

# Test Efficiency Analysis of Parametric, Nonparametric, Semiparametric Regression in Spatial Data

Diah Ayu Widyastuti\*, Adji Achmad Rinaldo Fernandes, Henny Pramoedyo, Nurjannah, Solimun

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

Received June 8, 2020; Revised July 16, 2020; Accepted August 10, 2020

## Cite This Paper in the following Citation Styles

(a): [1] Diah Ayu Widyastuti, Adji Achmad Rinaldo Fernandes, Henny Pramoedyo, Nurjannah, Solimun, "Test Efficiency Analysis of Parametric, Nonparametric, Semiparametric Regression in Spatial Data," *Mathematics and Statistics*, Vol. 8, No. 5, pp. 506 - 519, 2020. DOI: 10.13189/ms.2020.080503.

(b): Diah Ayu Widyastuti, Adji Achmad Rinaldo Fernandes, Henny Pramoedyo, Nurjannah, Solimun (2020). *Test Efficiency Analysis of Parametric, Nonparametric, Semiparametric Regression in Spatial Data*. *Mathematics and Statistics*, 8(5), 506 - 519. DOI: 10.13189/ms.2020.080503.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** Regression analysis has three approaches in estimating the regression curve, namely: parametric, nonparametric, and semiparametric approaches. Several studies have discussed modeling with the three approaches in cross-section data, where observations are assumed to be independent of each other. In this study, we propose a new method for estimating parametric, nonparametric, and semiparametric regression curves in spatial data. Spatial data states that at each point of observation has coordinates that indicate the position of the observation, so between observations are assumed to have different variations. The model developed in this research is to accommodate the influence of predictor variables on the response variable globally for all observations, as well as adding coordinates at each observation point locally. Based on the value of Mean Square Error (MSE) as the best model selection criteria, the results are obtained that modeling with a nonparametric approach produces the smallest MSE value. So this application data is more precise if it is modeled by the nonparametric truncated spline approach. There are eight possible models formed in this research, and the nonparametric model is better than the parametric model, because the MSE value in the nonparametric model is smaller. As for the semiparametric regression model that is formed, it is obtained that the variable  $X_2$  is a parametric component while  $X_1$  and  $X_3$  are the nonparametric components (Model 2). The regression curve estimation model with a nonparametric approach tends to be more efficient than Model 2 because the linearity assumption test results show that the relationship of all the predictor variables to the response variable shows a non-linear

relationship. So in this study, spatial data that has a non-linear relationship between predictor variables and responses tends to be better modeled with a nonparametric approach.

**Keywords** Parametric, Nonparametric, Spatial, Semiparametric, Heteroscedasticity

## 1. Introduction

Regression analysis is one method that can be used to determine the relationship between variables involved in a study (Draper & Smith, 1992). Regression analysis that involves one response variable and several predictor variables is multiple linear regression analysis. According to (Kutner et al, 2005), multiple linear regression analysis requires several assumptions that must be fulfilled, namely linearity, error normalization, homogeneity of various errors, non-autocorrelation, and non-multicollinearity. There are three approaches to regression analysis including parametric, nonparametric, and semiparametric approaches.

The parametric approach used in the shape of the regression curve is known as a linear, quadratic, cubic, or polynomial degree  $k$  (Fernandes et al, 2014). A nonparametric approach used in the shape of the regression curve is unknown, while the semiparametric approach is modeling the regression analysis used when some form of the regression curve is known and partly unknown. When

research data shows an unknown shape of the regression curve and the linearity assumption is not met, it is necessary to do statistical modeling with a nonparametric approach (Fernandes et al, 2015).

Some regression models with nonparametric approaches that are often used by researchers include Spline, Kernel, Fourier, and Wavelet (Eubank, 1999). Spline regression is a regression analysis method that can be used to estimate nonparametric regression models. Data that has a changing pattern at certain subintervals is very well modeled with splines (Hardle, 1990). Spline has piecewise polynomial properties in which a piece of the polynomial has segmented properties at the interval  $k$  formed at the knot point. One approach that can be used for parameter estimation in nonparametric regression models is truncated spline, where the truncated spline approach can overcome the changing data patterns at certain subintervals.

There is a development method of multiple linear regression, namely statistical modeling based on regional characteristics, where the statistical model that is formed is influenced by the geographical location between regions (Lu et al, 2014). Differences in geographical location affect the potential that is owned or used by an area. Therefore we need a statistical modeling method that considers geographical location or location observation factors.

Based on the three approaches in regression analysis, this study will discuss the comparison of the three approaches when used to that is formed is influenced by the geographical location between regions. The selection of the best model is seen based on the value of Mean Square Error (MSE). With this research, it is expected to be able to show the right regression analysis approach to be used if the research data does not follow a certain pattern. Regression curve estimation is done by the Weighted Least Square (WLS) method, and the estimated regression curve obtained applies globally and locally.

## 2. Literature Review

### 2.1. Spatial Data

Spatial data is data that contains geographical information so that it can be described in a map. In spatial data, there is a dependency between observation locations. The difference between spatial data and other data is that there are coordinates that indicate the location points according to geographical conditions (Anselin, 1988).

### 2.2. Regression Analysis

There are several regression analysis approaches based on data patterns, namely the parametric regression approach, nonparametric regression, and semiparametric regression. If the pattern of the relationship between the response variable and the predictor is known then it is called parametric regression analysis (Kutner et al, 2005), whereas if the pattern of the relationship between the

response variable and the predictor does not know the shape of the regression curve or there is no past information about the data pattern then the approach used is nonparametric regression (Fernandes et al, 2014). In addition to the two approaches, there is a semiparametric regression approach which states that this approach used in the shape of the regression curve is partially known and partly unknown (Eubank, 1999).

#### 2.2.1. Parametric Regression

In parametric regression, several classical assumptions must be fulfilled. One such assumption is that the shape of the regression curve is known, for example, as linear, quadratic, cubic,  $p$ -degree polynomials, exponents, and so on. Parametric regression functions can be written with the following equation:

$$y_i = f(x_i) + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (1)$$

$f(x_i)$  is a parametric regression function with  $\varepsilon_i$  as random errors that follow a normal distribution with zero mean and variance  $\sigma^2$ . Parametric regression functions are accompanied by linear functions as follows (Pramoedyo, 2013):

$$f(x_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} \quad (2)$$

where:

- $y_i$  : response variable at the  $i$ -th observation
- $\beta_0$  : intersection point between a regression line and the  $y$  axis (intercept)
- $\beta_1, \dots, \beta_p$  : regression coefficient for each  $p$ -predictor variable
- $X_{ip}$  : the value of the  $i$ -th observation on the  $p$ -predictor variable
- $\varepsilon_i$  : error in  $i$ -th observation
- $n$  : number of observations;  $i : 1, 2, \dots, n$
- $p$  : number of predictors variable

If parametric regression is applied to spatial data, the following equation will be obtained:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \quad (3)$$

Coordinates  $(u_i, v_i)$  indicate the location of each observation. In parametric modeling with spatial data, we will get as many models as the location of the observation, and this shows that each location of the observation shows a different effect between the predictor variables on the response variable (Lu et al, 2014). So that if equations (2) and (3) are combined, a parametric regression curve estimation model will be obtained globally for all locations and locally for each observation location with the following equation:

$$y_i = \beta_0 + \beta_0(u_i, v_i) + \sum_{k=1}^p [\beta_{1k} X_{ik} + \beta_{2k}(u_i, v_i) X_{ik}] + \varepsilon_i \quad (4)$$

where:

$X_{ik}$  : the value of the i-th observation on the k-th parametric predictor variable  
 $k : 1, 2, \dots, p$

2.2.2. Nonparametric Regression

Nonparametric regression is a statistical method used to estimate the pattern of relationships between predictor variables and responses when no information is obtained about the shape of a function or regression curve. In nonparametric regression, there are no classical assumptions as in parametric regression (Fernandes et al, 2019). Based on  $n$  independent observations, the relationship from variables  $y_i$  and  $x_i$  is unknown. According to (Hardle, 1990), a nonparametric regression model with more than one nonparametric predictor variable is as follows:

$$y_i = \sum_{j=1}^q f_j(x_{ij}) + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (5)$$

where:

$y_i$  : response variable from i-th observations  
 $f_j(x_{ij})$  : nonparametric regression function i-th observation on the j-th predictor variable  
 $\varepsilon_i$  : error is assumed independent with zero mean and variance  $\sigma^2$

Truncated Spline is one of the approach methods in the nonparametric regression model that is often used. Truncated Spline is polynomial pieces that have segmented and continuous properties. One of the advantages of the truncated spline approach in nonparametric regression tends to find its estimation of the regression function according to the data.

$f(x_i)$  is a function of a regression curve whose shape is unknown and assumed to be additive. If the nonparametric regression function is approximated by the truncated spline function, then it can be written in the equation as follows:

$$f(x_i) = \sum_{k=1}^m \beta_m X_i^k + \sum_{h=1}^r \alpha_{(h+m)} (X_i - k_h)_+^m \quad (6)$$

If the nonparametric predictor variable is more than one, the data pair arrangement is  $(y_i, x_{i1}, x_{i2}, \dots, x_{iq})$  so the nonparametric regression model that is formed is as follows.

$$y_i = \sum_{j=1}^q f_j(x_{ij}) + \varepsilon_i$$

$$y_i = \sum_{j=1}^q \sum_{k=1}^m \beta_{jk} X_{ij}^k + \sum_{j=1}^q \sum_{h=1}^r \alpha_{j(m+h)} (X_{ij} - k_{hj})_+^m + \varepsilon_i \quad (7)$$

and truncated functions as follows:

$$(X_{ij} - k_{hj})_+^m = \begin{cases} (X_{ij} - k_{hj}), & X_{ij} \geq k_{hj} \\ 0, & X_{ij} < k_{hj} \end{cases}$$

where:

$X_{ji}$  : j-th nonparametric predictor variable  
 $k_{hj}$  : the h-knot point of the j-th nonparametric predictor variable  
 $\beta_{jk}$  : the k-th parameter in the j-th nonparametric predictor variable  
 $\alpha_{j(m+h)}$  : parameters of the truncated spline function

The truncated spline nonparametric regression model in data with coordinate locations in the development of nonparametric regression, where the model is applied to spatial data (Sifriyani et al., 2018), so the estimated parameters generated are local for each observation location. The truncated spline approach is used to solve spatial data analysis problems for which the regression curve is unknown (Sifriyani et al, 2017).

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^q \sum_{k=1}^m \beta_{jk}(u_i, v_i) X_{ij}^k + \sum_{j=1}^q \sum_{h=1}^r \alpha_{j(m+h)}(u_i, v_i) (X_{ij} - K_{hj})_+^m + \varepsilon_i \quad (8)$$

Equation (9) shows a global nonparametric model so that in all locations of observation the predictor variables have the same effect on the response variable. If equation (9) is applied to spatial data, we will obtain a global nonparametric regression curve estimation model for all locations and locally for each observation location with the following equation:

$$y_i = \beta_0 + \beta_0(u_i, v_i) + \sum_{j=1}^q \left[ \beta_{1j} X_{ij} + \alpha_j (X_{ij} - k_{x_j}) + \beta_{2j}(u_i, v_i) X_{ij} + \alpha_{j1}(u_i, v_i) (X_{j1i} - k_{x_{j1}}) \right] + \varepsilon_i \quad (9)$$

2.2.3. Semiparametric Regression

Semiparametric regression combined between parametric regression and nonparametric regression. According to Eubank (1999), semiparametric regression states that the regression curve is partly known and partly unknown. The truncated spline semiparametric regression equation with more than one nonparametric predictor variable is as follows:

$$y_i = \beta_0 + \beta_0(u_i, v_i) + \sum_{k=1}^p [\beta_{1k} X_{ik} + \beta_{2k}(u_i, v_i) X_{ik}] + \sum_{j=1}^q [\beta_{1j} X_{ij} + \alpha_j (X_{1j} - k_{x_j})_+ + \beta_{2j}(u_i, v_i) X_{ij} + \alpha_{j1}(u_i, v_i) (X_{j1} - k_{x_{j1}})_+] + \varepsilon_i \tag{10}$$

where:

- $p$  : number of parametric predictors variable
- $q$  : number of nonparametric predictors variable

Based on equation (10), it can be seen that in this equation, there is a global and local influence. Global influence does not involve the coordinates of the point of observation so that all observation locations have the same influence (Fernandes et al, 2020). While the influence locally gives a different effect at each observation location. The regression curve estimation method used in the three approaches is the Weighted Least Square (WLS) method (Fernandes et al, 2017), where the weighting indicates that there is an influence of heterogeneity of variance globally for all observations and locally for each observation location.

### 2.3. Testing Linear Assumptions

The linearity assumption states that the relationship between the response variable and the predictor variable is appropriate, which means that the regression curve can be expressed in a linear, quadratic, or cubic form. If the linearity assumption is not met, then the linear regression analysis with the parametric approach is not suitable for use in data analysis. One method for testing linearity assumptions is the Regression Specification Error Test or RESET, which was first introduced in 1969 by Ramsey.

According to (Gujarati, 2003) steps to implement the RESET, namely:

1. Perform a regression analysis using one predictor variable to get the fitted value of the response variable from the following equation.

$$y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \tag{11}$$

From equation (11), parameter estimation using Ordinary Least Square (OLS) method and the coefficient determination is obtained with following equation:

$$R_1^2 = \frac{SSE_1}{SST_1} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{i1})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where:

- $SSE_1$  : sum square error from equation (11)
- $SST_1$  : sum square total from equation (11)

$y_i$  : response variable

$\hat{y}_{i1}$  : predicted value of response variable from equation (11)

$\bar{y}$  : mean value of response variable from equation (11)

2. Perform a regression analysis by entering the fitted value obtained from equation (11) as a new predictor variable with the regression equation model as follows:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \hat{y}_i^2 + \beta_3 \hat{y}_i^3 + \varepsilon_i \tag{12}$$

Based on equation (13) the coefficient determination is obtained with the following equation:

$$R_2^2 = \frac{SSE_2}{SST_2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{i2})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where:

$SSE_2$  : sum square error from equation (13)

$SST_2$  : sum square total from equation (13)

$\hat{y}_{i2}$  : predicted value of response variable from equation (13)

$\bar{y}$  : mean value of response variable from equation (13)

3. Then, the value  $R_1^2$  of equation (11) and the value  $R_2^2$  of equation (12) are obtained. After the values  $R_1^2$  and  $R_2^2$  of the two equations are known, then calculate the value of the test statistic  $F_{test}$  with the following equation:

$$F_{test} = \frac{(R_2^2 - R_1^2) / m}{(1 - R_2^2) / (n - k)} \tag{13}$$

where:

- $m$  : number of predictor variables that have been added
- $n$  : number of observations
- $k$  : number of parameters in the new equation

$R_1^2$  : coefficient determination from equation (11)

$R_2^2$  : coefficient determination from equation (12)

4. Based on the value  $F_{test}$  of equation (13), then compare the value with  $F_{table}$  the following hypothesis:

$H_0: \beta_2 = \beta_3 = 0$  vs

$H_1: \text{at least one of } \beta_j \neq 0; j = 2, 3$

If  $F_{test} > F_{table}$  or p-value  $< \alpha = 0,05$ ,  $H_0$  is rejected. The specification of the model used is a non-linear model.

### 3. Research Methodology

In the study, discussing farmer satisfaction with subsidized fertilizers from the government with the research variables used are as follows: Courage of a field counselor (Y), Nation Culture ( $X_1$ ), Reward Financial Courage of a field counselor ( $X_2$ ), and Leadership Role ( $X_3$ ). The composition of the research data consists of three cultures that exist in East Java Province, wherein each culture consists of five regencies that have the coordinates of the observation location. In each culture, analysis was carried out to determine the level of farmer satisfaction, research was conducted to model farmers' satisfaction with subsidized fertilizer as a whole culture and in each culture. Modeling is done with three approaches, namely parametric, nonparametric, and semiparametric so that the most appropriate modeling is obtained to represent farmer satisfaction data.

## 4. Results and Discussion

### 4.1. Testing Linear Assumptions

Linearity assumption testing is used to find out whether the relationship between response variables and predictor variables can be stated precisely. This means that the regression curve can be expressed in a linear, quadratic, or cubic form.

Based on Table 1, it can be seen that all the predictor variables involved have a non-linear relationship to the

response variable. Thus, insufficient evidence of the existence of a linear relationship pattern is used. The next step is modeling with three approaches at once, namely parametric, nonparametric, and semiparametric approaches. Based on the three approaches, modeling results will be obtained with the most efficient approach to represent data that has a non-linear relationship between the predictor variable and the response variable. Modeling on these three approaches is done by combining classical regression models with spatial data regression involving coordinates at each observation location.

**Table 1.** The Results of Linearity Test

Relationship	p-value	Information
$X_1 \rightarrow y$	7,8742e-10	Non-linear
$X_2 \rightarrow y$	1,1029e-11	Non-linear
$X_3 \rightarrow y$	1,8683e-14	Non-linear

Note: non-linear relationship (p-value  $< 0.05$ )

### 4.2. Parametric Regression

Following are the equations obtained with the parametric approach globally for all observations and locally according to each coordinate of the observation location:

**Table 2.** Estimation of Global Parametric Regression Curve

Parameter	Coefficients
	0,0096
	0,2454
	0,1819
	0,3005

Based on Table 2, the resulting curve estimates are valid for all observation locations without considering the coordinates of the location where the observations are. So the model formed is as follows:

$$y_i = 0,0096 + 0,2454X_{1i} + 0,1819X_{2i} + 0,3005X_{3i}$$

**Table 3.** Estimation of Local Parametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1442	$\beta_{02}$	-0,0493	$\beta_{03}$	-0,0852
$\beta_{211}$	-0,0294	$\beta_{212}$	0,1419	$\beta_{213}$	0,1329
$\beta_{221}$	0,0295	$\beta_{222}$	0,0720	$\beta_{223}$	0,0804
$\beta_{231}$	0,0563	$\beta_{232}$	0,1313	$\beta_{233}$	0,1129

Locally estimated regression curves are obtained because of the coordinates at each observation location. Based on Table 3, three regions produce estimations of the regression curve locally, namely regions with culture 1, culture 2, and culture 3 as follows:

Culture 1:

$$y_{1i} = 0,1442 - 0,0294 X_{11i} + 0,0295 X_{21i} + 0,0563 X_{31i}$$

Culture 2:

$$y_{2i} = -0,0493 + 0,1419 X_{12i} + 0,0720 X_{22i} + 0,1313 X_{32i}$$

Culture 3:

$$y_{3i} = -0,0852 + 0,1329 X_{13i} + 0,0804 X_{23i} + 0,1129 X_{33i}$$

**4.3. Nonparametric Regression**

**Table 4.** Estimation of Global Nonparametric Regression Curve

Parameter	Coefficients
	0,0475
	0,2182
	0,1687
	0,2615
	0,0512
$\alpha_2$	0,0232
$\alpha_3$	0,1345

Modeling with a nonparametric approach is done globally and locally so that all the predictor variables involved in the research are assumed to have non-linear relationships. Here are the results of estimating global nonparametric regression curves, in which there are three nonparametric predictor variables.

Estimation of nonparametric regression curves with a first-order truncated spline approach with the point of the optimum knot is as follows:

$$y_i = 0,0475 + 0,2182 X_{1i} + 0,1687 X_{2i} + 0,2615 X_{3i} + 0,0512(X_{1i} - 0,4671)_+ + 0,0232(X_{2i} - 0,4881)_+ + 0,1345(X_{3i} - 0,4950)_+$$

From Table 5 the estimation of the nonparametric regression curve using the first-order truncated spline approach in each culture as follows:

Culture 1:

$$y_{1i} = 0,1293 - 0,0127 X_{11i} - 0,0162 X_{21i} + 0,0335 X_{31i} - 0,0111(X_{11i} - 0,4462)_+ + 0,0548(X_{21i} - 0,6742)_+ + 0,0168(X_{31i} - 0,6556)_+$$

Culture 2:

$$y_{2i} = -0,0237 + 0,1211 X_{12i} + 0,0116 X_{22i} + 0,0664 X_{32i} - 0,0025(X_{12i} - 0,8048)_+ + 0,1131(X_{22i} - 0,7899)_+ + 0,0314(X_{32i} - 0,7357)_+$$

Culture 3:

$$y_{3i} = -0,0581 + 0,1098 X_{13i} + 0,0195 X_{23i} + 0,0687 X_{33i} - 0,0376(X_{13i} - 0,4684)_+ + 0,0937(X_{23i} - 0,7755)_+ + 0,0725(X_{33i} - 0,6841)_+$$

**Table 5.** Estimation of Local Nonparametric Regression Curve

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1293	$\beta_{02}$	-0,0237	$\beta_{03}$	-0,0581
$\beta_{211}$	-0,0127	$\beta_{212}$	0,1211	$\beta_{213}$	0,1098
$\beta_{221}$	-0,0162	$\beta_{222}$	0,0116	$\beta_{223}$	0,0195
$\beta_{231}$	0,0335	$\beta_{232}$	0,0664	$\beta_{233}$	0,0687
$\alpha_{11}$	-0,0111	$\alpha_{12}$	-0,0025	$\alpha_{13}$	-0,0376
$\alpha_{21}$	0,0548	$\alpha_{22}$	0,1131	$\alpha_{23}$	0,0937
$\alpha_{31}$	0,0168	$\alpha_{32}$	0,0314	$\alpha_{33}$	0,0725

**4.4. Semiparametric Regression**

Semiparametric regression modeling is used to model parametric and nonparametric regression simultaneously. The following will discuss six possible models that were formed using the semiparametric approach.

**4.4.1. Model 1**

The first model formed is assuming the variable  $X_1$  as a parametric component, while  $X_2$  and  $X_3$  as the nonparametric component.

**Table 6.** Estimation of Global Semiparametric Regression Curve

Parameter	Coefficient
	0,0649
	0,2502
	0,2266
	0,1155
$\alpha_2$	0,1470
$\alpha_3$	0,1075

Table 6 states that the semiparametric curve estimation is done globally for all observations. Estimation is carried out with the assumption that part of the regression curve has a known shape and part that has no / unknown shape. The equation obtained is as follows:

$$y_i = 0,0649 + 0,2502 X_{1i} + 0,2266 X_{2i} + 0,1155 X_{3i} + 0,1470(X_{2i} - 0,4881)_+ + 0,1075(X_{3i} - 0,4950)_+$$

Local semiparametric regression curve estimates cause each observation location to give different results. The semiparametric model is arranged based on optimal knot points, where each location gives different pieces of knot points with the following results:

Culture 1:

$$y_{1i} = 0,1789 - 0,0637 X_{11i} + 0,0080 X_{21i} - 0,0051 X_{31i} + 0,0424(X_{21i} - 0,6742)_+ + 0,0180(X_{31i} - 0,6556)_+$$

Culture 2:

$$y_{2i} = -0,0191 + 0,1169 X_{12i} + 0,0558 X_{22i} - 0,0069 X_{32i} + 0,0981(X_{22i} - 0,7899)_+ + 0,0263(X_{32i} - 0,7357)_+$$

Culture 3:

$$y_{3i} = -0,0949 + 0,1969 X_{13i} + 0,1629 X_{23i} - 0,0267 X_{33i} - 0,0250(X_{23i} - 0,7755)_+ + 0,0384(X_{33i} - 0,6841)_+$$

**Table 7.** Estimation of Local Semiparametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1789	$\beta_{02}$	-0,0191	$\beta_{03}$	-0,0949
$\beta_{211}$	-0,0637	$\beta_{212}$	0,1169	$\beta_{213}$	0,1969
$\beta_{221}$	0,0080	$\beta_{222}$	0,0558	$\beta_{223}$	0,1629
$\beta_{231}$	-0,0051	$\beta_{232}$	-0,0069	$\beta_{233}$	-0,0267
$\alpha_{21}$	0,0424	$\alpha_{22}$	0,0981	$\alpha_{23}$	-0,0250
$\alpha_{31}$	0,0180	$\alpha_{32}$	0,0263	$\alpha_{33}$	0,0384

4.4.2. Model 2

The second model formed to compile the estimation of the regression curve is by assuming  $X_2$  as a parametric component, while  $X_1$  and  $X_3$  as nonparametric components.

**Table 8.** Estimation of Global Semiparametrik Regression Curve

Parameter	Coefficients
	0,0787
	0,2326
	0,2164
	0,1147
	0,1635
$\alpha_3$	0,0974

Based on Table 8, the estimation is done globally for all

observation locations so that the estimated curve model generated by the order 1 truncated spline approach and the point of the optimum knot is as follows:

$$y_i = 0,0787 + 0,2326 X_{1i} + 0,2164 X_{2i} + 0,1147 X_{3i} + 0,1635 (X_{1i} - 0,4671)_+ + 0,0974 (X_{3i} - 0,4950)_+$$

Table 9 presents the results of the estimated semiparametric regression curves for each of the observation sites which include Culture 1, Culture 2, and Culture 3 with the following results:

Culture 1:

$$y_{1i} = 0,1436 - 0,0219 X_{11i} - 0,0606 X_{21i} + 0,0293 X_{31i} + 0,0531 (X_{21i} - 0,6742)_+ + 0,0168 (X_{31i} - 0,6556)_+$$

Culture 2:

$$y_{2i} = -0,0127 + 0,1079 X_{12i} + 0,0121 X_{22i} + 0,0619 X_{32i} + 0,0531 (X_{22i} - 0,6742)_+ + 0,0168 (X_{32i} - 0,6556)_+$$

Culture 3:

$$y_{3i} = -0,0523 + 0,1465 X_{13i} + 0,1298 X_{23i} + 0,1252 X_{33i} - 0,0322 (X_{23i} - 0,7755)_+ + 0,0256 (X_{33i} - 0,6841)_+$$

**Table 9.** Estimation of Local Semiparametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1436	$\beta_{02}$	-0,0127	$\beta_{03}$	-0,0523
$\beta_{211}$	-0,0219	$\beta_{212}$	0,1079	$\beta_{213}$	0,1465
$\beta_{221}$	-0,0606	$\beta_{222}$	0,0121	$\beta_{223}$	0,1298
$\beta_{231}$	0,0293	$\beta_{232}$	0,0619	$\beta_{233}$	0,1252
$\alpha_{11}$	0,0531	$\alpha_{12}$	0,0939	$\alpha_{13}$	-0,0322
$\alpha_{31}$	0,0168	$\alpha_{32}$	0,0240	$\alpha_{33}$	0,0256

4.4.3. Model 3

The third equation that may be formed if a semiparametric approach is used, namely  $X_3$  as a parametric component, and  $X_1$  and  $X_2$  as nonparametric components.

**Table 10.** Estimation of Global Semiparametric Regression Curve

Parameter	Coefficients
	0,0702
	0,2232
	0,1967
	0,1539
	0,1562
$\alpha_2$	0,1258

The estimation of the global regression curve states that

the effect of the predictor variables on the response variable is the same for all observation locations namely Culture 1, Culture 2, and Culture 3. Based on Table 10, then the equation can be written as follows:

$$y_i = 0,0702 + 0,2232 X_{1i} + 0,1967 X_{2i} + 0,1539 X_{3i} + 0,1562 (X_{1i} - 0,4671)_+ + 0,1258 (X_{2i} - 0,4881)_+$$

Based on Table 11, the estimated results of the regression curve for each observation location are obtained. The estimation of the regression curve locally shows that the influence of the predictor variable on the response variable at each location is different.

Culture 1:

$$y_{1i} = 0,1481 - 0,0246 X_{11i} - 0,0644 X_{21i} + 0,0276 X_{31i} - 0,0011 (X_{11i} - 0,4462)_+ + 0,0521 (X_{21i} - 0,6742)_+$$

Culture 2:

$$y_{2i} = -0,0140 + 0,1077 X_{12i} + 0,0109 X_{22i} + 0,0583 X_{32i} - 0,0030 (X_{12i} - 0,8048)_+ + 0,0951 (X_{22i} - 0,7899)_+$$

Culture 3:

$$y_{3i} = -0,0639 + 0,1400 X_{13i} + 0,1279 X_{23i} + 0,1109 X_{33i} - 0,0240 (X_{13i} - 0,4684)_+ + 0,0066 (X_{23i} - 0,6556)_+$$

**Table 11.** Estimation of Local Semiparametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1481	$\beta_{02}$	-0,0140	$\beta_{03}$	-0,0639
$\beta_{211}$	-0,0246	$\beta_{212}$	0,1077	$\beta_{213}$	0,1400
$\beta_{221}$	-0,0644	$\beta_{222}$	0,0109	$\beta_{223}$	0,1279
$\beta_{231}$	0,0276	$\beta_{232}$	0,0583	$\beta_{233}$	0,1109
$\alpha_{11}$	-0,0011	$\alpha_{12}$	-0,0030	$\alpha_{13}$	-0,0240
$\alpha_{21}$	0,0521	$\alpha_{22}$	0,0951	$\alpha_{23}$	0,0066

4.4.4. Model 4

The fourth model that is formed is by assuming that  $X_1$  and  $X_2$  as parametric components, while  $X_3$  as nonparametric components.

**Table 12.** Estimation of Global Semiparametric Regression Curve

Parameter	Coefficients
	0,0435
	0,2804
	0,2568
	0,1138
$\alpha_3$	0,1225

Estimates of the global regression curve where  $X_1$  and

$X_2$  are parametric components, while  $X_3$  as nonparametric components are as follows:

$$y_i = 0,0435 + 0,2804 X_{1i} + 0,2568 X_{2i} + 0,1138 X_{3i} + 0,1225 (X_{3i} - 0,4950)_+$$

Based on Table 13, the estimated regression curve can be stated as follows:

Culture 1:

$$y_{1i} = 0,1742 - 0,0603 X_{11i} + 0,0144 X_{21i} + 0,0491 X_{31i} + 0,0212 (X_{31i} - 0,6556)_+$$

Culture 2:

$$y_{2i} = -0,0408 + 0,1337 X_{12i} + 0,0674 X_{22i} + 0,1128 X_{32i} + 0,0323 (X_{32i} - 0,7357)_+$$

Culture 3:

$$y_{3i} = -0,0899 + 0,2069 X_{13i} + 0,1749 X_{23i} - 0,0481 X_{33i} + 0,0378 (X_{33i} - 0,6841)_+$$

**Table 13.** Estimation of Local Semiparametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1742	$\beta_{02}$	-0,0408	$\beta_{03}$	-0,0899
$\beta_{211}$	-0,0603	$\beta_{212}$	0,1337	$\beta_{213}$	0,2069
$\beta_{221}$	0,0144	$\beta_{222}$	0,0674	$\beta_{223}$	0,1749
$\beta_{231}$	0,0491	$\beta_{232}$	0,1128	$\beta_{233}$	-0,0481
$\alpha_{31}$	0,0212	$\alpha_{32}$	0,0323	$\alpha_{33}$	0,0378

## 4.4.5. Model 5

The fifth model is by assuming  $X_1$  and  $X_3$  as parametric components,  $X_2$  as nonparametric components.

**Table 14.** Estimation of Global Semiparametric Regression Curve

Parameter	Coefficients
	0,0370
	0,2637
	0,2275
	0,1622
$\alpha_2$	0,1580

Table 14 shows the results of the global semiparametric regression curve estimation. The effect on all observation locations is assumed to be the same as the following results:

$$y_i = 0,0370 + 0,2637 X_{1i} + 0,2275 X_{2i} + 0,1622 X_{3i} + 0,1580 (X_{2i} - 0,4881)_+$$

The local semiparametric regression curve estimation based on Table 15, states that each observation location shows a different effect between the predictor variables on the response variable.

**Table 15.** Estimation of Local Semiparametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1802	$\beta_{02}$	-0,0382	$\beta_{03}$	-0,1050
$\beta_{211}$	-0,0645	$\beta_{212}$	0,1308	$\beta_{213}$	0,1973
$\beta_{221}$	0,0109	$\beta_{222}$	0,0612	$\beta_{223}$	0,1555
$\beta_{231}$	0,0465	$\beta_{232}$	0,1120	$\beta_{233}$	0,0037
$\alpha_{21}$	0,0202	$\alpha_{22}$	0,0324	$\alpha_{23}$	0,0412

Culture 1:

$$y_{1i} = 0,1802 - 0,0645 X_{11i} + 0,0109 X_{21i} + 0,0465 X_{31i} + 0,0202 (X_{21i} - 0,6742)_+$$

Culture 2:

$$y_{2i} = -0,0382 + 0,1308 X_{12i} + 0,0612 X_{22i} + 0,1120 X_{32i} + 0,0324 (X_{22i} - 0,7899)_+$$

Culture 3:

$$y_{3i} = -0,1050 + 0,1973 X_{13i} + 0,1555 X_{23i} + 0,0037 X_{33i} + 0,0412 (X_{23i} - 0,7755)_+$$

The sixth model that was formed is by assuming variables  $X_2$  and  $X_3$  as parametric components, while  $X_1$  as nonparametric components.

Global semiparametric regression curve estimates from the sixth model based on Table 16 are stated as follows.

$$y_i = 0,0539 + 0,2445 X_{1i} + 0,2173 X_{2i} + 0,1558 X_{3i} + 0,1813 (X_{1i} - 0,4671)_+$$

Based on Table 17, semiparametric regression curve estimation locally gives a difference with the estimated parameter globally. The results of the estimation of the regression curve for each observation location are as follows:

Culture 1:

$$y_{1i} = 0,1443 - 0,0221 X_{11i} - 0,0607 X_{21i} + 0,0322 X_{31i} + 0,0569 (X_{11i} - 0,4462)_+$$

Culture 2:

$$y_{2i} = -0,0298 + 0,1201 X_{12i} + 0,0138 X_{22i} + 0,0668 X_{32i} + 0,1062 (X_{12i} - 0,8048)_+$$

Culture 3:

$$y_{2i} = -0,0298 + 0,1201 X_{12i} + 0,0138 X_{22i} + 0,0668 X_{32i} + 0,1062 (X_{12i} - 0,8048)_+$$

4.4.6. Model 6

**Table 16.** Estimation of Global Semiparametric Regression Curve

Parameter	Coefficients
$\beta_0$	0,0539
$\beta_{11}$	0,2445
$\beta_{12}$	0,2173
$\beta_{13}$	0,1558
$\alpha_1$	0,1813

**Table 17.** Estimation of Local Semiparametric Regression Curves

Culture 1		Culture 2		Culture 3	
Parameter	Coefficients	Parameter	Coefficients	Parameter	Coefficients
$\beta_{01}$	0,1443	$\beta_{02}$	-0,0298	$\beta_{03}$	-0,0605
$\beta_{211}$	-0,0221	$\beta_{212}$	0,1201	$\beta_{213}$	0,1465
$\beta_{221}$	-0,0607	$\beta_{222}$	0,0138	$\beta_{223}$	0,1314
$\beta_{231}$	0,0322	$\beta_{232}$	0,0668	$\beta_{233}$	0,1184
$\alpha_{11}$	0,0569	$\alpha_{12}$	0,1062	$\alpha_{13}$	-0,0073

**4.5. The Efficiency Model**

Based on the results obtained, there are three regression approaches curve estimation models that can be compiled. From the curve estimation models obtained, the best model will be selected. The best model is seen from the Mean Square Error (MSE) value for each model with the following results.

**Table 18.** MSE from Parametric and Nonparametric Models

Model	MSE
Parametric	0,5519
Nonparametric	0,3995

Based on Table 18, the model formed based on the estimated regression curve shows that the nonparametric model has a smaller MSE value than the parametric model.

**Table 19.** MSE from Semiparametric Models

Model	MSE	
Semiparametric	Model 1	0,4188
	Model 2	0,4063
	Model 3	0,4132
	Model 4	0,4132
	Model 5	0,4152
	Model 6	0,4202

Table 19 shows some semiparametric models that are formed based on the results of the estimation of the regression curve. Based on the six models formed, it is found that in Model 2 the MSE values are the smallest compared to the other models.

**5. Conclusions**

This study discusses the estimation of the regression curve carried out with three approaches. The approach is

parametric, nonparametric, and semiparametric. In this modeling combined with global and local curve estimation. In this study, there are three locations namely Culture 1, Culture 2, and Culture 3. Estimation of the regression curve when done globally will provide the same effect of the predictor variables on the response variable, while the estimation of the regression curve locally gives different results at each observation location. Then based on regression curve estimation models that are formed the best model is selected with the criteria for selecting the best model using MSE, the smaller the MSE value, the better the model formed. Of the eight possible models formed, the nonparametric model is better than the parametric model because the MSE value in the nonparametric model is smaller. As for the semiparametric regression model that is formed, it is obtained that Model 2 has the smallest MSE value, wherein Model 2 it is assumed that the variable  $X_2$  is a parametric component while  $X_1$  and  $X_3$  are the nonparametric components.

The regression curve estimation model with a nonparametric approach tends to be more efficient than Model 2 because the linearity assumption test results show that the relationship of all the predictor variables to the response variable shows a non-linear relationship. So in this study, spatial data that has a non-linear relationship between predictor variables and responses tends to be better modeled with a nonparametric approach.

**REFERENCES**

[1] Anselin, L., “*Spatial Econometrics: Method and Models*”, Netherlands, Kluwer Academic Publishers, 1988.  
 [2] Fernandes, A.A.R., Nyoman Budiantara, I., Otok, B.W., Suhartono, “Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data”, *International Journal of Mathematical Analysis*, Vol 8 No

- 40, pp 1951-1961, 2014.
- [3] Fernandes, A.A.R, Budiantara, I.N, Otok, B.W., and Suhartono, "Spline Estimator for Bi-Responses and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data", *Journal of Mathematics and Statistics*, Vol 11, No 2, pp. 61-69, 2015.
- [4] Fernandes, A.A.R., Solimun, & Arisoelaningsih, E, "Estimation of spline function in nonparametric path analysis based on penalized weighted least square (PWLS)", *AIP Conference Proceedings*, Vol 1913 No 1, pp 020037, 2017.
- [5] Fernandes, A.A.R., Hutahayan, B., Solimun, Arisoelaningsih, E., Yanti, I., Astuti, A.B., Nurjannah, & Amaliana, L, "Comparison of Curve Estimation of the Smoothing Spline Nonparametric Function Path Based on PLS and PWLS In Various Levels of Heteroscedasticity", *IOP Conference Series: Materials Science and Engineering*, Forthcoming Issue, 2019.
- [6] Fernandes, A.A.R., Widiastuti, D.A., Nurjannah, "Smoothing spline semiparametric regression model assumption using PWLS approach", *International Journal of Advanced Science and Technology*, 29(4), pp. 2059-2070, 2020.
- [7] Gujarati, D. N, "*Basic Econometrics* (Fourth Edi)", New York, McGraw-Hill, 2003.
- [8] Hardle, W, "*Applied Nonparametric Regression*", New York, Cambridge University Press, 1990.
- [9] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W, "*Applied Linear Statistical Models* (Fifth Edit)", New York, McGraw-Hill, 2005.
- [10] Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S, "Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data". *International Journal of Geographical Information Science*, 28(4), 660–681, 2014, <https://doi.org/10.1080/13658816.2013.865739>
- [11] Sifriyani, Haryatmi, Budiantara, I. N., & Gunardi, "Geographically Weighted Regression with Spline Approach". *Far East Journal of Mathematical Sciences (FJMS)*, 101(6), 1183–1196, 2017, <https://doi.org/http://dx.doi.org/10.17654/MS101061183>
- [12] Sifriyani, Kartiko, S. H., Budiantara, I. N., & Gunardi, "Development of nonparametric geographically weighted regression using truncated spline approach". *Songklanakarinn J. Sci. Technol.*, 40(4), 909–920, 2018.