

When this structure is estimated directly, however, neither ρ (SAR) nor λ (SEM) reaches statistical significance at α = 10%. Significant residual autocorrelation under OLS alongside non-significant spatial parameters once the spatial structure is modeled explicitly is not a contradiction. It suggests that a substantial share of the spatial clustering detected by Moran's I is mediated by the included covariates (GDI, poverty, log GRDP per capita, sanitation access, and nurse availability), which themselves tend to be spatially patterned across East Java's districts. Once these variables are held constant, little systematic spatial dependence remains for ρ or λ to capture. LeSage (2008) [25] and Pesaran et al. (2003) note that this is a common outcome in regional applications: apparent spatial autocorrelation is frequently a symptom of omitted, spatially patterned covariates rather than an intrinsic spatial process in the outcome itself, and it tends to attenuate once those covariates are included in the specification [28].

Although the SEM specification yields the marginally lowest AIC value (257.5429), the difference relative to the baseline OLS model (257.5454) is negligible (ΔAIC ≈ 0.003), and the difference relative to SAR (259.1116) also falls well within the range conventionally regarded as indicating no meaningful distinction between competing models (ΔAIC < 2). This is consistent with the earlier finding that neither ρ nor λ was statistically significant: the near-identical AIC values simply confirm, from a model-fit perspective, that explicitly incorporating spatial structure adds negligible explanatory value once GDI, poverty, log GRDP per capita, sanitation access, and nurse density are already included. Consequently, based on the principle of parsimony, the simpler classical OLS specification is preferred, since it achieves empirical performance equivalent to the spatial alternatives without the added complexity of estimating spatial dependence parameters.

3.6 Local Spatial Clustering Pattern (LISA)

To evaluate more directly whether the spatial pattern visible in Figure 1 constitutes genuine, statistically significant local clustering rather than only a broad visual association Local Indicators of Spatial Association (LISA) cluster maps were constructed for the OLS model's one significant covariate, access to adequate sanitation, and for stunting prevalence itself. Unlike the Natural Breaks classification in Figure 1, which only groups areas by value, LISA formally tests, district by district, whether an area's value is significantly similar to (High-High, Low-Low) or different from (High-Low, Low-High) the values of its neighbors. The results are presented in Figure 2.



(a) LISA Cluster Map — Access to Adequate Sanitation (b) LISA Cluster Map — Stunting Prevalence

Figure 2. LISA cluster maps of (a) access to adequate sanitation and (b) stunting prevalence

Figure 2a shows that access to adequate sanitation forms a clear, statistically significant Low-Low cluster covering a contiguous block of districts in the eastern peninsula and Madura/Tapal Kuda area, together with a smaller cluster along the northern coast, thereby confirming that low sanitation access in these districts is not an isolated condition but is spatially reinforced by similarly underserved neighbors. This matches the East-West divide already visible in Figure 1e and is consistent with sanitation access being the only structural covariate that remains statistically significant across all three model specifications (OLS, SAR, and SEM), echoing spatiotemporal evidence from India that LISA-detected sanitation clustering is a robust, recurring correlate of child growth-failure outcomes across survey rounds [29].

Figure 2b, by contrast, shows that stunting prevalence itself does not form an equivalent High-High cluster anywhere in the province. The only statistically significant local patterns are a small Low-Low pocket in the north-central area and a Low-High spatial outlier in the southeastern periphery, specifically a district with comparatively low stunting surrounded by higher-stunting neighbors. Despite the broad visual overlap between high stunting and high poverty, low sanitation, and low GDI in Figure 1, stunting prevalence does not, on formal LISA testing, form a significant hotspot (High-High cluster) in the same districts where sanitation access forms a significant cold-spot (Low-Low cluster); this kind of disconnect between a continuous prevalence map and its corresponding LISA cluster map has been documented elsewhere and is not by itself anomalous [30].

This asymmetry provides direct, map-based support for the study's central finding. If stunting itself possessed a strong, independent spatial process, a significant High-High cluster would be expected to coincide with the Low-Low sanitation cluster in the eastern periphery; instead, the significant local clustering is concentrated in the covariate (sanitation), not in the outcome (stunting). This is consistent with the non-significant ρ and λ parameters obtained in the SAR and SEM models (Sections 3.4–3.5), and parallels spatial-error evidence from India in which sanitation access, rather than an unmodeled spatial process, was the covariate most consistently associated with reduced child growth failure once spatial dependence was modeled directly [29]. The spatial dependence picked up by the global Moran's I test on OLS residuals appears to be substantially explained by the spatial pattern of sanitation access and related structural covariates, rather than reflecting a spatial process intrinsic to stunting that would additionally require a spatial lag or spatial error term. The LISA evidence therefore reinforces, at the level of individual districts, the model comparison conclusion that the classical OLS specification, estimated with sanitation access as its key significant predictor, is sufficient for inference on the determinants of stunting in East Java, directly supporting the study's objective of evaluating spatial econometric frameworks against a baseline classical regression model.

4. CONCLUSION

In conclusion, the Gender Development Index, Poverty, $\log(\text{GRDP per Capita})$, Access to Adequate Sanitation, and the number of nurses per 10,000 population simultaneously exert a significant joint effect on stunting rates. While the baseline OLS model shows marginal evidence of spatial autocorrelation in the residuals based on Moran's I ($p = 0.0556$) and verifies homoscedasticity via the Breusch Pagan test, the subsequent application of the full Spatial Autoregressive (SAR) and Spatial Error Models (SEM) yields spatial parameters that do not reach statistical significance. Specifically, both ρ (Rho) and λ (Lambda) yield p-values that exceed the 10% ($\alpha = 0.10$) significance level, indicating that there is insufficient statistical evidence of residual spatial dependence after accounting for the observed structural covariates in the comprehensive model specification. Given this, and following the principle of parsimony, the classical OLS specification is retained as the final model for inference in this study, with the SAR and SEM estimates reported as spatial diagnostic comparisons rather than as the preferred model.

The LISA cluster maps reinforce this conclusion at the district level: access to adequate sanitation forms a clear, statistically significant Low-Low cluster in the eastern and northern periphery, while stunting prevalence itself does not form a corresponding High-High cluster anywhere in the province. This asymmetry indicates that the spatial dependence detected by Moran's I on the OLS residuals is better explained by the spatial patterning of sanitation access and related structural covariates than by an independent spatial process in stunting, providing map-based, case-specific support alongside the AIC and parameter-significance evidence for retaining the classical OLS specification. Given the near-identical model fit across specifications, the absence of a significant ρ or λ, and the absence of a significant High-High cluster for stunting itself, the classical OLS specification is retained as the final model for inference, directly addressing this study's objective of evaluating spatial econometric frameworks against a baseline classical regression model. The SAR, SEM, and LISA results are reported as spatial diagnostics rather than as the preferred model.

A notable limitation of this study is its sample size, constrained to the 38 regencies and cities that constitute East Java Province ($n = 38$). This relatively small number of spatial units may itself limit the statistical power of the SAR and SEM parameter tests, thereby reducing the ability to detect residual spatial dependence after accounting for the observed covariates. In other words, the non-significance of ρ and λ observed here likely reflects a combination of two factors covariate absorption of spatial structure and limited test power at small n which this study's design cannot fully disentangle. Future research using a larger geographical scope, such as national or cross-provincial data, together with additional multidimensional determinant variables not yet captured in the current framework, would help clarify which explanation dominates and would allow spatial spillover effects to be tested with greater statistical power.

5. REFERENCES

[1] A. Iriany, W. Ngabu, D. Arianto, and A. Putra, "Classification Of Stunting Using Geographically Weighted Regression-Kriging Case Study Stunting In East Java," *Barekeng*, vol. 17, no. 1, pp. 495–504, Mar. 2023, https://doi.org/10.30598/barekengvol17iss1pp0495-0504.

[2] N. O. Nirmalasari, "Stunting pada Anak: Penyebab dan Faktor Risiko Stunting di Indonesia," *QAWWAM: Journal for Gender Mainstreaming*, vol. 14, no. 1, pp. 19–28, 2020, https://doi.org/10.20414/Qawwam.v14i1.2372.

[3] A. Fadliana and P. P. Drajat, "Pemetaan Faktor Risiko Stunting Berbasis Sistem Informasi Geografis Menggunakan Metode Geographically Weighted Regression," *IKRAITH-INFORMATIKA*, vol. 5, no. 3, pp. 91–102, Nov. 2021.

[4] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld, "A Review of Child Stunting Determinants in Indonesia," *Matern. Child Nutr.*, vol. 14, no. 4, pp. 1–10, Mar. 2018, https://doi.org/10.1111/mcn.12617.

[5] S. Siswanto, M. Mirna, M. Yusran, U. A. Syam, and A. S. I. Miolo, "Identification of Factors that Influence Stunting Cases in South Sulawesi using Geographically Weighted Regression Modeling," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 1, pp. 100–108, Sep. 2022, https://doi.org/10.20956/j.v19i1.21617.

[6] E. Rianti, A. Triwinarto, E. Lukman, and Sudarmi, "Aplikasi Cegah Anak Lahir Stunting Berbasis Android," in *Prosiding Forum Ilmiah Tahunan IAKMI (Ikatan Ahli Kesehatan Masyarakat Indonesia)*, Nov. 2020.

[7] United Nations, *The 17 goals*. Accessed: Jun. 22, 2026. [Online]. Available: https://sdgs.un.org/goals

[8] World Health Organization, "Reducing Stunting in Children: Equity Considerations for Achieving the Global Nutrition Targets 2025," 2018.

[9] M. Shekar, J. Kakietek, J. Dayton Eberwein, and D. Walters, "An Investment Framework for Nutrition: Reaching the Global Targets for Stunting, Anemia, Breastfeeding, and Wasting," Washington DC, 2017. doi: https://doi.org/10.1596/978-1-4648-1010-7.

[10] S. Liza Munira, "Hasil Survei Status Gizi Indonesia (SSGI) 2022," Jakarta, Feb. 2023.

[11] *Peraturan Republik Indonesia Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting*. 2021.

[12] Y. Karyati and A. Julia, "Pengaruh Jumlah Penduduk Miskin, Laju Pertumbuhan Ekonomi, dan Tingkat Pendidikan terhadap Jumlah Stunting di 10 Wilayah Tertinggi Indonesia Tahun 2010-2019," *Jurnal Riset Ilmu Ekonomi dan Bisnis*, vol. 1, no. 2, pp. 101–108, Dec. 2021, https://doi.org/10.29313/jrieb.v1i2.401.

[13] A. Salam, D. S. Pratomo, and P. M. A. Saputra, "Analisis Kemiskinan pada Rumah Tangga di Jawa Timur Melalui Pendekatan Multidimensi dan Moneter," *Jurnal Kependudukan Indonesia*, vol. 16, no. 2, p. 127, Mar. 2022, https://doi.org/10.14203/jki.v16i2.480.

[14] H. Pramoedyo, S. Astutik, F. Fauwziyah, and E. Pratiwi, "Modeling Geographically Weighted Negative Binomial Regression (GWNBR) on Stunting Incidence in Malang Regency," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 1, pp. 163–171, Sep. 2022, https://doi.org/10.20956/j.v19i1.21757.

[15] Trimono, Amri Muhaimin, P. C. Ekacitta, and A. E. Ardiani, "Spatial Autocorrelation Analysis of East Java Stunting Prevalence Cases in 2023," *Journal of Advances in Information and Industrial Technology*, vol. 7, no. 1, pp. 83–94, May 2025, https://doi.org/10.52435/jaiit.v7i1.689.

[16] D. Rantini *et al.*, "Understanding the Spatial Distribution of Stunting in East Java, Indonesia: A Comparison of GWR and MS-GWR Models," *Statistics, Optimization and Information Computing*, vol. 15, no. 1, pp. 311–323, Jan. 2026, https://doi.org/10.19139/soic-2310-5070-3066.

[17] Y. S. Dewi, S. Hastuti, and M. Fatekurohman, "Analysis of stunting in East Java, Indonesia using random forest and geographically weighted random forest regression," *Brazilian Journal of Biometrics*, vol. 42, no. 3, pp. 213–224, Aug. 2024, https://doi.org/10.28951/bjb.v42i3.679.

[18] I. Siramaneerat, E. Astutik, F. Agushybana, P. Bhumkittipich, and W. Lamprom, "Examining determinants of stunting in Urban and Rural Indonesian: a multilevel analysis using the population-based Indonesian family life survey (IFLS)," *BMC Public Health*, vol. 24, no. 1, Dec. 2024, https://doi.org/10.1186/s12889-024-18824-z.

[19] M. Usman and K. Kopczewska, "Spatial and Machine Learning Approach to Model Childhood Stunting in Pakistan: Role of Socio-Economic and Environmental Factors," *Int. J. Environ. Res. Public Health*, vol. 19, no. 17, Sep. 2022, https://doi.org/10.3390/ijerph191710967.

[20] G. F. Jenks, *The Data Model Concept in Statistical Mapping*. 1967. Accessed: Jun. 22, 2026. [Online]. Available: https://ci.nii.ac.jp/naid/10021899676

[21] S. I. Bangdiwala, "Regression: Multiple Linear," *Int. J. Inj. Contr. Saf. Promot.*, vol. 25, no. 2, pp. 232–236, Apr. 2018, https://doi.org/10.1080/17457300.2018.1452336.

[22] L. Anselin, "Spatial Econometrics," B. H. Baltagi, Ed., Blackwell Publishing Ltd, 2001, ch. Chapter Fourteen, pp. 310–330. https://doi.org/10.1002/9780470996249.

[23] K. Suryowati, R. D. Bekti, and A. Faradila, "A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Apr. 2018. https://doi.org/10.1088/1757-899X/335/1/012052.
[24] J.-M. Floch and L. R. Saout, "Spatial econometrics - common models," in *Handbook of Spatial Analysis*, Montrouge: INSEE-EUROSTAT, 2018, ch. Chapter 6, pp. 149–177. Accessed: Jun. 10, 2026. [Online]. Available: https://www.efgs.info/insee-handbook-of-spatial-analysis/
[25] J. P. LeSage, "An Introduction to Spatial Econometrics," Sep. 15, 2008, *Editions Techniques et Economiques*. https://doi.org/10.4000/rei.3887.
[26] R. Elvik, "Simpson's paradox: A collection of examples from road safety studies and emergency medicine," *Transp. Res. Interdiscip. Perspect.*, vol. 31, May 2025, https://doi.org/10.1016/j.trip.2025.101471.
[27] D. Curran-Everett and D. J. Benos, "Guidelines for reporting statistics in journals published by the American Physiological Society," 2004, *American Physiological Society*. https://doi.org/10.1152/ajprenal.00186.2004.
[28] M. H. Pesaran and M. Weeks, "Nonnested Hypothesis Testing: An Overview," in *A Companion to Theoretical Econometrics*, B. H. Baltagi, Ed., Wiley, 2003, pp. 279–309. https://doi.org/10.1002/9780470996249.ch14.
[29] L. Jain *et al.*, "Association of Child Growth Failure Indicators With Household Sanitation Practices in India (1998-2021): Spatiotemporal Observational Study," *JMIR Public Health Surveill.*, vol. 10, 2024, https://doi.org/10.2196/41567.
[30] R. Amir-ud-Din, S. Fawad, L. Naz, S. Zafar, R. Kumar, and S. Pongpanich, "Nutritional inequalities among under-five children: a geospatial analysis of hotspots and cold spots in 73 low- and middle-income countries," *Int. J. Equity Health*, vol. 21, no. 1, Dec. 2022, https://doi.org/10.1186/s12939-022-01733-1.