CAUCHY: Jurnal Matematika Murni dan Aplikasi
Volume 10(1) (2025), Pages 290-300
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Modelling Geographically Weighted Truncated Spline Regression Using Maximum Likelihood Estimation for Human Development Disparities

Adji Achmad Rinaldo Fernandes, La Ode Muhammad Saris*, Henny Pramoedyo, Adji Achmad Rinaldo Fernandes

Departement of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Indonesia

Email: fernandes@ub.ac.id; sarismuhammad838@gmail.com

ABSTRACT

A development of nonparametric truncated spline regression, Geographically Weighted Truncated Spline Regression (GWTSR) incorporates spatial effects in the modelling of nonlinear relationships between the response and predictor variables. This research utilizes the Maximum Likelihood Estimation (MLE) technique to estimate the parameters of the model. The first-order truncated spline with a single knot yielded a minimal Generalized Cross-Validation (GCV) value of 1.729781, suggesting a high level of accuracy in the model. Four weighting functions were evaluated: Gaussian Kernel, Exponential Kernel, Bi-Square Kernel, and Tri-Cube Kernel. Among these, the Bi-Square weighting function performed the best, achieving a coefficient of determination of 98,70474%, which demonstrates the model's ability to explain nearly all data variability effectively. The coefficient of determination, which is close to 100%, reflects the model's ability to explain data variations effectively. The model's preformance is influenced by the selection of optimal knot poins to capture nonlinear patterns, optimal bandwidth to ensure accurate spatial analysis, and appropriate weighting functions that enhance prediction accuracy and reliability. GWSTR proves to be a robust method for capturing complex nonlinear relationships while accounting for spatial variations, making it a valuable tool for spatial data analysis across various disciplines.

Keywords: Truncated Spline, Geographically Weighted Regression (GWR), Geographically Weighted Truncated Spline Regression (GWTSR), Weighting Spatial

INTRODUCTION

Human Development aims to improve the quality of life in various aspects, such as education, health, access to resources, and security, to achieve fair, inclusive, and sustainable prosperity [1]. In Indonesia. The Human Development Index(HDI) increased from 2020 to 2023. In 2023, Indonesia's HDI reached 74,39, an increase of 0,62 points or 0,84 percent compared to the previous year. This increase was supported by improved longevity, health, knowledge, and decent living standards [2]. In Southeast Sulawesi Province, the HDI has also increased, from 71,66 in 2021 to 72,74 in 2023. This increase reflects the economic recovery after the COVID-19 pandemic and the overall

Submitted: January 02, 2025 Reviewed: February 11, 2025 Accepted: March 11, 2025
DOI: http://dx.doi.org/10.18860/ca.v10i1.31381

improvement in human development. The success of equitable development can be seen from the change in the HDI status of several districts/cities from the "medium" to the "high" category [3]. However, in the context of research, the spatial influence on HDI in southeast Sulawesi has not been specifically discussed.

The relationship between HDI and Influencing factors can be known using regression analysis is used in statistics to measure the relationship between response and predictor[4] There are three methods used to describe the relationship between response variables and predictors, namely using parametric, nonparametric and semiparametric regression. Parametric regression is used when the shape of the curve has a pattern. At the same time, nonparametric regression can be done if both models of the relationship between two variables have no pattern. If the curve's shape is somewhat known and partially uncertain, semiparametric analysis is employed [5] The frequent nonparametric regression models are spline, Fourier series, wavelet, and kernel [6]. The reasearchers mostly use truncated spline because this method has high flexibility and can adjust according to the movement of data at certain sub-intervals[7].

Previous research on HDI has been conducted, including modelling the Human Development Indeks in Southeast Sulawesi in 2017-2020 using the panel data regression method. The factors that influence HDI are GRDP at constant prices, school enrollment rates, and the number of poor people [8]. Furthermore, research conducted by [9] with the theme The effect of the Human Development Index and Economic Growth on Unemployment in Southeast Sulawesi. This study uses multiple liner regression methods. The results showed the the Human Development Index was influenced by the unemployment Rate. Previous studies focused on linear regression research but for nonlinear data was not considered. Then, Southeast Sulawesi Province has a different geographical location in some districts/cities. Geographically Southeast Sulawesi province is a province that has several islands that cause disparities in the influence of development factors on HDI.

This study aims to develop a GWTSR model to analyze nonlinear patterns that influence spatial factors on HDI in Southeast Sulawesi. In addition, this study also evaluates the best model based on the optimal knot variation, kernel weight, and understanding the spatial patterns that effect HDI at the district/city level. Thus, the results of the study are expected to contrubute to understanding the dynamics of spatial-based human development and become the basis for more targeted and equitable developments policies.

METHODS

This study used secondary data from data from the southeast Sulawesi Central Bureau of Statistics (www.sultra.bps.go.id). The research data consisted of five predictor variables and one response variable. The response variable is the Human Development Index (HDI), while the predictor variable consist of: Average Years of schooling (x_1), Expected Years of Schooling (x_2), Life Expectancy (x_3), and per capita expenditure (x_4). The analysis in this study covers all districts/cities in Southeast Sulawesi with a total of 19 districts/cities included in the 2023 data. This research uses the nonparametric Geographically Weighted Truncated Spline Regression method. The research steps are as follows:

1. Linear assumption test using the Ramsey Test or (RESET) Test
2. Spatial Heterogeneity Test using Breusch Pagan Test (BP-TEST)
3. Estimating the Geographically Weighted Truncated Spline Regression curve function using MLE

4. Determining optimal knot points using the minimum value of Generalized used Cross Validation (GCV)
5. Determining the optimal bandwidth using the minimum value of Cross-Validation
6. Evaluate the accuracy of the curve estimation model using the coefficient of determination R^2_{adj}.

Determination of the Optimal Knot Point

Determination of optimal knot points using Generalized Cross-Validation (GCV) is done by selecting the number and location of knots the minimize the GCV function. A knot point is a change in the behavior pattern of the function or curve. However, the number of knot points also affects the complexity of the model due to the number of parameters used, so an appropriate method is needed to determine the optimal knot point. This method is determining by minimizing the GCV. The GCV method is generally defined as follows[10], [11]:

$$GCV\ (K) = \frac{MSE(K)}{(n^{-1} tr[\boldsymbol{I}-\boldsymbol{A}(K)])^2} \quad (1)$$

Where:

$\boldsymbol{I}$: Identity Matrix

n : The total number of observations

K : Designated knot point

$MSE(K)$: The Mean Square error associated with the truncated spline model

Spatial Heterogeneity Analysis

The spatial heterogeneity test is used to see the diversity between locations caused by the different structures in each area[12]. One of the methods used in this test is the Breusch-Pagan Test (BP Test). The hypothesis is as follow[13]s:

H_0 : $\sigma^2_{(u_1,v_1)} = \cdots = \sigma^2_{(u_n,v_n)} = \sigma^2$ (no spatial heterogeneity)

H_1 : there is at least one $\sigma^2_{(u_i,v_i)} \neq \sigma^2$ (spatial heterogeneity exists)

Test Statistic :

$$BP = \left(\frac{1}{2}\right) \boldsymbol{h^T Z}(\boldsymbol{Z^T Z})^{-1} \boldsymbol{Z^T h} \quad (2)$$

Where h is a vector with element $h_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$, e represents the error variance, σ^2 is the error variance, Z is the normalized matrix. Decision-making is done by comparing the $p - value$ and α. If $p - value < \alpha$. Then H_0 is rejected. In conclusion, there is spatial heterogeneity.

Spatial Weighting Method

The spatial weight matrix measures the spatial relationship between observation locations. The weightvalue depends on the distance between observation locations, where the elements on the diagonal of the matrix are weight functions. The spatial weighting function plays a role in determining the weight of observations based on the distance between locations[14], [15]. Some of weighting functions that can be used in this study include:

1. Gaussian Kernel : $W(d_{ij}) = exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right\}$
2. Exponential Kernel : $W(d_{ij}) = exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)\right\}$
3. Bi-Square : $W(d_{ij}) = \begin{cases} \left(1-\left(\frac{d_{ij}}{h}\right)^2\right)^2 d_{ij} < h \\ 0, d_{ij} > 0 \end{cases}$
4. Tri-cube kernel : $W(d_{ij}) = \left(1-\left(\frac{|d_{ij}|}{h}\right)^3\right)^3$

Where h is the bandwidth, d_{ij} is the distance between the j-th data point and the observed poin (u_i, v_i). The distance d_{ij} between locations i and j is calculated using the Euclidean distance formula:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (3)$$

The bandwidth h functions to determine how much influence other locations have on the observation location. The selection of the optimal bandwidth is crucial in GWR as it can affect the model's accuracy. One method to determine the optimal bandwidth is to minimize the Cross-Validation (CV) value. The CV can be calculated as follows[16]:

$$CV = \sum_{i=1}^{n}\left(y_i - \hat{y}_{\neq i}(h)\right)^2 \quad (4)$$

Where $\hat{y}_{\neq i}(h)$ is the estimated value of y_i when the observation at location i is excluded from the estimation process.

RESULTS AND DISCUSSIONS

Results

Estimation of Geographically Weighted Regression Spline Truncated Regression

A statistical technique called Geographically Weighted Truncated Spline Regression (GWTSR) takes into account spatial changes in the data while modelling the relationship between response variables and a nonparametric regression function, the exact form of which is unknown. The GWTSR model is an extension of the truncated spline nonparametric regression that accommodates spatial effect. The mathematical model of GWSTR can be as follows[17]:

$$f(x_i) = \beta_0(u_i, v_i) + \sum_{l=1}^{p}\sum_{j=1}^{m}\beta_{lj}(u_i, v_i)x_{li}^{j} + \sum_{l=1}^{p}\sum_{k=1}^{r}\delta_{p,m+k}(u_i, v_i)(x_{li} - K_{lk})_+^m \quad (5)$$

Where the truncated function is expressed as:

$$(x_{li} - x_{lk})_+^m = \begin{cases} (x_{li} - K_{lk}), & x_{li} > K_{lk} \\ 0 & x_{li} < K_{lk} \end{cases} \quad (6)$$

Equation (5) can be rewritten in the matrix as:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1}^m & (x_{11}-K_{11})_+^m & \cdots & (x_{11}-K_{1r})_+^m & \cdots & (x_{p1}-K_{p1})_+^m & \cdots & (x_{p1}-K_{pr})_+^m \\ 1 & x_{12} & \cdots & x_{p2}^m & (x_{12}-K_{11})_+^m & \cdots & (x_{12}-K_{1r})_+^m & \cdots & (x_{p2}-K_{p1})_+^m & \cdots & (x_{p2}-K_{pr})_+^m \\ 1 & x_{13} & \cdots & x_{p3}^m & (x_{13}-K_{11})_+^m & \cdots & (x_{13}-K_{1r})_+^m & \cdots & (x_{p1}-K_{p1})_+^m & \cdots & (x_{p3}-K_{pr})_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & (x_{p1}-K_{p1})_+^m & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn}^m & (x_{1n}-K_{11})_+^m & \cdots & (x_{1n}-K_{1r})_+^m & \cdots & (x_{p1}-K_{p1})_+^m & \cdots & (x_{pn}-K_{pr})_+^m \end{bmatrix} \begin{bmatrix} \beta_0(u_i,v_i) \\ \beta_{11}(u_i,v_i) \\ \vdots \\ \beta_{1m}(u_i,v_i) \\ \vdots \\ \beta_{p1}(u_i,v_i) \\ \vdots \\ \beta_{pm}(u_i,v_i) \\ \delta_{1,m+1}(u_i,v_i) \\ \vdots \\ \delta_{1,m+r}(u_i,v_i) \\ \vdots \\ \delta_{p,m+1}(u_i,v_i) \\ \vdots \\ \delta_{p,m+r}(u_i,v_i) \end{bmatrix} \quad (7)$$

Where:

$$\underset{\sim}{f}(x) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_n) \end{bmatrix};$$

$$\boldsymbol{A}(\underset{\sim}{k}) = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1}^m & (x_{11}-K_{11})_+^m & \cdots & (x_{11}-K_{1r})_+^m & \cdots & (x_{p1}-K_{p1})_+^m & \cdots & (x_{p1}-K_{pr})_+^m \\ 1 & x_{12} & \cdots & x_{p2}^m & (x_{12}-K_{11})_+^m & \cdots & (x_{12}-K_{1r})_+^m & \cdots & (x_{p2}-K_{p1})_+^m & \cdots & (x_{p2}-K_{pr})_+^m \\ 1 & x_{13} & \cdots & x_{p3}^m & (x_{13}-K_{11})_+^m & \cdots & (x_{13}-K_{1r})_+^m & \cdots & (x_{p1}-K_{p1})_+^m & \cdots & (x_{p3}-K_{pr})_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & (x_{p1}-K_{p1})_+^m & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn}^m & (x_{1n}-K_{11})_+^m & \cdots & (x_{1n}-K_{1r})_+^m & \cdots & (x_{p1}-K_{p1})_+^m & \cdots & (x_{pn}-K_{pr})_+^m \end{bmatrix};$$

$$\underset{\sim}{\theta}(u_i,v_i) = \begin{bmatrix} \beta_0(u_i,v_i) \\ \beta_{11}(u_i,v_i) \\ \vdots \\ \beta_{1m}(u_i,v_i) \\ \vdots \\ \beta_{p1}(u_i,v_i) \\ \vdots \\ \beta_{pm}(u_i,v_i) \\ \delta_{1,m+1}(u_i,v_i) \\ \vdots \\ \delta_{1,m+r}(u_i,v_i) \\ \vdots \\ \delta_{p,m+1}(u_i,v_i) \\ \vdots \\ \delta_{p,m+r}(u_i,v_i) \end{bmatrix};$$

The nonparametric GWSTR regression function in matrix notation is expressed as:

$$\underset{\sim}{f}(x) = \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i) \quad (8)$$

The coefficient $\beta(u_i,v_i)$ and $\delta(u_i,v_i)$. Estimated using the Maximum Likelihood Estimation (MLE)method.The response variables y_i have a mean of $\boldsymbol{A}(\underset{\sim}{k})\theta(u_i,v_i)$ and are distributed normally. Mathematically, the GWSTR mode is written as:

$$f(y_1,y_2,\dots,y_n) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{1}{2\sigma^2}(y_i - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta})^2\right) \quad (9)$$

The likelihood function is given by:

$$L\left(\underset{\sim}{y}\,;\underset{\sim}{\theta}(u_i,v_i),\sigma^2\right) = \prod_{i=1}^{n} f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))^2\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}\left\|\underset{\sim}{y} - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i)\right\|^2\right) \quad (10)$$

By incorporating the weight matrix $\boldsymbol{W}(u_i,v_i)$,He log-likelihood function is expressed as:

$$\ln L(\underset{\sim}{y};\ \underset{\sim}{\theta}(u_i,v_i),\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\|\boldsymbol{W}(u_i,v_i)(\underset{\sim}{y} - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i)\right\|^2 \quad (11)$$

To obtain the estimated value of $\underset{\sim}{\theta}(u_i, v_i)$, The likelihood function can be maximized by partially differentiating it concerning the parameters. The parameter estimation for $\underset{\sim}{\theta}(u_i, v_i)$Is carried out as follows:

$$\frac{\partial(\ln L(\underset{\sim}{y};\underset{\sim}{\theta}(u_i,v_i),\sigma^2))}{\partial\underset{\sim}{\theta}(u_i,v_i)} = \frac{-\frac{1}{2\sigma^2}\left\|\boldsymbol{W}(u_i,v_i)(\underset{\sim}{y} - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))\right\|^2}{\partial\underset{\sim}{\theta}(u_i,v_i)} = 0 \quad (12)$$

$$\frac{\boldsymbol{W}(u_i,v_i)(\underset{\sim}{y} - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))^T(\underset{\sim}{y} - \boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))}{\partial\underset{\sim}{\theta}(u_i,v_i)} = 0$$

$$\frac{\boldsymbol{W}(u_i,v_i)\left(\underset{\sim}{y}^T\underset{\sim}{y} - \underset{\sim}{y}^T(\boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i)) - (\boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))^T\underset{\sim}{y} - (\boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))^T(\boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i))\right)}{\partial\underset{\sim}{\theta}(u_i,v_i)} = 0$$

$$\boldsymbol{W}(u_i,v_i)(2\boldsymbol{A}(\underset{\sim}{k})^T\boldsymbol{A}(\underset{\sim}{k})\underset{\sim}{\theta}(u_i,v_i) - 2\boldsymbol{A}(\underset{\sim}{k})^T\underset{\sim}{y}) = 0$$

$$(\boldsymbol{A}(\underset{\sim}{k})^T\boldsymbol{W}(u_i,v_i)\boldsymbol{A}(\underset{\sim}{k}))\underset{\sim}{\theta}(u_i,v_i) = \boldsymbol{A}(\underset{\sim}{k})^T\boldsymbol{W}(u_i,v_i)\underset{\sim}{y}$$

$$\hat{\underset{\sim}{\theta}}(u_i,v_i) = (\boldsymbol{A}(\underset{\sim}{k})^T\boldsymbol{W}(u_i,v_i)\boldsymbol{A}(\underset{\sim}{k}))^{-1}\boldsymbol{A}(\underset{\sim}{k})^T\boldsymbol{W}(u_i,v_i)\underset{\sim}{y} \quad (13)$$

Based on Equation (8), the estimated GWSTR regression function is written as :

$$\hat{\underset{\sim}{\boldsymbol{f}}}(\boldsymbol{x}) = \boldsymbol{A}(\underset{\sim}{\boldsymbol{k}})\left(\boldsymbol{A}(\underset{\sim}{\boldsymbol{k}})^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{u_i},\boldsymbol{v_i})\boldsymbol{A}(\underset{\sim}{\boldsymbol{k}})\right)^{-\boldsymbol{1}}\boldsymbol{A}(\underset{\sim}{\boldsymbol{k}})^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{u_i},\boldsymbol{v_i})\underset{\sim}{\boldsymbol{y}} \quad (10)$$

Where $W_i = diag(W_1(u_i,v_i), W_2(u_i,v_i), \dots, W_n(u_i,v_i))$ is the spatial weight matrix that varies for each predictor parameter at the i-th location.

Linear Assumption Test

The purpose of the linear assumption test is to assess how the predictor variables (X) and response variables (Y). Relate to one and This assessment determines whether the relationship follows a linear, quadratic, or cubic pattern. The method used for this evaluation is the Regression Specification Error Test (RESET).

Table 1 Ramsey RESET Linearity Tes

Relationship	**P-Value**	**Conclusion**
X vs Y	0.0001803	Non-Linear

The linearity test results utilizing the Ramsey RESET method are shown in **Table 1** Based on the table, the relationship between **X** and **Y** has a p-value of 0.0001803, indicating that the relationship is not linear, as the p-value is less than 0.05. The linearity assumption test was conducted with multiple observations, involving predictor variables such as X_1, X_2, X_3, and X_4 against the response variable Y. The test results show that the relationship between the response variable and the predictor variables is non-linear. Therefore, a nonparametric regression approach is needed to model this relationship.

Spatial Heterogeneity Assumption Test

The spatial heterogeneity determines whether the regression model residuals vary spatially, showing that the residual variances vary throughout the observed regions. This test utilises the Breush-Pagan Test (BP Test) with the following hypotheses:

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_i^2$ (no spatial heterogeneity).

H_1 : At least one $\sigma_i^2 \neq \sigma_j^2$ (spatial heterogeneity exists).

The test results show a p-value of 0.01586 < $\alpha(0.05)$. Thus, it can be concluded that spatial heterogeneity exists in the data.

Spatial Weighting

The estimation of parameters in the geographically Weighted Spline Truncated Regression (GWSTR) model requires spatial weighting based on observation point location. These observation points are determined by considering geographic coordinates, namely longitude and latitude. To select the optimal weighting function, the distances between locations are calculated using Euclidean distance. Subsequently, bandwidth optimization is conducted by minimizing the cross-validation (CV) value.

Table 2. Determination of Optimal Bandwidth

Weighting Function	Bandwidth	Cross Validation (CV)
Fixed Gaussian	0.5311736	2.693233
Adaptive Gaussian	0.2940822	2.76508
Fixed Bisquare	1.729781	2.581049
Adaptive Bisquare	0.75427	3.073763
Fixed Tricube	1.767424	2.670245
Adaptive Tricube	0.8826577	2.912786

Based on **Table 2.** the optimal bandwidth is obtained using the Fixed Bisquare weighting function with a value of 1.729781, yielding a minimum cross-validation (CV) value of 2.581049. The selection of bandwidth plays a crucial role in determining the spatial weighting function used in the model. Whether using Maximum Likelihood Estimation (MLE) or the Weighted Least Squares (WLS) approach, this spatial weighting determination is a crucial part of parameter estimation for the geographically weighted regression (GWR) model.

Determining the Optimal Knot Point

The ideal knot point identifies the best nonparametric spline regression model. A knot signifies a point where the function's behavioural pattern shifts. Generali zed cross-validation (GCV) is a technique to pick the optimal knot point. In this investigation, the optimal knot was considered to be one-knot,two-knot, and three-knot setups, with the polynomial order limited to linear. The GCV values for the truncated spline nonparametric regression model are represented below.

Table 3 Selection of the Optimal Knot Point

Polynomial Order	Knot Point	R^2	GCV
	One Knot	0.9765355	1.15888
Order One	Two Knots	0.9726351	4.324017
	Three Knots	0.9765355	Inf

Based on **Table 3,** the lowest recorded value of Generalized Cross Validation (GCV) is 1.15888, indicating that the selection of the optimal knot points has been successfully achieved. The optimal knot points for each variable in the nonparametric spline regression model with a first-order polynomial and a single knot are located at the following positions: $K_{11} = 7.34, K_{21} = 11.9, K_{31} = 70.14, K_{41} = 7564$. The selection of these knot points is a crucial step in the model estimation process, where a lower GCV value reflects a reduction in prediction errors and an improvement in model fit to the data. Thus, the ideal knot points play an important role in improving the accuracy and validity of the nonparametric spline regression result used in this work.

Determining Optimal Knot Point

The selection of the best model for spatially-based nonparametric spline regression in this study is determined by the coefficient of determination (R^2). A model with a higher R^2 value than others is considered superior. Four types of kernel weights are employed to estimate the curve function in the Geographically Weighted Truncated Spline Regression (GWTSR) model: Gaussian Kernel, Exponential Kernel, Bi-Square Kernel, and Tri-Cube Kernel. The R^2 values for each kernel are presented in Table 4.

Table 4. Selection of the Best Model

Weight Function	R^2
GWSTR Gaussian	98,38712%
GWSTR Gaussian	98,32781%
GWSTR Bi-Square	98,70474%
GWSTR Tri-Cube	98,52369%

As shown in Table **Table 4**, the GWSTR model with the Bi-Square kernel yields the highest R^2, outperforming the Gaussian, Exponential, and Tri-Cube Kernel models. This indicates that the Bi-Square kernel best capture the spatial variation inthe data, resulting in the most accurate predictions of the Human Development Indeks (HDI) at the district/city level in SoutheastSulawesi for 2023. Thus, The Bi-Square kernel is eemed the most optimal for modelling HDI in the region.

Determining the Optimal Knot Point

After selecting the optimal truncated spline model, which is a linear spline with knot poins, the next step is determine the latitude and longitude coordinates for each location, calculate the Euclidean distence, and determine the optimal bandwidth value based on the Cross Validation (CV) criteria. The optimal bandwidth value is 1,929781. The next step is to form a weight matrix using the Fixed Bi-Square keren Function and estimate the coefficient of the nonparametric regression curve function using one knot. The weight matrix used for model estimation is very important, as it ensures that the coefficients for each observation location are different. The coefficient estimation results obteined through R software for the spline truncated GWR model are represented in Table 5.

Table 5. Coefficients of Nonparametric Regression Curve Function for GWTSR

Region	β_0	β_1	β_2	β_3	β_4	δ_1	δ_2	δ_3	δ_4
Buton	1.902e-06	1.002e-05	-3.89e-06	2.23e-05	-3.16e-07	1.25e-04	-8.04e-06	0.0088	-0.0055
Muna	1.59e-06	7.932e-06	-3.735e-06	1.050e-05	-8.489e-06	1.109e-04	-1.041e-06	0.00857	-0.0035
Konawe	2.360e-06	2.421e-05	6.966e-06	3.085e-05	2.769e-06	1.657e-04	2.339e-07	0.0093	-0.0084
Kolaka	2.360e-06	2.421e-05	6.966e-06	3.085e-06	2.769e-06	1.657e-04	2.339e-07	0.0093	-0.0084
Konawe Selatan	1.890e-06	8.665e-06	-5.152e-06	1.653e-05	-5.960e-06	1.288e-04	-3.696e-06	0.00853	-0.0057
Bombana	1.923e-06	1.038e-05	3.676e-06	1.816e-05	-4.725e-06	1.303e-04	-4.520e-06	0.00849	-0.0060
Wakatobi	2.042e-06	1.257e-05	2.356e-06	2.348e-05	-8.267e-07	1.391e-04	-4150e-06	0,00873	-0.0067
Kolaka Utara	2.124e-06	1.433e-05	-1.195e-06	2.632e-05	1.040e-06	1.475e-04	-1.484e-06	0.00874	-0.007
Buton Utara	2.156e-06	1.720e-05	1.434e-06	2.600e-05	3.401e-07	1.497e-04	-1.488e-06	0.00906	-0.0072

Konawe Utara	2.049e-06	1.351e-05	-1.469e-06	2.321e-05	1.170e-06	1.389e-04	-4.740e-06	0.0088	-0.0066
Kolaka Timur	2.103e-06	1.337e-05	-2.001e-06	2.312e-05	-1.900e-06	1.527e-04	5.230e-06	0.0089	-0.0069
Konawe Kepulauan	2.137e-06	1.947e-05	3.844e-06	2.419e-05	1.243e-06	1.456e-04	-4.297e-06	0.0090	-0.0071
Muna Barat	1.596e-06	7.932e-06	-3.735e-06	1.050e-05	-8.489e-06	1.109e-04	1.041e-06	0.0085	-0.0035
Buton Tengah	1.991e-06	9.695e-06	-4.859e-06	2.456e-05	8.735e-07	1.352e-04	-4.357e-06	0.00871	-0.0063
Buton Selatan	1.866e-06	9.057e-06	-4.585e-06	2.131e-05	-0.904e-07	1.229e-04	-7.911e-06	0.00878	-0.0053
Kota Kendari	1.890e-06	8.665e-06	-5.152e-06	1.653e-05	-5.960e-06	1.288e-04	-3.696e-06	0.00853	-0.0057
Kota Baubau	1.853e-06	8.740e-06	-4.812e-06	2.098e-05	-1.078e-06	1.221e-04	-7.864e-06	0.00875	-0.0052

Based on **Table 5,** the coefficient values for the parameters in the GWR truncated spline model show both positive and negative signs, indicating the presence of both positive and negative effects at different observation locations. The parameter estimates at each location, derived using the Fixed Bi-Square Kernel weight function, yield an *lue* of 0,9870474 which means the model explains 99,999% of the variance in the data. The GWR Truncated Spline model with a linear order and one-knot, using the Bi-Square kernel weight function, can be expressed in the following equation:

$$\hat{y}_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \beta_2(u_i, v_i)x_{2i} + \beta_3(u_i, v_i)x_{3i} + \beta_4(u_i, v_i)x_{4i} + \delta_{1i}(u_i, v_i)(x_i - 7.34)_+ + \delta_{2i}(x_i - 11.9)_+ + \delta_{3i}(x_i - 70.14)_+ + \delta_{4i}(x_i - 7564)_+ \quad (11)$$

This equation illustrates the model structure with the estimated parameters at each observation location, emphasizing the spatial variability captured by the model. The positive and negative values of the parameters indicate varying influences at different geographical locations.

Discussion

The results showed the Human Development Index (HDI) is influenced by Life Expectancy, Average Years of Schooling, Expected Years of Schooling, and Per Capita Income. Analysis using Geographically Weighted Truncated Spline Regression (GWTSR) confirmed the nonlinear relationship between HDI and its predictor variables based on the Ramsey test, so the spline regression method was used to capture changes in the structure of the data relationship. Compared to linear approaches such as the classical Geographically Weighted Regression (GWR), GWTSR is more effective in accommodating nonlinear patterns and spatial heterogeneity, in line with previous studies showing the superiority of nonparametric modelling in the analysis of nonstationary spatial data.

The spatial heterogeneity test shows variations in the relationship between HDI and predictor factors in each region, confirming the need for HDI improvement policies tailored to local conditions. Factors such as regional policies, access to educations, and economic structure influence this heterogeneity. This study has limitations in the use of the truncated spline method which only reaches order one with three knot points. Futher development can use methods such as Fourier series, wavelet regression, or Bayesian Spatial Spline to increase the flexibility and accuracy of the model in capturing nonlinear patterns and more complex spatial variations.

CONCLUSIONS

The findings of this study provide valuable insights for policymakers and practitioners in designing more effective development strategies to improve the Human Development Index (HDI) in Southeast Sulawesi. By considering local conditions and socioeconomic dynamics, the results support the formulation of targeted, data-driven policies to enhance the well-being of the population. The study also demonstrates that a single knot, identified based on the minimal GCV value, and the Bi-Square weighting function, which achieved the highest R^2 value, are optimal for modelling HDI using Geographically Weighted Truncated Spline Regression.

REFERENCES

[1] United Nations Development Programmer (UNDP), "Human Development Report 2023-24," UNDP. Accessed: Jan. 01, 2025. [Online]. Available: https://hdr.undp.org/content/human-development-report-2023-24

[2] Badan Pusat Statistik, "Indeks Pembangunan Manusia 2023," 2023. Accessed: Dec. 20, 2024. [Online]. Available: https://www.bps.go.id/id/publication/2024/05/13/8f77e73a66a6f484c655985a/indeks-pembangunan-manusia-2023.html

[3] Badan Pusat Statistik Sultra, "Indeks Pembangunan Manusia (IPM) Provinsi Sulawesi Tenggara tahun 2023 mencapai 72,94," Badan Pusat Stistik Sultra. Accessed: Dec. 20, 2024. [Online]. Available: https://sultra.bps.go.id/id/pressrelease/2023/12/01/1058/the-human-development-index--hdi--of-southeast-sulawesi-province-in-2023-will-reach-72-94.html

[4] Norman R. Draper and Harry Smith, *Applied Regression Analysis*. John Wiley & Sons, Inc, 1998. doi: doi.org/10.1002/9781118625590.

[5] Adji Achmad Rinaldo Fernandes and Solimun, *Analisis Regresi Dalam Pendekatan Fleksibel*. Malang: UB Press, 2021. Accessed: Jan. 26, 2025. [Online]. Available: http://www.ubpress.ub.ac.id

[6] Randall L. Eubank, *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

[7] A. S. Suriaslan, I. N. Budiantara, and V. Ratnasari, "Nonparametric regression estimation using multivariable truncated splines for binary response data," *MethodsX*, vol. 14, Jun. 2025, doi: 10.1016/j.mex.2024.103084.

[8] W. Somayasa and G. Ngurah Adhi Wibawa, "PEMODELAN INDEKS PEMBANGUNAN MANUSIA DI SULAWESI TENGGARA TAHUN 2017-2020 MENGGUNAKAN METODE REGRESI DATA PANEL." [Online]. Available: www.sultra.bps.go.id

[9] M. N. Afiat, Z. Saenong, and Dewangga Puspa, "PENGARUH INDEKS PEMBANGUNAN MANUSIA DAN PERTUMBUHAN EKONOMI TERHADAP PENGANGGURAN DI SULAWESI TENGGARA," *JEMBA(Jurnal Ekonomi, Manajemen, Bisnis, dan Akuntansi*, pp. 571–580, 2024, Accessed: Feb. 03, 2025. [Online]. Available: http://bajangjournal.com/index.php/JEMBA

[10] R. Putra, M. G. Fadhlurrahman, and Gunardi, "Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation," *MethodsX*, vol. 10, Jan. 2023, doi: 10.1016/j.mex.2022.101994.

[11] P. J. . Green and B. W. . Silverman, *Nonparametric regression and generalized linear models : a roughness penalty approach*. Chapman & Hall/CRC, 2000.

[12] J. P. Lesage, "Spatial Econometrics," 1998. [Online]. Available: http://www.econ.utoledo.edu.

[13] L. Anselin, *Spatial Econometrics: Methods and Models*, vol. 4. in Studies in Operational Regional Science, vol. 4. Dordrecht: Springer Netherlands, 1988. doi: 10.1007/978-94-015-7799-1.

[14] H. Ilmi, S. Prangga, U. Mulawarman Jl Barong Tongkok No, K. Gunung Kelua, and K. Timur, "Geographically Weighted Spline Nonparametric Regression dengan Fungsi Pembobot Bisquare dan Gaussian Pada Tingkat Pengangguran Terbuka Di Pulau Kalimantan," 2021. [Online]. Available: www.unipasby.ac.id

[15] S. Sifriyani, "Simultaneous Hypothesis Testing of Multivariable Nonparametric Spline Regression in the GWR Model," *Int J Stat Probab*, vol. 8, no. 4, p. 32, 2019, doi: 10.5539/ijsp.v8n4p32.

[16] C. P. S. Purnamasari and Y. Widyaningsih, "Perbandingan Performa Bandwidth CV, AICc, dan BIC pada Model Geographically Weighted Regression (Aplikasi pada Data Pengangguran di Pulau Jawa)," *Inferensi*, vol. 1, no. 1, p. 71, Oct. 2023, doi: 10.12962/j27213862.v1i1.19130.

[17] S. H. Kartiko and I. N. Budiantara, "Development of nonparametric geographically weighted regression using truncated spline approach."