ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 6, Issue 1, April 2026, pp. 25-36

Utilizing Geographically Weighted Regression with a Gaussian Kernel to Analyze Unemployment

Ma’rufah Hayati [a,1,*], Nora Madonna [a,2], Erica Grace Simanjuntak [a,3], Rohmatun Nikmah [b,4]

[a] Department of Actuarial Science, Institut Teknologi Sumatera, South Lampung 35365, Indonesia
[b] Department of Statistic, Nahdlatul Ulama University of Lampung, East Lmpung 34192, Indonesia
[1] marufah.mt@at.itera.ac.id *, [2] nora.madonna@at.itera.ac.id, [3] erica.simanjuntak@at.itera.ac.id, [4] rohmatunikmah003@gmail.com
* Corresponding author

ARTICLE INFO

History
Submitted: 3 November 2025
Revised: 8 March 2026
Accepted: 10 March 2026

Keywords
Optimal Bandwidth
Fixed Gaussian Kernel
Geographically Weighted
Regression Spatial Heterogeneity
Unemployment Rate
Lampung Province

ABSTRACT

Unemployment is a major challenge in economic development, reflecting an imbalance between labor supply and available job opportunities. This study aimed to examine the spatial variation of factors influencing the open unemployment rate (OUR) in Lampung Province, Indonesia, and to compare the performance of a global regression model with the geographically weighted regression (GWR) model in explaining these variations. The GWR method, using a fixed Gaussian kernel, was applied to capture spatial heterogeneity across regions. Secondary data were obtained from the Statistics Indonesia of Lampung Province in 2023, including economic growth (EG), human development index (HDI), and labor force participation rate (LFPR). The results showed that in the global regression model, LFPR was the only variable that significantly reduced unemployment, while EG and HDI were not statistically significant. The Breusch–Pagan test confirmed spatial heterogeneity, supporting the use of the GWR. The GWR model performed better, with Akaike information criterion (AIC) of 40.8262 and R^2 of 0.6059. Spatial analysis indicated that EG and HDI positively affected unemployment in several districts, suggesting limited job absorption and possible skill mismatches, whereas LFPR consistently showed a negative relationship with the open unemployment rate (OUR) across regions.

1. Introduction

Data analysis is a crucial stage in statistical research, as it aims to identify patterns, relationships, and trends from the collected data. In the context of spatial data, analysis becomes more complex due to spatial factors such as variations in characteristics across locations and geographic proximity. This aligns with the principle of spatial data, which states that everything is related to everything else, but near things are more related than distant things [1], Therefore, selecting an analytical method that can capture spatial variation is essential for producing representative and relevant insights.

One important issue requiring a spatial analysis approach is unemployment. In Lampung Province, the open unemployment rate (OUR) shows variations between regions, which can be influenced by various local social and economic factors [2]. A global statistical approach tends to be

p-ISSN 2798-253X
e-ISSN 2798-3153

inappropriate because it ignores inter-regional heterogeneity. Therefore, a model that can accommodate spatial differences in the data is needed.

Geographically weighted regression (GWR) is a regression analysis method that allows regression parameters to vary across locations, making it more adaptive to spatial variation than global regression models [3]. This method is an effective alternative for analyzing phenomena that are not geographically homogeneous. Prior research has indicated that spatial regression models are more accurate than global regression in identifying factors influencing unemployment across regions [4]. Meanwhile, another study demonstrated that using GWR) with an adaptive kernel weighting function could improve model accuracy in analyzing unemployment in Papua [5].

Based on these findings, the GWR method is considered relevant for analyzing unemployment in Lampung Province, which has geographically diverse characteristics. Therefore, this study aimed to analyze the factors influencing the open unemployment rate (OUR) in Lampung Province using the GWR model equipped with a Gaussian kernel weighting function. It is expected that the results of this study will provide a more accurate picture of unemployment conditions in each district/city and serve as a consideration in formulating more effective and targeted employment policies.

2. Method

This study aimed to analyze the factors influencing the unemployment rate in Lampung Province using a statistical approach through descriptive analysis, multiple linear regression analysis, and GWR. Before discussing further, it is important to describe the supporting theories, such as the basic concept of multiple linear regression, classical assumption testing, and the concept of GWR that considers spatial aspects. These theories serve as the foundation for understanding the relationship between dependent and independent variables and spatial variations at each observation location.

2.1. Multiple Linear Regression

Multiple linear regression is an equation model used to describe the influence between one dependent variable (Y) and two or more independent variables (X_1, X_2, ... , Xp). The main objective of this analysis is to predict the value of the dependent variable (Y) based on the known values of the independent variables [6]. In addition, this regression is also used to understand the direction of the relationship between the dependent variable and its independent variables [7]. This model assumes a linear relationship between the predictor and the dependent variable, which is usually represented [8]. The classical assumption tests applied in this study included multicollinearity tests, heteroscedasticity tests, autocorrelation tests, normality tests, and linearity tests. If all classical assumptions are met, the regression model can be considered valid [9]. Parameter estimation in a multiple linear regression model can be performed using the ordinary least squares (OLS) method, which aims to minimize the sum of squared errors.

The multiple linear regression equation model can be written as in [8]:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \dots + \beta_P X_{ip} + \varepsilon_i; i = 1,2, \dots n \tag{1}$$

where Y_i denotes value of the dependent variable in the ith observation, β_k denotes the kth regression parameter, X_{ik} denotes the value of the kth independent variable in the ith observation, and ε_i denotes error on the ith observation

2.2. Spatial Regression

The spatial regression is a statistical analysis method used to model the relationship between dependent and independent variables by considering spatial aspects. Spatial analysis is performed if the data used meet spatial requirements, namely, having correlated errors or spatial heterogeneity [10]. Spatial heterogeneity indicates that the relationship between variables can vary across locations within a geographic space.

One way to detect spatial heterogeneity is to perform the Breusch-Pagan test [11], as follows:

Hypothesis:

$H_0 = \sigma_1 = \sigma_2 = \cdots = \sigma_n$ (no spatial heterogeneity occurs)

H_1 = There is at least one $\sigma_i^2 = \sigma_j^2$ (spatial heterogeneity occurs)

Statistical test:

$$\beta P = (1\,/\,2) f^T z\,(\,z^T z\,)^{-1} z^T I. \tag{2}$$

The element of the vector f is $\left(\frac{e_i^2}{\sigma^2} - 1\right)$, e_i is the residual for the ith observation, Z is a matrix of size $n \times (p + 1)$ containing standardized vectors for each observation H_0 is rejected if $BP > X^2\,(a, p)$.

2.3. Geographically Weighted Regression (GWR)

The GWR model is the development of the global regression model, which does not take geographical aspects into account. Unlike multiple linear regression, which assumes that the relationship between dependent and independent variables is uniform across the study area, the GWR model allows regression parameters to vary at each observation location, thus capturing spatial influences more accurately. The GWR model is the development of the global regression model, where the basic idea is taken from non-parametric regression [12]. The response variable Y in the GWR model is predicted by predictor variables, each of which has a regression efficiency that depends on the location where the data are observed. The GWR model can be written as in (3) [13]:

$$y_i = \beta_0\,(U_i, V_i) + \Sigma_{k=1}^{P}\,\beta_k\,(U_i, V_i)\,x_{ik} + \varepsilon_i\,, i = 1,2, \dots\,, n \tag{3}$$

where y_i is the observed value of the dependent variable at the ith observation, x_{ik} is the observed value of the ith independent variable at the observation location, $\beta_0\,(U_i, V_i)$ is the constant/intercept at the ith observation, U_i, V_i states the geographic coordinates of the ith observation location, $\beta_k\,(U_i, V_i)$ is the regression coefficient of the kth independent variable at the observation location ith, and ε_i is an error at the ith observation assumed to be independently and identically distributed with mean zero and a constant variance of σ^2 [14].

2.4. Gaussian Kernel Weight Matrix

In the GWR model, weighting is applied to account for spatial proximity between locations. Observations closer to a given location will have a greater influence on the parameter estimation process at that location. To control this influence, a kernel function is used, and one commonly used to be the Gaussian kernel function. The Gaussian kernel function has the form:

$$w_{ij} = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \tag{4}$$

where w_{ij} denotes the weight between location i and j, d_{ij} denotes the Euclidean distance between location i and j, and b denotes Bandwidth parameter that controls how quickly the weights decrease with distance [15].

The weights w_{ij} form a weight matrix $W\,(U_i, U_i)$ for each location i, where (U_i, U_i) are the coordinates of location i. This weight matrix is diagonal, with diagonal elements being the weights for each observation location relative to location i, while non-diagonal elements are zero.

$$w_{(U_i, U_i)} = \begin{bmatrix} w_{ij} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix}. \tag{5}$$

This weight matrix is used to estimate the regression parameters at each location using a local regression approach using the weighted least squares method. Thus, each location has a different regression model according to its spatial conditions. The use of a fixed Gaussian kernel means that each location has a weighted influence determined by a fixed distance (constant bandwidth), making it suitable for use when data distribution is relatively even across regions.

2.5. Analysis Flow and Research Variables

After determining the appropriate estimation model, the next step was to outline the research analysis flow. This flow describes the systematic stages, starting from data collection and processing, selecting the analytical method, testing assumptions, to interpreting the results. The purpose of developing the analysis flow was to ensure that each stage was carried out in a structured and consistent manner in accordance with the research objectives. After determining the appropriate estimation model, the next step was to outline the research analysis flow; describe the response variable and the predictor variables related to unemployment cases in Lampung Province; and analyze the classical linear regression model, or ordinary linear regression (OLR), for unemployment cases in Lampung Province. To analyze the GWR model for unemployment cases in Lampung Province, the following steps were conducted.

a. Determine the longitude and latitude coordinates of each city/regency in Lampung Province.
b. Calculate the Euclidean distance between cities/regencies in Lampung Province.
c. Determine the bandwidth based on the minimum cross-validation (CV) criterion.
d. Calculate the weighting matrix for each city/regency using a kernel function.
e. Estimate the GWR parameters using the optimal bandwidth.
f. Compare the Akaike information criterion (AIC) values of the global regression and GWR fixed kernel Gaussian models.
g. Interpret and draw conclusions from the results obtained.

The data for this study were obtained from the 2023 publication available on the official website of the Statistics Indonesia of Lampung Province. The study employed four main variables, consisting of one dependent variable and three independent variables. The details of these variables are presented in Table 1.

Table 1. Research Variables

Variable	Description	Abbreviation
Y	Open unemployment rate (%)	OUR
X1	Economic growth (%)	EG
X2	Human development index	HDI
X3	Labor force participation rate (%)	LFPR

Table 1 shows that OUR (Y) as the response variable is explained by EG (X_3), HDI (X_2), and labor force participation rate (LFPR) (X_3). The selection of these variables was based on [4], and considers the availability of data at the regency/municipality level.

3. Result and Discussion

After conducting data analysis using multiple linear regression and spatial regression methods, this section presents the research findings. The discussion begins with a spatial data exploration to examine the geographical distribution of variables such as OUR variable Y, Economic growth (EG) variable (X_1), HDI (human development index) variable (X_2), and LFPR variable (X_3). Subsequently, the results of applying the GWR model with a Gaussian kernel weighting function are presented, along with a comparison to the global regression model to assess the model's ability to capture spatial variations.

3.1. Data Exploration

The data exploration stage was carried out to understand the characteristics and distribution patterns of OUR, EG, HDI, and LFPR variables in Lampung Province. Through descriptive statistical analysis and map-based visualization of Lampung's regions, spatial patterns, outliers, and preliminary relationships between variables could be identified. These findings formed the basis for applying the GWR model with a Gaussian kernel weighting function in analyzing unemployment levels. The maps for each variable at the regency/municipality level are presented in Fig. 1 until Fig. 4.

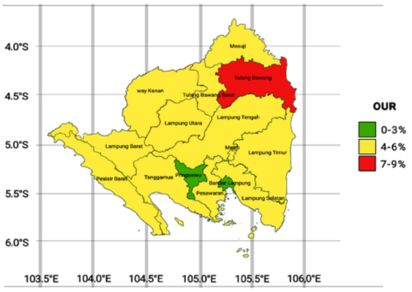

Fig 1. Map of OUR variable percentage (Y).

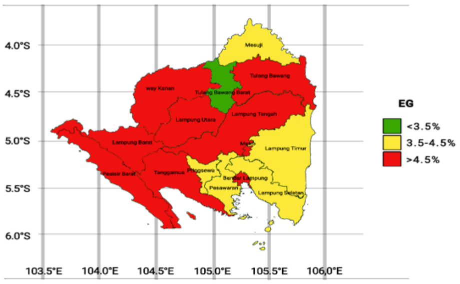

Fig 2. Map of EG variable percentage (X_1).

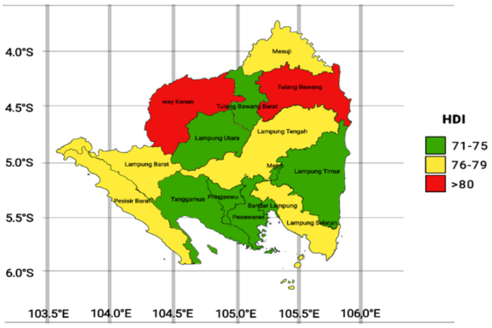

Fig 3. Map of HDI variable percentage (X_2).

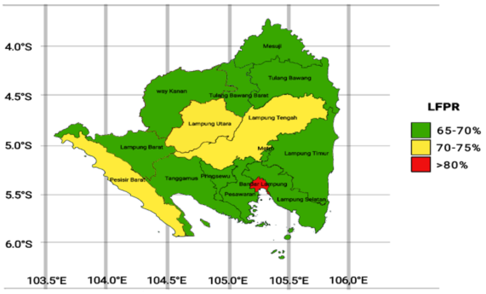

Fig 4. Map of LFPR variable percentage (X_3).

Fig. 1 shows that OUR (Y) reached its highest value in Tulang Bawang (7–9%) and the lowest in Bandar Lampung and Pringsewu (0–3%), while most regions fell within moderate category (4–6%). Fig. 2 indicates that EG (X_1) reached its highest value (> 4.5%) in several regions such as Pesisir Barat and Bandar Lampung, while the lowest values (< 3.5%) were found in Tulang Bawang Barat, and moderate in other areas. Fig. 3 shows that HDI (X_2) reached its highest value (> 80) in Tulang Bawang and Way Kanan and the lowest (71–75) in several regions including Lampung Utara and

Tanggamus, while other regions fell within medium category. Fig. 4 indicates that LFPR (X_3) reached its highest value (>80%) in Bandar Lampung and the lowest (65–70%) in many districts, while the remaining regions fall within the moderate range (70–75%).

3.2. Global Regression

Prior to implementing the GWR model, a global regression analysis was first carried out as a comparison to examine the relationship between the dependent and independent variables while disregarding spatial effects. The outcomes of the global regression estimation are shown in Table 2.

Table 2. Global Regression Model Estimation Results

Variable	Coefficient	P-value
Variable intercept (Constant)	-0.21268	0.00922
EG variable (X_1)	0.86789	0.43467
HDI variable (X_2)	0.07015	0.43467
LFPR variable (X_3)	-0.21268	0.00922**

The global regression estimation results in Table 2 showed that only the LFPR (X_3) had a significant effect on the OUR (Y) with a p-value of <0.01, whereas EG (X1) and HDI (X2) are not statistically significant (p-value > 0.05). Next, a multicollinearity test was conducted using the variance inflation factor (VIF) to ensure there was no strong linear relationship between the independent variables. A relationship between predictor variables in a multiple linear regression model is indicated when the VIF value is greater than 10 [16]. The results of the VIF test are presented in Table 3.

Table 3. Multicollinearity Test Results (VIF)

Variable	Coefficient
EG Variable (X_1)	1.285788
HDI Variable (X_2)	1.346917
LFPR Variable (X_3)	1.090023

The test results in Table 3 show that all variables yielded a VIF value of < 10, indicating no serious multicollinearity. In conclusion, globally, only LFPR variable (X_3) significantly influenced unemployment. However, this analysis did not consider spatial factors, so the GWR method was used to capture spatial variation in the relationships between variables.

3.3. Spatial Heterogeneity Test

A spatial heterogeneity test is necessary to determine the presence of spatial diversity in observations. Spatial heterogeneity can result in different regression parameters for each observation location [17]. To identify spatial diversity, the Breusch–Pagan test can be used to examine whether the variance of the residuals from the global regression model is constant or varies across observations. In this study, the Breusch–Pagan test was performed using the *R* statistical software with the lmtest package, evaluating the presence of heteroscedasticity based on the squared residuals of the global regression model. The results of the Breusch–Pagan analysis are presented Table 4.

Table 4. Breusch-Pagan test

Bruch-Pagan (BP)	df	p-value
9.0119	3	0.02913

Based on Table 4, with a significance level of $\alpha = 5\%$, the obtained p-value = 0.02913, which is smaller than $\alpha = 0.05$. Therefore, it can be concluded that there was spatial heterogeneity, indicating that the GWR model was appropriate for data analysis.

3.4. Determination of Optimal GWR Bandwidth

In GWR analysis, bandwidth is a key parameter that defines the spatial extent of estimation by weighting observations based on distance. Selecting an appropriate bandwidth is essential, as it influences the model's ability to capture spatial variation. The optimal bandwidth is typically determined by minimizing the CV score, where a smaller CV indicates better predictive performance.

In this study, the optimal bandwidth was obtained through an iterative process using a fixed Gaussian kernel, with selection based on minimizing the AIC. The procedure was implemented in *R* using a GWR package, which automatically identified the optimal bandwidth. Although determined iteratively, only the final result is reported in Table 5.

Table 5. Optimal Bandwidth Selection Results in GWR Model

CV	Bandwidth
27.20537	1.992522

Based on Table 5, the optimal bandwidth was 1.992522 (CV = 27.20537), indicating the best performance in capturing spatial variation. This bandwidth improved the model's estimation accuracy and enabled a better representation of the relationship between Y and X_1, X_2, and X_3 *i*n the Lampung region.

3.5. Fixed Kernel Gaussian

The choice of kernel function affects GWR estimation results. The fixed Gaussian kernel used a constant bandwidth of 1.992522, giving each location the same spatial influence and ensuring consistent analysis, especially for relatively uniform data distributions. Before GWR, global regression showed that only X_3 was significant ($p < 0.01$), while X_1 and X_2 were not ($p > 0.05$), with $R^2 = 0.5959$. After applying GWR, model performance improved with R^2 increasing to 0.6059, indicating a better explanation of variation in Y. Overall, GWR with a fixed Gaussian kernel better captures spatial variation than global regression and is suitable for data with relatively uniform spatial distribution. In this study, the evenly distributed observation points across districts in Lampung Province support the application of this method. Therefore, GWR is effective for analyzing local relationships and supporting spatial-based decision-making.

3.6. Selection of the Best Model Between the Global Regression Model and the GWR Model

Model selection is an evaluation process to determine the most appropriate model in explaining the relationships between variables based on the data [18]. In this study, a comparison between global regression and GWR using the fixed Gaussian kernel approach was conducted based on AIC and R^2 values. Prior to GWR, global regression was estimated as a baseline, showing that only X_3 was statistically significant ($p < 0.01$), while X_1 and X_2 were not significant ($p > 0.05$). After applying GWR with a fixed bandwidth of 1.992522, the comparison results of both models are presented in Table 6.

Table 6. Best Model Selection Results

Model	AIC	R^2
Global Regression	46.94093	0.5959
GWR Fixed	40.8262	0.6059

Based on Table 6, the GWR fixed model produced a lower AIC (40.8262) compared to the global regression model (46.94093) and a higher R^2 (0.6059 vs 0.5959). This indicates that the GWR model provides a better fit and is more capable of capturing spatial variation. Therefore, the fixed GWR model is more appropriate for explaining the relationship between Y, X_1, X_2, and X_3 in a spatial context [18].

3.7. Test of of Significance of Parameters for Each Location

Based on the *t*-test results in the GWR model, the statistical significance of each variable at each location was determined using the critical value of $t > 2$ or $t < -2$ at a certain confidence level. In the GWR model, each location had its own R^2 or often referred to as the local R^2 value [19], [20]. The *t*-test results showed variations in the significance of the coefficients at each location, reflecting the importance of considering spatial variance in the analysis. Summary of the *t*-test results for each variable can be seen in Table 7.

Table 7. Estimated Regression Parameters for Each Regency Using the GWR Method

Regency	*t*-Intercept	*t*-X_1	*t*-X_2	*t*-X_3
Lampung Barat	1.187637	1.422503	0.8337382	-3.097390
Tanggamus	1.274956	1.427955	0.7279492	-3.110620
Lampung Selatan	1.136875	1.426853	0.8695559	-3.064326
Lampung Timur	1.240757	1.396661	0.8168363	-3.148636
Lampung Tengah	1.210679	1.392876	0.8482455	-3.139014
Lampung Utara	1.275928	1.396623	0.7811305	-3.160086
Way Kanan	1.351602	1.311775	0.8022263	-3.268.009
Tulang Bawang	1.220063	1.402160	0.8294441	-3.134141
Pesawaran	1.198054	1.430776	0.8134967	-3.092687
Pringsewu	1.272707	1.436630	0.7086400	-3.089572
Mesuji	1.212439	1.424157	0.8086718	-3.106380
Tulang Bawang Barat	1.213705	1.439614	0.7842909	-3.086403
Pesisir Barat	1.294612	1.362578	0.8046565	-3.203361
Bandar Lampung	1.297993	1.371549	0.7905338	-3.195025
Metro	1.332462	1.384261	0.7324272	-3.184316

Interpretation of *t*-test outcomes based on Table 7for each variable can be described as follows.

3.7.1. Intercept (Constant)

The *t*-value for the intercept across all districts ranged from 1.13 to 1.35. Since this value is still below the threshold of 2, the constant is not statistically significant in any location. This means that without considering other variables, the average response value is not significantly different from zero across all regions.

3.7.2. EG Variable X_1

The *t*-value for X_1 ranged from 1.31 to 1.44. All *t*-values are still below 2, so X_1 has not shown a statistically significant effect on the response variable in all locations. However, *t*-values approaching 2 in some regions (e.g., Pringsewu and Tulang Bawang Barat) indicate that X_1 may be starting to show a positive trend in influencing the response in those areas.

3.7.3. HDI Variable X_2

The *t*-value for X_1 ranged from 0.70–0.87, indicating that the effect of X_2 is not significant in all districts. This means that this variable does not contribute significantly to explaining the variation in the response variable across locations.

3.7.4. LFPR Variable X_3

The *t*-values for X_3 ranged from -3.06 to -3.26, which are below -2. Thus, variable X_3 has a significant negative effect across all locations. This means that the larger the value of X_3, the smaller the value of the response variable in each district. This consistent negative direction indicates that the influence of X_3 on the response is strong and stable across all regions.

3.8. Estimated Regression Parameters for Each Location

The GWR model allows for different regression coefficients across locations, providing more accurate results in understanding the patterns of variable relationships across regions. This section will discuss the GWR regression coefficients for each district in Lampung Province. The parameters used in the model included the intercept, EG variable (X_1), HDI variable (X_2), and LFPR variable (X_3). The Table 8 shows the estimated GWR regression parameters for each district.

Table 8. Estimated Regression Parameters for Each Regency Using the GWR Method

Regency	Intercept	X_1	X_2	X_3
Lampung Barat	9.705.274	0.8968344	0.07308131	-0.2115983
Tanggamus	10.427.767	0.9012736	0.06384040	-0.2127213
Lampung Selatan	9.314.518	0.9020416	0.07645195	-0.2098235
Lampung Timur	10.132.996	0.8795227	0.07150320	-0.2149608
Lampung Tengah	9.896.930	0.8777890	0.07433964	-0.2143930
Lampung Utara	10.420.184	0.8795046	0.06835697	-0.2158297
Way Kanan	11.072.123	0.8311007	0.07046875	-0.2241770
Tulang Bawang	9.966.689	0.8832210	0.07264058	-0.2139946
Pesawaran	9.784.782	0.9019254	0.07127692	-0.2112104
Pringsewu	10.423.774	0.9084955	0.06225411	-0.2115104
Mesuji	9.900.304	0.8973123	0.07082210	-0.2120931
Tulang Bawang Barat	9.910.436	0.9077968	0.06871801	-0.2107596
Pesisir Barat	10.580.051	0.8590002	0.07045916	-0.2189426
Bandar Lampung	10.606.089	0.8643113	0.06920130	-0.2183663
Metro	10.905.389	0.8731182	0.06418381	-0.2180626

Based on Table 8, coefficients varied across districts. EG (X_1) and HDI (X_2) had positive effects on unemployment, indicating growth and HDI did not always reduce unemployment. Meanwhile, LFPR (X_3) had a negative effect, meaning higher participation lowers unemployment. This confirms spatial variation, so region-specific policies are needed. The results can be used to form regression models for each district, such as Lampung Barat.

$$\hat{y}_{Lampung\ barat} = 9.705274 + 0.8968344\ (X_1) + 0.07308131(X_2) - 0.2115983(X_3)$$

The final estimated model represents the predicted value of the dependent variable for Lampung Barat obtained from the GWR estimation. Therefore, the error component is not included in the final equation, as the model expresses the estimated relationship between the dependent and independent variables at that specific location. Interpretation of intercept (constant) showed that when economic growth, the HDI, and LFPR were zero, then the OUR in Lampung Barat was estimated at 9.705%. This is the baseline unemployment rate without the influence of explanatory variables.

Interpretation of EG variable (X_1) indicates that every 1-unit increase in economic growth in Lampung Barat will increase the OUR by 0.897 percentage points, assuming other factors remain constant. This positive trend indicates that economic growth is not yet inclusive, perhaps only occurring in sectors that do not absorb a large workforce (e.g., modernized agriculture or wholesale trade).

Interpretation of HDI variable (X_2) indicates that every 1-unit increase in the HDI will increase the OUR by 0.073, assuming other factors remain constant. The positive trend indicates that an increase in the HDI does not automatically reduce unemployment. This could indicate that improvements in education and quality of life have not been matched by the availability of jobs that match the population's competencies.

Interpretation of LFPR variable (X_3) is that every 1-unit increase in LFPR (employment allowance) reduces the OUR (employment allowance) by 0.212 percentage points. This negative coefficient aligns with theory that the more people actively working or looking for work, the lower unemployment will be due to increased labor absorption. Fig. 5 until Fig. 8 exhibits map of the GWR regression coefficients for each variable.

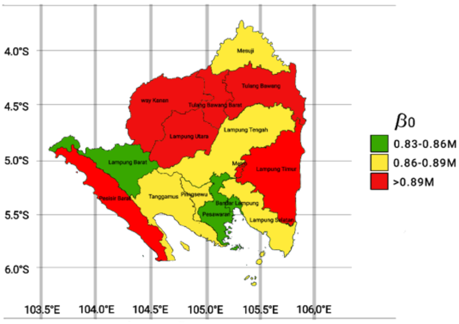

Fig 5. Coefficient of intercept (β_0).

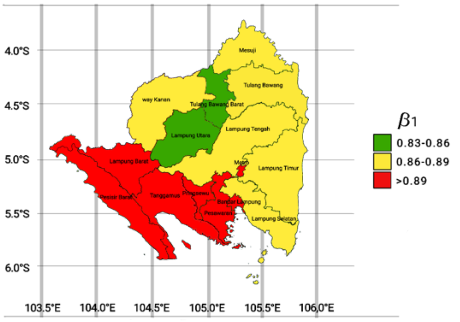

Fig 6. Coefficient of the EG variable (β_1).

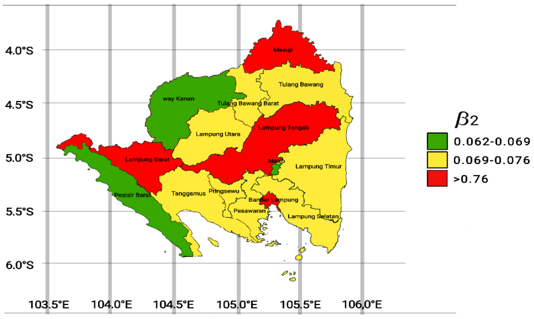

Fig 7. Coefficient of the HDI variable (β_2).

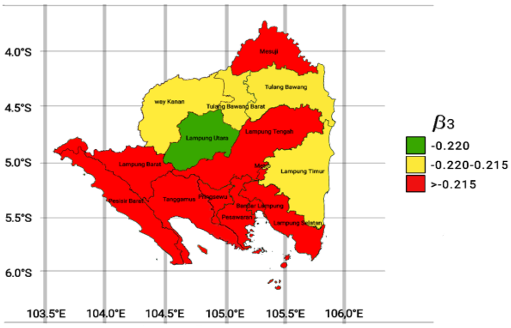

Fig 8. Coefficient of the LFPR variable (β_3).

Spatial analysis using the GWR model was conducted to examine variation in the influence of socioeconomic factors on the OUR in Lampung Province. The results show spatial heterogeneity, meaning the relationships between explanatory variables and unemployment differ across regions.

Fig. 5 shows that central and southern areas such as Lampung Tengah, Lampung Timur, Lampung Selatan, and Pesawaran had higher coefficients (>10.4%), indicating higher unemployment, while western and northern regions such as Lampung Barat and Way Kanan had lower coefficients (<9.3%), indicating lower unemployment.

Fig. 6 indicates that the effect of economic growth varies spatially. Higher coefficients (>0.69%) in southern regions suggest a positive relationship between growth and unemployment, meaning growth has not been fully inclusive. Lower coefficients (0.33–0.49%) in areas such as Lampung Utara and Way Kanan indicate a more moderate effect in reducing unemployment.

Fig. 7 shows that the impact of HDI also varies. Higher coefficients (>0.076%) in regions such as Lampung Tengah and Tanggamus indicate that increased HDI may coincide with higher unemployment, possibly due to limited job absorption, while lower coefficients (0.062–0.069%) indicate a better relationship in reducing unemployment.

Fig. 8 shows that LFPR consistently has a negative relationship with unemployment across all regions. Coefficients below –0.215 indicate that increasing labor force participation significantly reduces unemployment.

Overall, GWR results confirm that the effects of economic growth, HDI, and LFPR on unemployment are spatially heterogeneous. Therefore, policies to reduce unemployment should be region-specific and based on local socioeconomic condition.

4. Conclusion

The analysis combining global regression and GWR provides several important conclusions. The global regression results indicate that only the LFPR significantly affects the OUR, with a negative coefficient, while EG and HDI are not statistically significant at the global level. However, the Breusch–Pagan test confirms the presence of spatial heterogeneity, suggesting that regression parameters vary across regions and that a global model may not adequately capture local variations. The GWR model using a fixed Gaussian kernel demonstrates better performance (AIC = 40.8262; $R^2 = 0.6059$) compared to the global regression model (AIC = 46.9409; $R^2 = 0.5959$), indicating its stronger ability to explain spatial variation. The spatial distribution maps further reveal that economic growth generally has a positive effect on unemployment in most districts, implying that economic expansion may not be sufficiently labor-intensive. Similarly, HDI shows a positive coefficient in several areas, suggesting that improvements in education and living standards are not always aligned with job creation. In contrast, LFPR consistently exhibits a negative coefficient across all regions, confirming its important role in reducing unemployment. Specifically, in West Lampung District, LFPR significantly reduces unemployment, while EG and HDI tend to increase it, reflecting a structural mismatch between economic development and labor market conditions. Overall, the findings confirm that the relationship between unemployment and its determinants is spatially heterogeneous; therefore, policy interventions should be region-specific. A place-based economic policy that accounts for local characteristics, labor structures, and development capacity is recommended to effectively reduce unemployment disparities across Lampung Province.

References

[1] R.J. Djami and G. Haumahu, “Pemodelan indeks pembangunan manusia di Provinsi Maluku dan Maluku Utara dengan menggunakan spatial autoregressive model (SAR),” *VARIANCE J. Stat. Its Appl.*, vol. 5, no. 1, pp. 37–46, Apr. 2023, doi: 10.30598/variancevol5iss1.

[2] A.N.M. Lumi and G.N. Sari, “Analisis tingkat pengangguran terbuka di Kota-Kota Pulau Sulawesi,” *J. Pembang. Ekon. Keuang. Drh,* vol. 26, no. 1, pp. 74–88, Dec. 2025, doi: 0.35794/jpekd.65781.26.1.2025.

[3] Y. Dai et al., “Geographically Weighted regression enhances spectral diversity – biodiversity relationships in inner Mongolian Grasslands,” *Diversity*, vol. 17, no. 8, 2025, Art. no 541, doi: 10.3390/d17080541.

[4] R. Ilahi, “Model spatial autoregressive pengangguran Provinsi Kepulauan Bangka Belitung,” *Sem. Nas.l Off. Stat.*, vol. 2021, no. 1, pp. 517–526, Nov. 2021, doi: 10.34123/semnasoffstat.v2021i1.955.

[5] Y. Taek, R.D. Bekti, and K. Suryowati, “Penerapan model geograpgically weighted regression (GWR) menggunakan fungsi pembobot adaptive kernel Gaussian dan adaptive kernel bisquare padatingkat pengangguran terbuka di Pulau Papua,” *J. Stat. Ind. Komputasi*, vol. 8, no. 2, pp. 84–101, Jul. 2023, doi: 10.34151/statistika.v8i2.4459.

[6] S.L. Khoiriah and W.U. Dewi, “The impact of tobacco production and cultivated area on the GRDP of the agricultural sector in Lampung Province: A multiple linear regression analysis,” Symmetry Sigma J. Math. Struct. Stat. Patterns, vol. 1, no. 2, pp. 165–182, Dec. 2024, doi: 10.58989/symmerge.v1i2.28.

[7] M. Muhidin and T.L. Situngkir, “Pengaruh rasio profitabilitas terhadap harga saham perusahaan perbankan yang terdaftar di Bursa Efek Indonesia pada tahun 2015 - 2021,” Transform. Manageria J. Islamic Edu. Manag., vol. 3, no. 1, pp. 15–27, Sep. 2022, doi: 10.47467/manageria.v3i1.2093.

[8] U.N. Faiza and M. Hayati, “Factors influencing the human development index (HDI) in Lampung Province: A multiple linear regression analysis,” *Symmetry Sigma J. Math. Struct. Stat. Patterns*, vol. 1, no. 2, pp. 148–164, 2024, doi: 10.58989/symmerge.v1i2.27.

[9] A.A. Khoiriah and R. Azizah, “The effect of minimum wage, population, and human development index on the open unemployment rate in Lampung Province,” *Symmetry Sigma J. Math. Struct. Stat. Patterns*, vol. 1, no. 2, pp. 132–147, Dec. 2024, doi: 10.58989/symmerge.v1i2.26.

[10] Marsono, “Deteksi spasial pada model indeks ketimpangan gender Indonesia,” *BUANA GENDER J. Stud. Gend. Anak*, vol. 6, no. 1, Jan.–Jun. 2021, doi: 10.22515/bg.v6i1.3482.

[11] F. Mala and M.F. Hidayat, “Pemodelan indeks pembangunan manusia Nusa Tenggara Barat menggunakan geographically weighted regression,” *Euler J. Ilm. Mat. Sains Tekno.*, vol. 11, no. 2, pp. 339–350, Dec. 2023, doi: 10.37905/euler.v11i2.23042.

[12] W. Kang and T.M. Oshan, “Scale and correlation in multiscale geographically weighted regression (MGWR),” *J. Geogr. Syst.*, vol. 27, no. 3, pp. 399–424, 2025, doi: 10.1007/s10109-025-00468-1.

[13] A. Sulekan and S.S.S. Jamaludin, “Review on geographically weighted regression (GWR) approach in spatial analysis,” *Malays. J. Fundam. Appl. Sci.*, vol. 16, no. 2, pp. 173–177, Mar.–Apr. 2020, doi: 10.11113/mjfas.v16n2.1387.

[14] P.F. Utami, A. Rusgiyono, and D. Ispriyanti, “Pemodelan semiparametric geographically weighted regression pada kasus pneumonia balita Provinsi Jawa Tengah,” *J. Gaussian*, vol. 10, no. 2, pp. 250–258, 2021, doi: 10.14710/j.gauss.v10i2.30945.

[15] K. Septiani, “Pemodelan GWR (geographically weighted regression) menggunakan pembobot fixed Gaussian kernel dan fixed tricube kernel (Studi kasus: prevalensi stunting di Provinsi Jawa Timur tahun 2017),” B.S. thesis, Fakultas Mat. Sains., Univ. Brawijaya, Malang, Indonesia, 2020.

[16] K. Puteri and A. Silvanie, “Machine learning untuk model prediksi harga sembako,” *J. Nas. Inform.*, vol. 1, no. 2, pp. 82–94, 2020.

[17] R. Amalah, A.K. Jaya, and N. Sirajang, “Pemodelan geographically weighted logistic regression dengan metode ridge,” *ESTIMASI J. Stat. Its Appl.*, vol. 4, no. 2, pp. 130–143, 2023, doi: 10.20956/ejsa.v4i2.12250.

[18] D.W.S. Yusuf, E.M.P. Hermanto, and W. Pramesti, “Pemodelan geographically weighted regression (GWR) pada persentase kriminalitas di Provinsi Jawa Timur Tahun 2017,” *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, pp. 156–163, 2020, doi: 10.29244/ijsa.v4i1.557.

[19] M.N. Lessani and Z. Li, “SGWR: Similarity and geographically weighted regression,” *Int. J. Geogr. Inf. Sci.*, vol. 38, no. 7, pp. 1232–1255, 2024, doi: 10.1080/13658816.2024.2342319.

[20] C. Feng, Y. Leung, Q. Wang, and Y. Zhou, “Spatial non-stationarity test of regression relationships in the multiscale geographically weighted regression model,” *Spat. Stat.*, vol. 62, 2024, Art. no 100846, doi: 10.1016/j.spasta.2024.100846.