JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 3, July 2024, pp. 937-949

Implementation of Gamma Regression and Gamma Geographically Weighted Regression on Case Poverty in Bengkulu Province

Ilham Alifa Azagi[1], I Made Sumartejaya[1*], Asep Saefuddin[1]
[1]Department of Statistics, IPB University, Indonesia
imsjaya@apps.ipb.ac.id

Article History:
Received : 20-04-2024
Revised : 23-06-2024
Accepted : 25-06-2024
Online : 19-07-2024

Keywords:
Spatial Analysis;
Poverty;
Gamma Regression;
Gamma Geographically;
Weighted Regression.

ABSTRACT

Spatial analysis involves leveraging spatial references inherent in the data being analyzed. The method to be used in spatial analysis is the Geographically Weighted Regression (GWR) method. GWR is an extension of the linear regression model at each location by adding a weighting function to the model. Generally, the GWR model uses residuals with a normal distribution in its analysis. One distribution that can be used is the gamma distribution. With the development of methods in statistics, when a response variable follows a gamma distribution, analysis is performed using Gamma Regression (GR). GR analysis is conducted because the response variable meets the gamma distribution assumption. One method used for spatial effects with a gamma-distributed response variable is the Gamma Geographically Weighted Regression (GGWR) method. In 2022, Bengkulu Province was among the ten poorest provinces in Indonesia. Therefore, the main objective is to compare the GR and GGWR models and analyze the factors affecting poverty in Bengkulu Province using these models. The results of this study show that the GR model has an R^2 accuracy of 87.93%, while the GGWR model has an R^2 accuracy of 95.87%. This indicates that the best model for the analysis is the GGWR. An example of the GGWR model equation for poverty in Bengkulu Province is $Y = \exp(-6.039 + 3.15 \times 10^{-6}X_1 - 0.055X_2 + 0.156X_4 - 0.00021X_5 + 0.004X_7 - 0.021X_8 - 0.006X_9 + 4.794 \times 10^{-5}X_{10})$. The factors influencing the GGWR model in Bengkulu Province are Population, Life Expectancy, Average Years of Schooling, Adjusted Per Capita Expenditure, School Participation Rate, Per Capita Expenditure on Food, Households Receiving Rice for the Poor, and Gross Regional Domestic Product. The benefit of this research is to serve as a reference for the provincial government of Bengkulu regarding the variables that influence poverty. It is expected that this will help the government reduce the poverty rate in Bengkulu Province.

Crossref
https://doi.org/10.31764/jtam.v8i3.22930

A. INTRODUCTION

Spatial statistics, also known as geostatistics, involves techniques used to analyze and model spatial data (Rash et al., 2024). Spatial data pertains to information linked with geographical locations and finds wide-ranging applications in various domains. Geospatial information can be categorized as data representing specific points or areas, and any information associated with spatial details, collectively falling under the umbrella of saptial data. It typically includes location coordinates (logitude, latitude) and associated variables (Yamagata & Seya, 2019), which are essential for subsequent spatial analysis.

In spatial analysis, researchers commonly employ Geographically Weighted Regression (GWR) models and global spatial econometric models to explore spatial heterogeneity and spatial dependence separately (Zhao et al., 2022). GWR extends the traditional regression approach by incorporating spatial coordinates into the model, thereby improving the understanding of local spatial effects and addressing issues like spatial autocorrelation (Isazade et al., 2023; Tang et al., 2022). Gamma Regression (GR) is another statistical method used when response variables exhibit a gamma distribution pattern, offering advantages in modeling asymmetric data distributions (Dewi, 2019; Dunder et al., 2016). It belongs to the generalized gamma distribution family and provides flexible modeling frameworks that can better approximate certain types of data distributions compared to more standard approaches.

Research has explored various applications of GWR, including its use in analyzing poverty levels in specific regions like Central Java Province (Agustina et al., 2015). Studies have also extended this framework to include spatial effects, resulting in methods like Gamma Geographically Weighted Regression (GGWR), which remains less explored but holds promise in enhancing spatial modeling capabilities (Putri et al., 2017). Indonesia continues to face significant challenges with poverty, impacting various aspects of daily life, particularly among vulnerable groups such as children (Yasin et al., 2015). Recent statistics from the Central Statistics Agency (CSA) highlight persistent poverty rates, with provinces like Bengkulu consistently among the poorest (CSA, 2022). This research aims to investigate and analyzing influential factors poverty in Bengkulu Province using both GR and GGWR models. By comparing these approaches, the study seeks to enhance understanding of spatial variations in poverty and contribute insights that could inform more effective policy interventions.

B. METHODS

1. Gamma Regression (GR)

The probability density function, the expected value and the variance of the gamma distribution are obtained as follows: $E(Y) = \mu = \alpha\theta$ and $Var(Y) = \alpha\theta^2$. Suppose $Y_i \sim gamma(\alpha, \theta)$, where $i = 1, \dots, k$ are random sample sizes. The mean form in GR is as follows:

$$g(\mu_i) = \phi_i = x_i^T \beta \tag{1}$$

In the GR setup, with g as the link function, the vector of regression mean parameters is represented as $\beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_k]^T$, where x_i denotes the value vector of predictor variables, and ϕ_i stands linear predictor. Here, $g(.): (0, \infty) \rightarrow \mathbb{R}$ signifies a real-valued function. Common mean link functions in GR include the logarithm function, $g(\mu) = \ln(\mu)$; the identity function, $g(\mu) = \mu$, and the inverse function, $g(\mu) = \frac{1}{\mu}$. The inverse function is the canonical link mean in the general linear model (Bossio & Cuervo, 2015). The model used in GR is log-linear, thus yielding a positive expected value. The model is as follows:

$$g(\mu_i) = \ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \tag{2}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) \tag{3}$$

2. Gamma Geographically Weighted Regression (GGWR)

GGWR is an extension of gamma regression, with the key difference in the weighting that incorporates the latitude and longitude coordinates of the observed data points. In the GGWR model, the response variable (Y) following a gamma distribution is influenced by predictor variables (X) whose coefficient values are affected by the geographical location based on latitude and longitude coordinates. The GGWR model yields locally parameterized parameter estimators. This implies that the formed regression applies only to specific locations. The probability density function of GGWR, where $Y_i \sim gamma(\alpha, \theta(u_i, v_i))$ is as follows (Putri et al., 2017):

$$f(y_i) = \begin{cases} \dfrac{1}{\left(\frac{\exp\left(x_i^T \beta(u_i,v_i)\right)}{\alpha}\right)^{\alpha} \Gamma(\alpha)} y_i^{\alpha-1} e^{-\frac{y_i}{\left(\frac{\exp\left(x_i^T \beta(u_i,v_i)\right)}{\alpha}\right)}} & , y > 0 \text{ and } \alpha > 0 \\ 0 & , \text{ others } y \end{cases}$$

$$= \begin{cases} \dfrac{\left(\frac{\alpha}{\exp\left(x_i^T \beta(u_i,v_i)\right)}\right)^{\alpha} y_i^{\alpha-1} \exp\left(\frac{-y_i \alpha}{\exp\left(x_i^T \beta(u_i,v_i)\right)}\right)}{\Gamma(\alpha)} & , y > 0 \text{ and } \alpha > 0 \\ 0 & , \text{others } y \end{cases} \quad (4)$$

3. Data Sources

The research relies on secondary data from the Central Statistics Agency (CSA) and data analysis uses program R. The dataset compares poverty data as the response variable and ten predictor variables, including population, life expectancy, literacy rate, average years of schooling, adjusted per capita expenditure, Human Development Index (HDI), school enrolment rate, per capita expenditure on food and housing, poor rice recipient household, and Gross Regional Domestic Product (GRDP). The data pertains to ten districts/cities within Bengkulu Province and is structured as panel data spanning the years 2015 to 2022, as shown in Table 1.

Table 1. Variables Used

Variable	Information	Unit
Y	Poor people	Percent
X_1	Population	People
X_2	Life expectancy	Percent
X_3	Literacy rate	Percent
X_4	Average years of schooling	Percent
X_5	Adjusted per capita expenditure	Thousand rupiah
X_6	HDI	Percent
X_7	School enrolment rate	Percent
X_8	Per capita expenditure on food and housing	Percent
X_9	Poor rice recipient household	Percent
X_{10}	GRDP	Billion rupiah

4. Analysis Procedures

The research aims to investigate the GR and GGWR models in the context of analyzing poverty levels in Bengkulu Province. The methodology involves the following steps:

a. Conduct descriptive and exploratory analyses with both response and predictor variables. Descriptive data includes summary statistics and exploratory analysis using boxplots.
b. Perform data pre-processing to address any issues or anomalies in the dataset.
c. Test the gamma distribution assumption of the response variable (Y). Uses statistical test Anderson-Darling (AD).
d. Analyses the data using GR, involving: (1) Test for multicollinearity in predictor variables using VIF criteria. Address multicollinearity using Principal Component Analysis (PCA) if detected; (2) Determine the GR model; (3) Test the significance of parameters both simultaneously and partially in the GR model; and (4) Interpret the model.
e. Conduct a spatial heterogeneity test to assess spatial aspects in the data. Uses spatial statistical methods Breusch Pagan (BP) test.
f. GGWR model by:
   1) Calculating Euclidean distance (d_{ij}) between *i-th* and *j-th* locations.
      The Euclidean distance formula is as follows:

$$\boldsymbol{d_{ij}} = \sqrt{(\boldsymbol{u_i} - \boldsymbol{u_j})^2 + (\boldsymbol{v_i} - \boldsymbol{v_j})^2} \tag{5}$$

      Here, d_{ij} represents the Euclidean distance between observation locations i and j (Ma et al., 2021).
   2) Determining the optimum spatial bandwidth (b_S) based on minimum cross-validation (CV) value.

$$CV(\delta) = \sum_{i=1}^{k} [y_i - \hat{y}_{\neq i}(\delta)]^2 \tag{6}$$

      Furthermore, the parameter estimators η_1 and η_2 can be obtained through iterative methods based on the estimation results of δ that yield the minimum CV (Kardiana et al., 2022).
   3) Calculating model parameter estimates for each location.
   4) Performing partial model parameter testing.

g. Evaluate model performance between GR and GGWR using AIC, BIC, AICc, RMSE, and R^2 values. In the AIC, the term $-2\ \ln L(\hat{\theta})$ based on the empirical log-likelihood is referred to as the goodness-of-fit term. It assesses how well the fitted model $L(\hat{\theta})$ aligns with the data used in its construction, y. The correction term 2M serves as a penalty for model complexity. Models that are too simplistic to adequately capture the data's characteristics tend to have large goodness-of-fit terms but small penalty terms. Conversely, models that fit the data well but include unnecessary parameters have small goodness-of-fit terms but large penalty terms. Ideally, models striking a balance between fidelity to the data and simplicity exhibit small AIC values, reflecting a compromise between these two considerations (Cavanaugh & Neath, 2019). Formula AIC as follows:

$$AIC = -2\ln L(\hat{\theta}) + 2M \tag{7}$$

Schwarz (1978) introduced the Bayesian Information Criterion (BIC) within a Bayesian context. BIC applies a more stringent penalty term, M log k, compared to the penalty term of 2M in AIC, where k donates the number of observations. BIC is known for being asymptotically consistent in various model selection situations, unlike AIC, which has been demonstrated to be inconsistent (Heo et al., 2020). Formula BIC as follows:

$$BIC = -2\ln L(\hat{\theta}) + Mln(k) \tag{8}$$

AICc was developed by Hurvich and Tsai in 1989. Model fitness criteria for parameter estimation using AIC are considered biased for samples with limited sample size. However, AICc can correct for this. The criteria for selecting the best model using AICc are defined as follows (DelSole & Tippett, 2021):

$$AICc = AIC + \frac{2M(M+1)}{k-M-1} \tag{9}$$

RMSE is a tool for measuring the goodness of a model based on the residual values of the estimation results. Uncertainty in the residuals of observations or the method used to compare models and observations should be considered. RMSE assumes that the residual sample set is unbiased. The criteria for selecting the best model using RMSE are defined as follows (Chai & Draxler, 2014):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{k}\left(Y_i - \hat{Y}_i\right)^2}{k}} \tag{10}$$

The coefficient of determination, often symbolized as R^2, signifies the proportion of variability in the dependent or response variable explained by the independent or explanatory variables in regression analysis. It is serves as a widely used metric to gauge the strength of the relationship in regression models. This coefficient is defined as follows (Kasuya, 2018):

$$R^2 = \frac{\sum_{i=1}^{k}\left(\hat{Y}_i - \bar{Y}_i\right)^2}{\sum_{i=1}^{k}(Y_i - \bar{Y}_i)^2} \tag{11}$$

These steps collectively provide a comprehensive approach to understanding and modeling poverty levels in Bengkulu Province, utilizing both traditional GR and the spatially focused GGWR model.

C. RESULT AND DISCUSSION

1. Descriptive and Exploratory Analysis

Descriptive data analysis is conducted to observe the range or scope of each variable used in the data. Additionally, descriptive data can provide information regarding the data. A total of ten observation points were made over eight years, from 2015 to 2022, resulting in 80 observations obtained. Descriptive data analysis is conducted on the entire dataset used in the analysis. Descriptive statistics can be seen in Table 2.

Table 2. Descriptive Statistics

Variable	Minimum	Mean	Maximum	Standard Deviation
Y	8.2	15.55	22.98	4.037
X_1	106290	197160	385140	84791.836
X_2	62.31	67.2	70.46	1.777
X_3	87.51	99.24	100	1.846
X_4	6.88	8.386	11.82	1.199
X_5	7077	9862	14503	1586.148
X_6	63.41	68.78	80.99	3.994
X_7	69.14	94.38	100	6.802
X_8	44.91	57.89	65.33	4.426
X_9	3.51	40.66	81.72	17.259
X_{10}	2331	6834	28090	5532.203

Table 2 shows that for variable Y, the lowest percentage of the poor population was 8.2% in Central Bengkulu Regency in 2018, and the highest percentage of the poor population was 22.98% in Seluma Regency in 2015. The mean value of the poor population is 15.55%, and the standard deviation for the poor population indicates the extent of variation from the mean value, which is 4.037%. Data was explored using a boxplot on the dataset to check for outliers and correlations. The exploration conducted to identify outliers in Bengkulu Province is as shown in Figure 1.

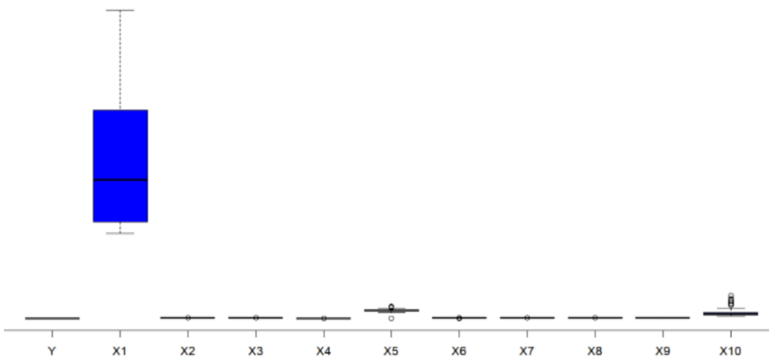

Figure 1. Boxplot of the Variable Used

Based on Figure 1, the data used has different units. Different units cause the values to have various ranges. This gives the boxplot a wide range. In the boxplot, it is found that several variables have outliers. The following are the Correlation Values among Variables, as shown in Figure 2.

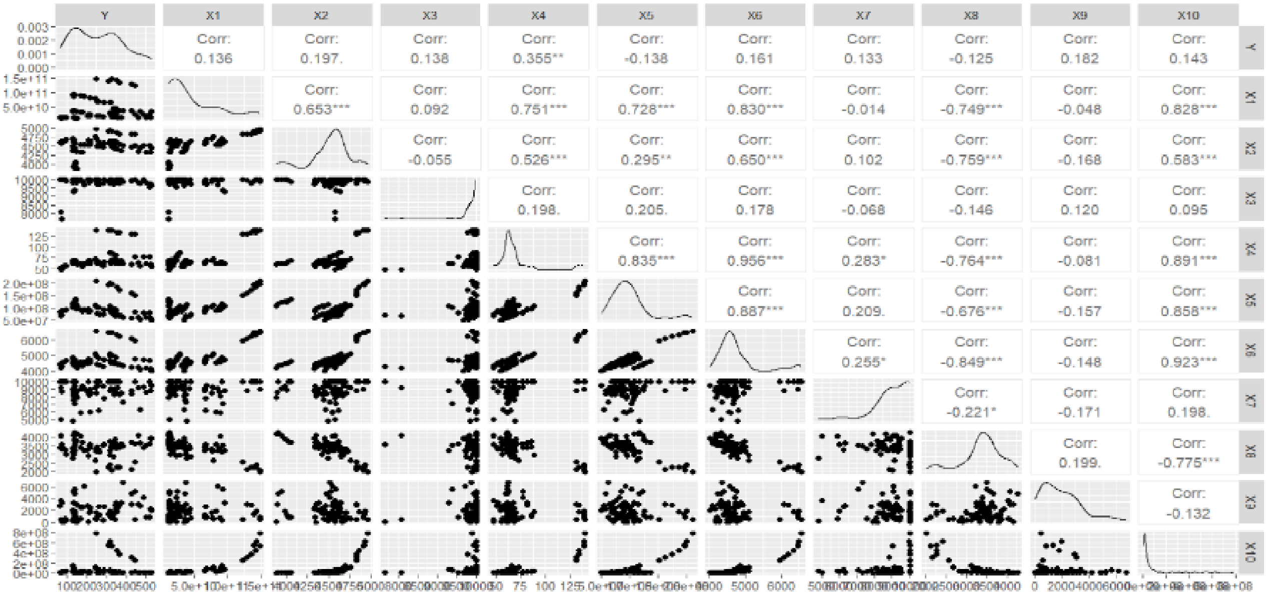

Figure 2. Correlation Values among Variables

Based on Figure 2, it can be seen that there is a correlation between predictor variables with significant correlation coefficients, indicated by asterisks. These significant correlation values signify a notable relationship between one predictor variable and another. In contrast, insignificant correlations suggest no relationship between the predictor variables (Jamil et al., 2018). These predictor variables will be used to model Gamma Regression in further analysis.

2. Gamma Distribution Test

Distribution testing will be conducted on the response variable (Y), the percentage of the poor population, to determine the distribution of the poor population variable. The testing will use the 'goft' package in the R program. This test determines whether variable Y follows a gamma distribution or not. Two widely recognized methods exist for devising tests to ascertain the fundamental distribution of data. The first traditional method relies on the "distance" between the empirical distribution function and the assumed distribution, as seen in tests like Kolmogorov–Smirnov (KS), Anderson–Darling (AD), and Cramér–von Mises (CM). The second method involves assessing specific properties of the distribution under scrutiny, such as the gamma distribution in this study (Plubin & Siripanich, 2017). Here, the research employs the Anderson-Darling (AD) test. The single-sample variant of the Anderson-Darling (AD) test for a ranked test sample $Y = (y_1, y_2, \dots, y_n)$ and empirical cumulative distribution function F is described in reference provided (Halme & Koivunen, 2018):

$$AD = -k - \frac{1}{k}\sum_{i=1}^{k}(2i-1)[logF(y_i) + \log\{1 - F(y_{k+1-i})\}] \tag{12}$$

Table 3. Distribution Testing for Gamma Distribution

Variable	Test Statistics	*p-value*
Y^2	-1.3529	0.3387

Based on Table 3, the results indicate that after data transformation, the variable Y^2 exhibits a distribution similar to a gamma distribution. This finding is crucial for the context of poverty modeling analysis in Bengkulu Province. The gamma distribution is often suitable for variables with characteristics such as positive skewness and unlimited values starting from zero, commonly found in economic and social data like poverty rates. Overall, the results from Table 3 indicate that the data used in this analysis meets the necessary assumptions for Gamma Regression.

3. Gamma Regression Modelling

Estimating parameters at each observation location yields consistent (global) properties. This study encompasses factors influencing each observation location. After obtaining the GR model, assumptions will be tested. Testing will include multicollinearity and heterogeneity tests. Multicollinearity testing aims to detect correlations among the predictor variables used. This is achieved by calculating the Variation Inflation Factor (VIF). The VIF formula is $VIF_k = \frac{1}{1-R_k^2}$, where k denotes the number of predictor variables. R_k^2 represents the coefficient of determination of predictor variable k chosen as the response variable and other predictor variables as predictors for the response variable (Marcoulides & Raykov, 2018). The following is the multicollinearity test for predictor variables, as shown in Table 4.

Table 4. The Multicollinearity Test for Predictor Variables

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
VIF	4.99	35.65	1.27	66.28	74.31	278.9	1.39	5.49	1.2	11.04

Based on Table 4, VIF values for each predictor variable are shown. VIF values greater than 10 indicate a common threshold for detecting multicollinearity. High VIF values increase sample variance and result in dependency among the variables used (Ahmad et al., 2021). In Table 4, variables with $VIF > 10$ indicate multicollinearity. Principal Component Analysis (PCA) will be employed to address multicollinearity. PCA allows the transformation of initially information from initially correlated data into new orthogonal components, condensing a large amount of information. This process facilitates the discovery of hidden relationships, enhances data visualization, aids in outlier detection, and improves classification within the newly defined dimensions. (Zhu et al., 2019). PCA will be used for Gamma Regression modeling.

During PCA testing, multicollinearity issues yielded VIF values of 1 for each predictor variable. This means all predictor variables can be used for the response variable (percentage of the poor population). When modeling with Gamma regression, the influential predictor variables obtained from PCA are $PC_1, PC_2, PC_4, PC_5, PC_7, PC_8, PC_9,$ and PC_{10}. These predictor variables will be modeled in Gamma regression, resulting in the following Gamma regression model:

$$Y = \exp(\ 25.314 + 2.825 \times 10^{-6}X_1 - 0.06X_2 + 0.016X_3 + 0.091X_4 - 0.00023X_5 + 0.051X_6 + 0.005X_7 - 0.014X_8 - 0.005X_9 + 3.71 \times 10^{-5}X_{10})$$

4. Spatial Heterogeneity Test

Heterogeneity testing utilizes the Breusch-Pagan (BP) statistical test. This test is used to determine the presence of spatial influence heterogeneity. Simultaneously, the test is conducted on 10 districts/cities in Bengkulu Province from 2015 to 2022. The formula for the BP test follows a chi-square distribution. The formula is as follows (Halunga et al., 2017):

$$\boldsymbol{BP} = \frac{\mathbf{1}}{\mathbf{2}}\left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i\, f_i}\right)^{\mathbf{T}}\left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i x_i^T}\right)\left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i\, f_i}\right) \sim \boldsymbol{\chi}^2_{(\mathbf{p-1})} \quad (13)$$

Table 5. Heterogeneity Test

BP Value	db	*p-value*
15.537	8	0.0495

Table 5 shows the p-value < 0.05, indicating significant at the 5% level. Therefore, there is spatial heterogeneity in poverty data for districts/cities in Bengkulu Province from 2015 to 2022. The presence of heterogeneity effects suggests significant variation in poverty levels across different regions within Bengkulu Province during this period. Thus, these findings support the continuation of analysis by incorporating spatial variables into the model, such as using Gamma Geographically Weighted Regression (GGWR). This approach can provide deeper insights in to the factors influencing poverty in Bengkulu and aid in designing targeted policies to address these issues effectively.

5. Gamma Geographically Weighted Regression Modelling

Gamma Geographically Weighted Regression (GGWR) has two assumptions to be met before modeling: spatial heterogeneity and the distribution of the response variable (Y) being gamma distributed. This study employs weighting with Gaussian and Bisquare kernel functions, both fixed and adaptive. The purpose is to determine which weighting method will be used for the case of the percentage of the poor population in Bengkulu Province. The method for selecting the best weighting is cross validation (CV). The best weighting functions are made across all years using the GGWR model. The comparison between weighting functions for the GGWR model is as shown in Table 6.

Table 6. Comparison of Weighting Functions

Weighting Function	Fixed Gaussian	Adaptive Gaussian	Fixed Bisquare	Adaptive Bisquare
AIC	-59.08	-48.70	-52.21	-93.84

Based on Table 6, the best weighting function is the one with the lowest AIC value. The best weighting function for the case of the percentage of the poor population is Adaptive Bisquare. Weighting in the GGWR model is conducted for each year, so each year's model has different weights. Modelling with the best weighting will be done for the 2015-2022. This is because the AIC value for those years is the smallest. GGWR modeling will be performed using the Adaptive Bisquare function. The equation for the GGWR model is as follows:

$$Y = \exp(-6.039 + 3.15 \times 10^{-6} X_1 - 0.055 X_2 + 0.156 X_4 - 0.00021 X_5 + 0.004 X_7 \\ -0.021 X_8 - 0.006 X_9 + 4.794 \times 10^{-5} X_{10})$$

Map of the distribution of variables based on Districts/Cities in Bengkulu Province. Mapping of the GGWR model using parameters for the year 2022. The spatial distribution of parameter estimators is performed on predictor variables $X_1, X_2, X_4, X_5, X_7, X_8, X_9$ and X_{10}. Parameters are categorized in red colour when the parameter values are negative and blue when the parameter values are positive, as shown in Figure 3.

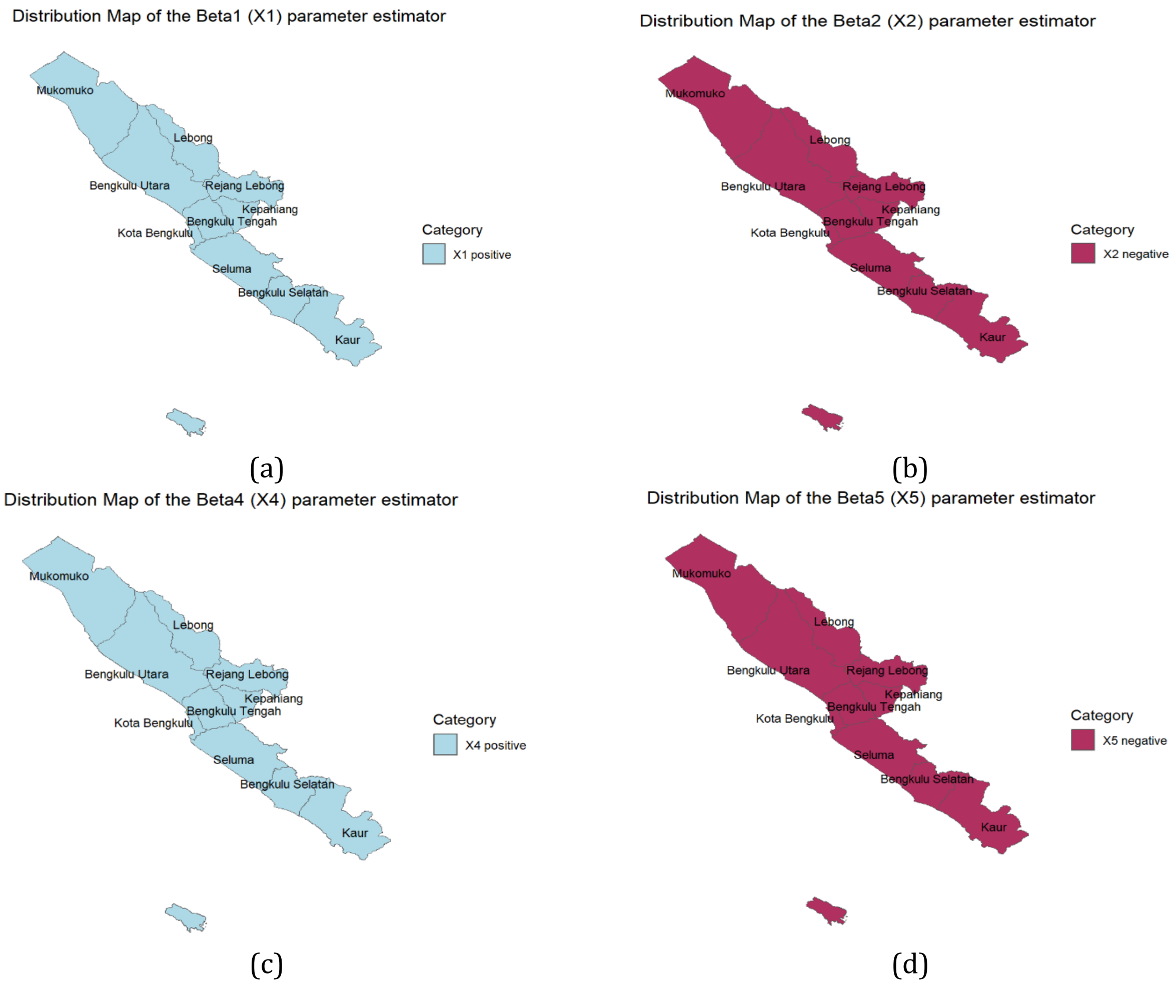

Figure 3. Mapping of the GGWR Model for Beta1 (a), Beta2 (b), Beta4 (c), Beta5 (d), Beta7 (e), Beta8 (f), Beta9 (g), and Beta10 (h) parameter estimator

6. Model Evaluation

Model evaluation is used to determine the best model among Gamma Regression and Gamma Geographically Weighted Regression models. These models will be evaluated using five criteria. Model evaluation results are as shown in Table 7.

Table 7. Comparison Model

Model	AIC	BIC	AICc	RMSE	R^2
GR	846.52	849.71	870.34	46.54	0.8793
GGWR	-93.84	-69.85	-17.09	0.11	0.9587

Table 7 suggest that the best model is determined by having the smallest values of AIC, BIC, AICc, and RMSE, alongside the largest value of R^2. According to these criteria, the GGWR model is identified as the optimal model. The coefficient of determination (R^2) for the GGWR model in modeling poverty in Bengkulu Province is 0.9587. This indicates that 95.87% of the variance is explained by the variables Total Population, Life Expectancy, Average Length of Schooling, adjusted per Capita Expenditure, School Participation Rate, per Capita Expenditure for Food, Household Receiving Raskin, and Regional Gross Domestic Product, while the remaining variance is explained by other variable. This suggests that the GGWR model is the best model.

D. CONCLUSION AND SUGGESTIONS

Gamma Geographically Weighted Regression (GGWR) Model for Poverty in Bengkulu Province. One of the models is the form of the GWGR model equation for Bengkulu City in 2022 as follows: $Y = \exp(-6.039 + 3.15 \times 10^{-6}X_1 - 0.055X_2 + 0.156X_4 - 0.00021X_5 + 0.004X_7 - 0.021X_8 - 0.006X_9 + 4.794 \times 10^{-5}X_{10})$. Based on the performance evaluation of the Gamma Regression and GGWR models, it is found that the best model is the GGWR model. For future research on addressing multicollinearity issues, it is recommended to consider using other methods such as LASSO or ridge regression. The benefit of this research is to serve as a reference for the provincial government of Bengkulu regarding the variables that influence poverty. It is expected that this will help the government reduce the poverty rate in Bengkulu Province.

REFERENCES

Agustina, M. F., Wasono, R., & Darsyah, Y. (2015). Pemodelan Geographically Weighted Regression (GWR) Pada Tingkat Kemiskinan di Provinsi Jawa Tengah Monica. *Statistika*, *3*(2), 67–74. https://doi.org/10.26714/jsunimus.3.2.2015.%25p

Ahmad, A. U., Balakrishnan, U. V, & Jha, P. S. (2021). A Study of Multicollinearity Detection and Rectification under Missing Values. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(1S), 399–418. https://doi.org/10.17762/turcomat.v12i1s.1880

Bossio, M. C., & Cuervo, E. C. (2015). Gamma regression models with the Gammareg R package. *Comunicaciones En Estadística*, *8*(2), 211–223. https://doi.org/10.15332/s2027-3355.2015.0002.05

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(3), 1–11. https://doi.org/10.1002/wics.1460

CSA. (2022). Katalog: 3205014. In *Data dan Informasi Kemiskinan Kabupaten/Kota di Indonesia Tahun 2022*. https://www.bps.go.id/id/publication/2022/11/30/3b084878f782dfa44e0025e0/data-dan-informasi-kemiskinan-kabupaten-kota-tahun-2022.html

DelSole, T., & Tippett, M. K. (2021). Correcting the corrected AIC. *Statistics and Probability Letters*, *173*(1), 1–7. https://doi.org/10.1016/j.spl.2021.109064

Dewi, D. K. (2019). *Penaksiran Parameter Dan Pengujian Hipotesis Mixed Geographically Weighted Bivariate Generalized Poisson Regression Testing of Geographically Weighted*. http://repository.its.ac.id/id/eprint/45755

Dunder, E., Gumustekin, S., & Cengiz, M. A. (2016). Variable selection in gamma regression models via artificial bee colony algorithm. *Journal of Applied Statistics*, *45*(1), 8–16. https://doi.org/10.1080/02664763.2016.1254730

Halme, T., & Koivunen, V. (2018). Distributed Nonparametric Inference Using a One-Sample Bootstraped Anderson-Darling Tesst and P-Value Fusion. *2018 IEEE Data Science Workshop, DSW 2018 - Proceedings*, 56–60. https://doi.org/10.1109/DSW.2018.8439920

Halunga, A. G., Orme, C. D., & Yamagata, T. (2017). A heteroskedasticity robust Breusch–Pagan test for Contemporaneous correlation in dynamic panel data models. *Journal of Econometrics*, *198*(2), 209–230. https://doi.org/10.1016/j.jeconom.2016.12.005

Heo, J., Lee, J. Y., & Kim, W. (2020). Bayesian information criterion accounting for the number of covariance parameters in mixed effects models. *Communications for Statistical Applications and Methods*, *27*(3), 300–311. https://doi.org/10.29220/CSAM.2020.27.3.301

Isazade, V., Qasimi, A. B., Dong, P., Kaplan, G., & Isazade, E. (2023). Integration of Moran's I, geographically weighted regression (GWR), and ordinary least square (OLS) models in spatiotemporal modeling of COVID-19 outbreak in Qom and Mazandaran Provinces, Iran. *Modeling Earth Systems and Environment*, *9*(4), 3923–3937. https://doi.org/10.1007/s40808-023-01729-y

Jamil, S. A. M., Abdullah, M. A. A., Long, K. S., Jupri, N. F. M., & Mamat, M. (2018). A stepwise logistic regression analysis: An application toward poultry farm data in Johor. *International Journal of Engineering and Technology(UAE)*, *7*(3.28), 68–71. https://doi.org/10.14419/ijet.v7i3.28.20968

Kardiana, A., Wigena, A. H., Djuraidah, A., & Soleh, A. M. (2022). Statistical Downscaling Modeling for Monthly Rainfall Estimation Using Geographical and Temporal Weighted Gamma Regression. *Asian Journal of Mathematics and Computer Research*, *29*(February), 43–55. https://doi.org/10.56557/ajomcor/2022/v29i27923

Kasuya, E. (2018). On the use of r and r squared in correlation and regression. *Ecological Research*, *34*(1), 235–236. https://doi.org/10.1111/1440-1703.1011

Ma, Z., Xue, Y., & Hu, G. (2021). Geographically Weighted Regression Analysis for Spatial Economics Data: A Bayesian Recourse. *International Regional Science Review*, *44*(5), 582–604. https://doi.org/10.1177/0160017620959823

Marcoulides, K. M., & Raykov, T. (2018). Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *Educational and Psychological Measurement*, *79*(5), 874–882. https://doi.org/10.1177/0013164418817803

Plubin, B., & Siripanich, P. (2017). An alternative goodness-of-fit test for a gamma distribution based on the independence property. *Chiang Mai Journal of Science*, *44*(3), 1180–1190. https://www.thaiscience.info/Journals/Article/CMJS/10985864.pdf

Putri, A. N., & Wardhani, A. K. (2020). Penerapan Metode Single Moving Average Untuk Peramalan Harga Cabai Rawit Hijau. *Indonesian Journal of Technology, Informatics and Science (IJTIS)*, *2*(1), 37–40. https://doi.org/10.24176/ijtis.v2i1.5653

Putri, D. E., Purhadi, & Prastyo, D. D. (2017). Parameter estimation and hypothesis testing on geographically weighted gamma regression. *IOP Conference Series: Earth and Environmental Science*, *893*(1), 1–7. https://doi.org/10.1088/1755-1315/880/1/012044

Rash, A. J. H., Khodakarami, L., Muhedin, D. A., Hamakareem, M. I., & Ali, H. F. H. (2024). Spatial modeling of geotechnical soil parameters: Integrating ground-based data, RS technique, spatial statistics and GWR model. *Journal of Engineering Research (Kuwait)*, *12*(1), 75–85. https://doi.org/10.1016/j.jer.2023.10.026

Tang, Y., Xie, S., Huang, L., Liu, L., Wei, P., Zhang, Y., & Meng, C. (2022). Spatial Estimation of Regional PM2.5 Concentrations with GWR Models Using PCA and RBF Interpolation Optimization. *Remote Sensing*, *14*(21), 1–26. https://doi.org/10.3390/rs14215626

Yamagata, Y., & Seya, H. (2019). Spatial analysis using big data: Methods and urban applications. In *Spatial Analysis Using Big Data: Methods and Urban Applications*. Japan:Elsevier. https://doi.org/10.1016/C2016-0-02101-0

Yasin, H., Sugito, S., & Prahutama, A. (2015). Analisis Data Kemiskinan Di Jawa Tengah Menggunakan Metode Mixed Geographically and Temporally Weighted Regressions (MGTWR). *BIAStatistics*, *9*(1), 15–23. http://eprints.undip.ac.id/47841/

Zhao, Q., Fan, Q., & Zhou, P. (2022). An integrated analysis of GWR models and spatial econometric global models to decompose the driving forces of the township consumption development in Gansu, China. *Sustainability (Switzerland)*, *14*(1), 1–16. https://doi.org/10.3390/su14010281

Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked*, *17*(3), 1–7. https://doi.org/10.1016/j.imu.2019.100179