JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 3, July 2026, pp. 914-925

Hourly Wage Modeling in Indonesia using Spatial Durbin Model Approach

Suliyanto[1*], Dita Amelia[1], Ilham Maulana Al Hasri[1], Nashwa Carista[1]
[1]Department of Mathematics, Airlangga University, Indonesia
suliyanto@fst.unair.ac.id

Article History:
Received : 17-10-2025
Revised : 02-01-2026
Accepted : 03-01-2026
Online : 01-07-2026

Keywords*:*
Spatial Durbin Model;
Hourly Wage;
Maximum Likelihood Estimation;
Regression Analysis.

ABSTRACT

Hourly wage disparities in Indonesia reflect complex regional economic conditions that vary across provinces. These disparities are closely related to spatial factors, as economic conditions in one region may influence neighboring regions. This study aims to compare the performance of the Ordinary Least Squares (OLS) linear regression model and the Spatial Durbin Model (SDM) in identifying the determinants of hourly wages in Indonesia. The study uses secondary data from BPS Indonesia for 2023, covering 34 provinces. Predictor variables used including the poverty gap index, expected years of schooling, GRDP per capita, and the percentage of poor population. Spatial effects were examined using Moran's I and the Breusch-Pagan test. The test results indicate the presence of both spatial dependence and heterogeneity in provincial hourly wages, suggesting that the OLS model is insufficient to capture spatial interactions between regions. Therefore, the Spatial Durbin Model is applied to accommodate both direct effects and spatial spillover effects. The empirical results of the SDM show that the poverty gap index and GRDP per capita have significant direct effects on hourly wages at the provincial level. In addition, the poverty gap index and expected years of schooling exhibit significant indirect effects. Model performance was evaluated using the coefficient of determination (R-Square), Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results show that the Spatial Durbin Model outperforms the OLS model, as indicated by a higher R-Square value and lower MSE, MAE, and MAPE values.

Crossref
https://doi.org/10.31764/jtam.v10i3.35666

A. INTRODUCTION

Hourly wages are a fundamental indicator of the well-being of workers and the quality of life of a country's population (Kasy, 2022). In Indonesia, significant disparities in hourly wages across regions indicate structural imbalances in regional economic development. Wage differences across provinces are often associated with variations in productivity, human capital, and regional economic capacity (Elhorst, 2014; Dorjnyambuu & Galambosné Tiszberger, 2024). According to the Central Bureau of Statistics (BPS), the national hourly wage increased by 9.8% year-on-year to an average of IDR 19,270. However, this increase was not evenly distributed among provinces. DKI Jakarta recorded the highest average hourly wage of IDR 42,354, while West Nusa Tenggara (NTB) had the lowest average hourly wage of IDR 12,933 (BPS, 2025). Wage disparity represents a major obstacle to achieving the Sustainable Development Goals (SDGs) Goal 8, which focuses on economic growth and decent work. Therefore, further analysis is needed to determine the predominant factors contributing to wage inequality using an appropriate regression model approach.

Regression analysis is a statistical approach used to investigate the relationship between covariate variables and a response variable. In regional economic studies, parametric regression models are commonly applied because the functional form of the regression is explicitly specified. However, conventional linear regression using Ordinary Least Squares (OLS) assumes that observations are independent across regions. This assumption may be violated when spatial dependence exists, where economic and labor market conditions in one region influence those in neighboring regions. To address this issue, spatial regression models are employed to analyze data that exhibit spatial dependence (Anselin, 1988).

There are two main models in spatial regression, namely the Spatial Lag Model (SLM) and the Spatial Error Model (SEM). The selection of the appropriate model depends on data characteristics and the results of diagnostic tests performed. Spatial regression models that consider the influence of covariate variables in a region and response variables in other regions are known as Spatial Lag Models (Ramli et al., 2024). Meanwhile, regression models that involve covariates in a region and spatial dependence on error components in other regions are referred to as Spatial Error Models (SEM) (Anselin, 1988). In addition, nonparametric regression models assume an unknown form of the regression function, allowing greater flexibility in capturing complex relationships between variables. These models can be approximated using several approaches, such as Fourier series functions (Mardianto et al., 2021; Ramli et al., 2024; Suliyanto et al., 2025), spline functions (Putra et al., 2023; Sriliana et al., 2022), and polynomial local estimators (López-Ureña & Yáñez, 2024; Wu et al., 2023).

An extension of the Spatial Lag Model is the Spatial Durbin Model (SDM), which incorporates spatial lags of both the response variable and the explanatory variables. This model is able to capture direct effects as well as indirect or spillover effects between regions, which are particularly relevant in the context of wage formation where labor mobility and human capital interactions occur across provincial boundaries (Elhorst, 2014). Empirical studies have shown that spatial regression models, including the SDM, perform better than OLS in explaining regional economic phenomena by capturing spatial dependence and reducing estimation bias (Takezawa, 2005). Several previous studies have applied SLM, SEM, and SDM to analyze regional economic conditions (Paramita et al., 2021; Shang et al., 2025; Varlamova & Kadochnikova, 2023; Wu et al., 2023; Zhong et al., 2023). However, studies that specifically compare OLS and SDM in modeling hourly wage disparities across Indonesian provinces remain limited. Therefore, this study aims to analyze hourly wages in Indonesia using the Spatial Durbin Model and to compare its performance with the OLS regression model, in order to highlight the role of spatial spillover effects in explaining interregional wage inequality.

B. METHODS

This study uses secondary data from Statistics Indonesia (BPS) in 2023. The data used is cross-sectional data containing 34 provinces in Indonesia with 1 response variable and 4 predictor variables presented in the following Table 1.

Table 1. Research Variable

Variable	Description	Scale
y	Hourly Wage	IDR
x_1	Poverty Gap Index	Ratio
x_2	Expected Years of Schooling	Year
x_3	GDRP per Capita	IDR
x_4	Percentage of Poor Population	Percentage

The procedures carried out in this study were modeling with a global linear regression model, testing for spatial effects, modeling with a Durbin spatial regression model, and evaluating the model. In detail, the research procedures can be described as follows.

1. Spatial Effect Test

a. Moran's I Test

Moran's I index is utilized to quantify spatial autocorrelation, defined as the extent to which a variable at a specific location exhibits a pattern of association with neighboring locations (Moran, 1950). The Moran's I test statistic is calculated by the equation (1) below (Cartone et al., 2022).

$$Z_I = \frac{I - E(I)}{\sqrt{var(I)}} \text{ with } I = \frac{n}{\sum_i \sum_j w_{ij}} \times \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \quad (1)$$

where n is the number of observations; w_{ij} is the element of the i-th row and j-th column spatial weighting matrix; y_i is the value of the response variable at location i; y_j is the value of the response variable at location j; and $\bar{y}$ is the average of the response variable y. with hypothesis:

$H_0: I = 0$ (No spatial dependencies)

$H_1: I \neq 0$ (There is a spatial dependency)

The critical region of the test is that H_0 is rejected if the value $|Z_I| > Z_{\alpha/2}$ or *p-value* < α. If Moran's I index is statistically significant, it is necessary to employ a spatial regression approach for the purpose of capturing the effects of interregional linkages in the estimation model (Tobler, 1889). A recent study by Elhorst (2014) demonstrated that neglecting spatial autocorrelation can result in bias in the estimation results.

b. Breusch-Pagan Test

The Breusch-Pagan test is used to identify the presence of heteroscedasticity, which is a condition in which residual variants are not constant across regions. In the context of spatial regression, heteroscedasticity often arises due to differences in economic characteristics between regions (Breusch & Pagan, 1979). The statistics of the Breusch-Pagan test are expressed in equation (2) as follows.

$$BP = \frac{1}{2}\boldsymbol{f}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{f} \tag{2}$$

Where the element of vector $\boldsymbol{f}$ is $f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$, e_i is the least square observation for the i-th observation. $\boldsymbol{Z}$ is an $n \times p$ sized matrix containing a vector of standardized covariates variables. The test hypothesis uses the Breusch-Pagan test statistics as follows.
H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (Homoskedasticity)
H_1 : there is at least one $\sigma_i^2 \neq \sigma^2; i = 1,2,\dots,n$ (Heteroscedasticity)

The critical region is that H_0 is rejected if the value of $BP > \chi_p^2$ or *p-value* < α with p is number of covariates. In the event of heteroscedasticity being indicated by the test results, corrections must be implemented. Such corrections may take the form of robust estimation αp or the Generalized Least Squares (GLS) method, which serves to enhance the validity of the regression results (Anselin, 2005).

2. Modeling with Spatial Durbin Model

The general form of the spatial durbin model according to Anselin (1988) is expressed in equation (3) as follows.

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}) \tag{3}$$

With $\boldsymbol{y}$ is a vector of dependent variables with dimension $n \times 1$; $\boldsymbol{X}$ **is** a matrix of is the matrix of the covariate variable sized $n \times (p+1)$; ρ is a coefficient of spatial lag from response variable; $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ is a vector of spatial durbin model coefficient with dimension $(p+1) \times 1$; $\boldsymbol{W}$ is a spatial weighting matrix sized $n \times n$; and $\boldsymbol{\varepsilon}$ is a random error vector with dimension $n \times 1$ that assumed to be multivariate normally distributed with mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}$. If it is defined $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, then from equation (3) the PDF of $\boldsymbol{\varepsilon}$ is obtained that is expressed in equation (4) below.

$$f(\boldsymbol{\varepsilon}) = \frac{1}{\left(\sqrt{2\pi}\right)^n\sqrt{|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\right\} \tag{4}$$

The likelihood function of $\boldsymbol{y}$ that obtained from equation (4) is multiplied by the determinants of the Jacobian transformation matrix as follows.

$$L\left(\rho, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2 \mid \boldsymbol{y}\right) = \frac{1}{\left(\sqrt{2\pi}\right)^n\sqrt{|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\right\} |\boldsymbol{J}| \tag{5}$$

From equation (3) it is obtained

$$\boldsymbol{\varepsilon} = (\boldsymbol{I} - \rho\boldsymbol{W})\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{W}\boldsymbol{X}\boldsymbol{\eta} \tag{6}$$

By differentiating $\boldsymbol{\varepsilon}$ to vector $\boldsymbol{y}$, then the Jacobian transformation matrix is obtained as follows.

$$\boldsymbol{J} = \frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{y}^T} = \begin{pmatrix} 1 & -\rho_{W12} & \cdots & -\rho_{W1n} \\ -\rho_{W21} & 1 & \cdots & -\rho_{W2n} \\ \vdots & \vdots & \ddots & \vdots \\ -\rho_{Wn1} & -\rho_{Wn2} & \cdots & 1 \end{pmatrix} = \boldsymbol{I} - \rho \boldsymbol{W} = \boldsymbol{A} \tag{7}$$

From equation (7) the log-likelihood function is obtained as follows.

$$\ell = \ln|\mathbf{A}| - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta})'(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta}) \tag{8}$$

The sufficient conditions for the log-likelihood function in equation (8) reach the maximum value are $\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{0}$, $\frac{\partial \ell}{\partial \sigma^2} = 0$, $\frac{\partial \ell}{\partial \boldsymbol{\eta}} = \mathbf{0}$, and $\frac{\partial \ell}{\partial \rho} = 0$. Next estimate the $\boldsymbol{\beta}$ as follows.

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^2}(-2\boldsymbol{X}'\mathbf{A}\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{W}'\mathbf{X}\boldsymbol{\eta} + 2\boldsymbol{X}'\boldsymbol{X\beta}) = \mathbf{0}$$

$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{A}\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}'\mathbf{X}\,\boldsymbol{\eta} \tag{9}$$

Next estimate the σ^2 as follows.

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta})'(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta}) = 0$$

$$\sigma^2 = \frac{1}{n}(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta})'(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta}) \tag{10}$$

Then estimate the $\boldsymbol{\eta}$ as follows.

$$\frac{\partial \ell}{\partial \boldsymbol{\eta}} = -\frac{1}{2\sigma^2}(-2\boldsymbol{X}'\boldsymbol{W}'\mathbf{A}\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{W}'\mathbf{X}\boldsymbol{\beta} + 2\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{WX\eta}) = \mathbf{0}$$

$$\boldsymbol{\eta} = (\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{WX})^{-1}\boldsymbol{X}'\boldsymbol{W}'\mathbf{A}\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{WX})^{-1}\boldsymbol{X}'\boldsymbol{W}'\mathbf{X}\boldsymbol{\beta} \tag{11}$$

The vector $\boldsymbol{\beta}$ in equation (9) is a function of ρ. Next, by substituting equation (10) to equation (8) it is obtained

$$\ell = \ln|\mathbf{A}| - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{1}{\mathrm{n}}(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta})'(\mathbf{A}\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{WX\eta})\right) - \frac{n}{2} \tag{12}$$

To obtain the value of ρ that maximized the pseudo log-likelihood function in equation (12), the univariate Newton-Raphson is employed with the following steps.

a. Differentiating $\ell(\rho)$ in (19) to the parameter ρ, namely $\ell'(\rho)$
b. Setting the initial value of ρ_0
c. Calculating $\ell(\rho_0)$ and $\ell'(\rho_0)$

d. Calculating $\rho_{i+1} = \rho_i - \frac{\ell(\rho_i)}{\ell'(\rho_i)}$ for $i = 0, 1, 2, \dots$
e. If $|\rho_{i+1} - \rho_i| < \delta$ for δ a small positive number, then proceed to step (6). If $|\rho_{i+1} - \rho_i| \geq \delta$,then go back to step (4)
f. Calculating $\hat{\rho} = \rho_{i+1}$

The value of $\hat{\rho}$ that obtained is used to estimate $\boldsymbol{\beta}$ in equation (9), estimate σ^2 in equation (10), and estimate $\boldsymbol{\eta}$ in equation (11), then it is obtained the estimation of SDM as follows.

$$\hat{\boldsymbol{y}} = \hat{\rho}\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\eta}} \quad (13)$$

3. Best Regression Model Selection

In spatial regression analysis, the assessment of model goodness is of paramount importance to ensure the accurate representation of the relationship between covariate and response variables. The evaluation of spatial regression models is commonly undertaken by means of several key performance indicators, including the coefficient of determination R-Square, Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The coefficient of determination, R-Square, is a measure of the extent to which a model can explain variations in the data. Higher R-Square values indicate a stronger model capacity to explain variations in the data. In the context of spatial regression, R-Square is particularly useful in assessing the extent to which the model captures spatial patterns in the data. The R-Square value is defined in equation (14) as follows (Serefoglu et al., 2024).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (14)$$

with y_i is actual value of the i-th observation of response variable; and $\hat{y}_i$ is predicted value of the i-th observation of response variable. It has been demonstrated in previous research that R-Square is a more informative measure than other evaluation indices, such as MAE, MSE and MAPE, in the context of regression analysis (Chicco et al., 2021). The utilisation of R-Square as a regression model evaluation standard facilitates a more precise interpretation of model performance. MSE is utilised to quantify the square mean of the discrepancy between the observed value and the predicted value within a regression model (Arum et al., 2024). The MSE value is indicative of the quality of the model estimate, with lower values indicating a more accurate estimation. In addition to MSE, MAE is also used as an evaluation metric in this study. Additionally, a model's accuracy is typically determined by the value of its MAPE. MSE, MAE, and MAPE are used to calculate the performance of the model in predicting responses. MSE, MAE, and MAPE will be smaller if the values predicted by the model are closer to the actual values (Eeuwijk, 2022). Therefore, these metrics are also used in this study to analyze how far the predicted values generated by the model are from the actual values. The calculations for MSE, MAE, and MAPE are as follows.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (15)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{16}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{17}$$

C. RESULT AND DISCUSSION

1. Descriptive Statistics

This study employs a cross-sectional dataset consisting of observations across 34 administrative regions. The dataset is sourced from the official national statistics agency and contains several covariates commonly associated with socioeconomic disparity analysis. An overview of hourly wages in Indonesia (y) and the variables that are thought to influence it, namely the poverty gap index (x_1), expected years of schooling (x_2), GRDP per capita (x_3), and the percentage of poor people (x_4) are presented with descriptive statistics in Table 2.

Table 2. Descriptive Statistics of Research Variable

Variable	Mean	Standard Deviation	Min	Max
y	19.662	5.664	12.933	42.354
x_1	102.40	10.00	88.65	135.56
x_2	13.296	0.738	11.150	15.66
x_3	48181	36034	13513	192133
x_4	10.089	5.184	4.250	26.03

Based on Table 2, it shows that hourly wages in Indonesia (y) have a mean of 12933 and a standard deviation of 5664 with a minimum value of 12933 in NTB and a maximum value of 42354 in DKI Jakarta. Prior to modeling, each variable was examined to identify the presence of outliers, missing values, and non-normal distributions. Standardization was applied to all covariates to eliminate scale effects, allowing for comparable interpretation of estimated coefficients.

2. Modelling with Multiple Linear Regression using OLS

Before applying a regression model, it is essential to perform a test for multicollinearity. This test helps determine whether there exists a significant linear correlation among the predictor variables in the model. The assessment of multicollinearity is typically carried out using the Variance Inflation Factor (VIF). If the VIF for any variable exceeds a value of 10, it indicates the presence of multicollinearity, as shown in Table 3.

Table 3. Multicollinearity Test Results

Variable	VIF
x_1	1.57
x_2	1.06
x_3	1.39
x_4	1.33

Based on Table 3, the VIF value is less than 10 for all covariate variables. Therefore, it can be concluded that there is no multicollinearity. Then, the modeling stage with linear regression can be carried out, beginning with parameter estimation as follows, as shown in Table 4.

Table 4. Linear Regression Model Parameter Estimation and Testing Results

Parameter	Coefficient	Standard Deviation	t-value	p-value
$\hat{\beta}_0$	1.766	12.71	0.139	0.890
$\hat{\beta}_1$	0.2629	0.6655	3.950	0.000
$\hat{\beta}_2$	-0.8729	0.7420	-1.176	0.249
$\hat{\beta}_3$	8.004×10^{-5}	1.862×10^{-5}	4.299	0.000
$\hat{\beta}_4$	-0.1265	118.3	-1.069	0.294

From the simultaneous testing, the p-value $0.000 < \alpha$ (0.05) or the F test value of 21.14 > $F_{(0.05;29;4)} = 2.7014$ so it can be concluded that the poverty gap index, expected years of schooling, GRDP per capita, and the percentage of poor people simultaneously affect hourly wages in Indonesia. Based on the modeling with classical regression in Table 3, the individual test results show that the poverty gap index and GRDP per capita have an individually significant effect on the value of hourly wages because they have p-value $< \alpha$ (0.05). In addition, the R-Square value of the linear regression model is 74.46%, the MSE value is 7.9519, the MAE value is 1.9395 and the MAPE value is 9.76%. Based on the coefficient of determination, the covariate variables of the multiple linear regression model with linear regression can explain the variation of hourly wages by 74.46%.

3. Testing for Spatial Effects

The results from the linear regression model do not account for spatial effects. To enhance the realism of the model, spatial models will be considered, namely the SDM approach. Spatial assumption testing will be performed, including testing for spatial dependence using Moran's I test and spatial heterogeneity through the Breusch-Pagan test. The spatial dependency test is designed to evaluate whether data points from one location are affected by values from other nearby locations. By using the Moran's I test, a p-value of $0.013 < \alpha$ (0.05) was obtained. It can be concluded that spatial dependence exists. In addition, spatial heterogeneity testing was conducted to test whether there are differences in conditions between one location and another. The Breusch-Pagan test result showed that the p-value of $0.009 < \alpha$ (0.05) was obtained, so it can be concluded that there are differences in conditions between regions.

4. Modeling with Spatial Durbin Model

In modeling hourly wages in Indonesia in 2023 using the Spatial Durbin Model (SDM), several stages were undertaken, including SDM model estimation, model adequacy testing, inference or parameter significance testing, and evaluate the regression model.

a. Parameters Estimation

The parameter estimation results using maximum likelihood estimator (MLE) for the SDM applied to hourly wage data in Indonesia are shown in Table 5 below.

Table 5. Parameters Estimation Results of the SDM Model

Parameter	Estimated Value
$\hat{\rho}$	0.3994
$\hat{\beta}_0$	-14.473
$\hat{\beta}_1$	0.3124
$\hat{\beta}_2$	-0.0189
$\hat{\beta}_3$	7.617×10^{-5}
$\hat{\beta}_4$	-0.1627
$\hat{\eta}_1$	-0.2269
$\hat{\eta}_2$	0.8267
$\hat{\eta}_3$	3.636×10^{-5}
$\hat{\eta}_4$	0.2974

Based on the parameter presented in Table 5, the SDM model for hourly wages across Indonesian provinces is expressed as follows:

$$\begin{aligned}\hat{y}_i = {} & -14473 + 0.3994W_iy + 312.42x_{i,1} - 18.944x_{i,2} + 0.076x_{i,3} - 162.71x_{i,4} \\ & - 226.93W_ix_{i,1} + 826.79W_ix_{i,2} + 0.0364W_ix_{i,3} + 297.40W_ix_{i,4}\end{aligned} \quad (18)$$

where W_i denotes the i-th row vector within the spatial weighting matrix W.

b. Model Adequacy Test

The adequacy test for the SDM model was conducted. The obtained p-value was 0.022 < α (0.05) and LRT value of 5.213 > $\chi^2_{(0.05;1)} = 3.481$. Therefore, it can be concluded that the SDM model is adequacy.

c. Individual Parameter Test

Subsequently, the individual test of the SDM model was carried out to examine each parameter β_j and η_j for $j = 1,2,3,4$. The individual test results are shown in Table 6 below.

Table 6. Individual Parameter Testing of the SDM Model

Parameter	z-value	p-value
$\hat{\beta}_0$	-1.8302	0.067
$\hat{\beta}_1$	5.3803	0.000
$\hat{\beta}_2$	-0.0493	0.961
$\hat{\beta}_3$	4.7116	0.000
$\hat{\beta}_4$	-1.5742	0.115
$\hat{\eta}_1$	-2.7358	0.006
$\hat{\eta}_2$	1.9987	0.045
$\hat{\eta}_3$	1.0247	0.305
$\hat{\eta}_4$	1.8088	0.070

Based on Table 6, it shows that the poverty gap index, GRDP per capita, lag poverty gap index, and lag school expectancy significantly affect hourly wages because they have a $|z| > z_{0.05}(1.96)$ or p-value $< \alpha(0.05)$. Meanwhile, school expectancy, the percentage of poor population, lag GRDP per capita, and lag percentage of the poor population do not significantly affect hourly wages as they have $|z| < z_{0.05}(1.96)$ or p-value $> \alpha(0.05)$.

The school expectancy variable in a particular region does not significantly affect hourly wages, but the lag school expectancy or the school expectancy in directly bordering regions significantly impacts hourly wages in that region.

5. Best Model Selection

The best model was determined by comparing R-Square, MSE, MAE, and MAPE values across linear regression and SDM. The selected model is the one with the highest R-Square value, smallest MSE, smallest MAE, and smallest MAPE, as shown in Table 7.

Table 7. Model Comparison Based on Goodness-of-Fit Criteria

Model	R-Square	MSE	MAE	MAPE
Linear Regression	74.46%	7.9519	1.9395	9.76%
SDM	81.08%*	5.8918*	1.6176*	8.15%*

Based on Table 7, the SDM model is identified as the most appropriate model, as it yields the highest R-squared value (81.08%) and the lowest values for MSE (5.8918), MAE (1.6176), and MAPE (8.15%). These results indicate that the SDM provides the best model fit and predictive accuracy among the compared models.

6. Best Model Interpretation

The model interpretation was carried out on the optimal model, namely the spatial durbin model. If the hourly wage in *i*-th province increases by 1 unit while keeping other variables constant, the hourly wage in that province will increase by 0.31242 thousand rupiahs. However, the hourly wage will decrease by 0.22693 thousand rupiahs if the poverty gap index increases by 1 unit in all directly bordering regions. If the school expectancy in *i*-th province increases by 1 unit while keeping other variables constant, the hourly wage in that province will decrease by 0.01894 thousand rupiahs, but it will increase by 0.82679 thousand rupiahs if the school expectancy increases by 1 unit in all directly bordering regions. If the GRDP per capita in *i*-th province increases by 1 unit while keeping other variables constant, the hourly wage in that province will increase by 7.617×10^{-5} thousand rupiahs, and it will further increase by 3.636×10^{-5} thousand rupiahs if the GRDP per capita increases by 1 unit in all directly bordering regions. If the percentage of the poor population in *i*-th province increases by 1 unit while keeping other variables constant, the hourly wage in that province will decrease by 0.16271 rupiahs, but it will increase by 0.2974 rupiahs if the percentage of the poor population increases by 1 unit in all directly bordering regions. If the hourly wage in provinces directly bordering *i*-th province increases by 1 unit while keeping other variables constant, the hourly wage in *i*-th province will increase by 0.39941 thousand rupiahs.

These results align with several studies on wage disparities. Dorjnyambuu & Galambosné Tiszberger (2024) also found in their study that education is one of the factors influencing wages. However, wages earned by people with different education levels contribute to different income classes. Moreover, the impact of education factors varies across regions. This can be linked to other factors such as migration rates and the cost of living in an area. Additionally, Goh (2025) used the ARDL method to examine income inequality in ASEAN countries and found that GRDP per capita in several countries, including Indonesia, positively correlates with wages or income for the upper-middle class. However, the study also noted that the increase in GRDP per capita does not benefit all social groups equally. Therefore, income and economic inequality cannot be completely resolved even with an increase in factors affecting hourly wages, particularly in Indonesia.

D. CONCLUSION AND SUGGESTIONS

This study shows that the distribution of hourly wages in Indonesia is influenced by several economic and social factors that exhibit significant spatial effects. The spatial effect test showed that there was spatial dependence, so the Durbin spatial regression model was used to accommodate the spatial effect. In comparison with the linear regression model using Ordinary Least Squares (OLS), the Spatial Durbin Model (SDM) is proven to be more effective in capturing interregional linkages, as reflected by its better model performance.

In terms of policy implications, the findings suggest that efforts to reduce wage inequality should not be limited to individual regions. The government is encouraged to promote investment in low-wage areas, improve the equitable distribution of infrastructure, and expand access to education and job training to enhance workforce competitiveness across regions. For future research, it is recommended to explore other spatial regression models and to include additional factors, such as labor mobility and infrastructure investment, in order to obtain a more comprehensive understanding of the spatial dynamics of wage inequality in Indonesia.

ACKNOWLEDGEMENT

The authors would like to thank the Faculty of Science and Technology, Airlangga University for providing academic support and facilities throughout the research process. Appreciation is also extended to the Central Bureau of Statistics (BPS) for supplying the data used in this study.

REF ERENCES

Anselin, L. (1988). *Spatial econometrics: methods and models (Vol. 4).* Springer Science & Business Media. https://doi.org/10.2307/2290042

Anselin, L. (2005). Asymptotically optimum equidistant grouping for the normal distribution. *Scandinavian Actuarial Journal*, *2005*(3–4), 157–161. https://doi.org/10.1080/03461238.1963.10410606

Arum, P. R., Ridwan, M., Alfidayanti, I., & Wasono, R. (2024). Robust Geographically Weighted Regression With Least Absolute Deviation (Lad) Estimation and M-Estimation on Grdp of West Java Province. *Barekeng*, *18*(3), 1573–1584. https://doi.org/10.30598/barekengvol18iss3pp1573-1584

BPS. (2025). *Catalog : 1101001.* Statistik Indonesia 2020, 53,2025, 790. https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html

Breusch, T., & Pagan, A. (1979). BreuschPagan_1979.pdf. *Econometrica 74*(5), 1287–1294. https://edu.hansung.ac.kr/~jecon/art/BreuschPagan_1979.pdf%0Ahttps://opus.lib.uts.edu.au/handle/10453/13901

Cartone, A., Panzera, D., & Postiglione, P. (2022). Regional economic disparities, spatial dependence and proximity structures. *Regional Science Policy and Practice*, *14*(5), 1034–1050. https://doi.org/10.1111/rsp3.12482

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Dorjnyambuu, B., & Galambosné Tiszberger, M. (2024). The sources and structure of wage inequality changes in the selected Central-Eastern European Countries. In *Journal of Economic Inequality* (Vol. 22, Issue 4). Springer US. https://doi.org/10.1007/s10888-024-09621-0

Eeuwijk, F. Van. (2022). *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer. https://www.deep-mind.org/2023/03/26/the-universal-approximation-theorem/

Elhorst, J. P. (2014). From Cross-Sectional Data to Spatial Panels. In *Spatial Econometrics: Vol. VII*.

Fariz Fadillah Mardianto, M., Gunardi, & Utami, H. (2021). An analysis about fourier series estimator in nonparametric regression for longitudinal data. *Mathematics and Statistics*, *9*(4), 501–510. https://doi.org/10.13189/ms.2021.090409

Goh, L. T. (2025). Assessing the impact of human capital development on income inequality in the ASEAN region: an empirical analysis. *Eurasian Economic Review, 15*(3), 813-844. https://doi.org/10.1007/s40822-024-00309-4

Kasy, M. (2022). Who wins, who loses? Identification of conditional causal effects, and the welfare impact of changing wages. *Journal of Econometrics*, *226*(1), 155–170. https://doi.org/10.1016/j.jeconom.2021.02.001

López-Ureña, S., & Yáñez, D. F. (2024). Subdivision Schemes Based on Weighted Local Polynomial Regression: A New Technique for the Convergence Analysis. *Journal of Scientific Computing*, *100*(1), 1–44. https://doi.org/10.1007/s10915-024-02573-7

Moran, P. (1950). Notes on Continuous Stochastic Phenomena Published by : Biometrika Trust Stable. *Biometrika*, *37*(1), 17–23. http://www.jstor.org/stable/2332142.

Paramita, N., Masjkur, M., & Indahwati. (2021). Spatial Regression Model with Optimum Spatial Weighting Matrix on GRDP Data of Sulawesi Island. *Journal of Physics: Conference Series*, *1863*(1). https://doi.org/10.1088/1742-6596/1863/1/012045

Putra, R., Fadhlurrahman, M. G., & Gunardi. (2023). Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation. *MethodsX*, *10*(October 2022), 101994. https://doi.org/10.1016/j.mex.2022.101994

Ramli, M., Budiantara, I. N., & Ratnasari, V. (2024). Statistical inference on nonparametric regression model with approximation of Fourier series function: Estimation and hypothesis testing. *MethodsX*, *13*(June). https://doi.org/10.1016/j.mex.2024.102922

Serefoglu Cabuk, K., Cengiz, S. K., Guler, M. G., Topcu, H., Cetin Efe, A., Ulas, M. G., & Poslu Karademir, F. (2024). Chasing the objective upper eyelid symmetry formula; R2, RMSE, POC, MAE, and MSE. *International Ophthalmology*, *44*(1), 1–9. https://doi.org/10.1007/s10792-024-03157-y

Shang, Y., Lee, H., & Ma, J. (2025). A Study on the Spatial Effects of the Digital Economy on Regional Economic Growth in China. *Sustainability (Switzerland)*, *17*(5), 1–21. https://doi.org/10.3390/su17052259

Sriliana, I., Budiantara, I. N., & Ratnasari, V. (2022). A Truncated Spline and Local Linear Mixed Estimator in Nonparametric Regression for Longitudinal Data and Its Application. *Symmetry*, *14*(12). https://doi.org/10.3390/sym14122687

Suliyanto, Saifudin, T., Rifada, M., & Amelia, D. (2025). Statistical inferences and applications of nonparametric regression models based on fourier series. *MethodsX*, *14*(February). https://doi.org/10.1016/j.mex.2025.103217

Takezawa, K. (2005). *Introduction to Nonparametric Regression*. John Wiley & Sons.

Tobler, A. W. R. (1889). Clark University. *Science*, *ns-13*(332), 462–465. https://doi.org/10.1126/science.ns-13.332.462

Varlamova, J., & Kadochnikova, E. (2023). Modeling the Spatial Effects of Digital Data Economy on Regional Economic Growth: SAR, SEM and SAC Models. *Mathematics*, *11*(16). https://doi.org/10.3390/math11163516

Wu, S., Sun, Y., Jia, W., Wang, F., Lu, S., & Zhao, H. (2023). Estimation of Above-Ground Carbon Storage and Light Saturation Value in Northeastern China's Natural Forests Using Different Spatial Regression Models. *Forests*, *14*(10). https://doi.org/10.3390/f14101970

Zhong, H., Zhang, J., & Song, W. (2023). The relationship between rural information consumption and regional economic development with spatial Durbin model—A case study of Zhejiang, China. *PLoS ONE*, *18*(2 February), 1–18. https://doi.org/10.1371/journal.pone.0281405