Far East Journal of Mathematical Sciences (FJMS)

http://www.pphmj.com
http://dx.doi.org/10.17654/MS101061183
Volume 101, Number 6, 2017, Pages 1183-1196 **ISSN: 0972-0871**

GEOGRAPHICALLY WEIGHTED REGRESSION WITH SPLINE APPROACH

Sifriyani[1], Haryatmi[1], I. N. Budiantara[2] and Gunardi[1]

[1]Gadjah Mada University
Indonesia

[2]Institut Tekhnologi Sepuluh Nopember
Indonesia

Abstract

Geographically weighted truncated spline nonparametric regression is a new method of statistical science. It is used to solve the problems of regression analysis of spatial data whose regression curve is unknown. This method is the development of nonparametric regression with truncated spline function approach to the analysis of spatial data. Truncated spline approach can be a solution for the problem of modeling spatial data analysis. The data patterns between the response variable and the predictor variable are unknown or regression curve is not known. This study is focused on finding estimator of truncated spline nonparametric regression in geographically weighted regression models with weighted maximum likelihood estimator (MLE) method. The characteristic of the unbiased estimator is also investigated. The results show that the nonparametric regression with truncated spline

Received: August 4, 2016; Accepted: August 20, 2016

2010 Mathematics Subject Classification: 62G05, 62G08.

Keywords and phrases: geographically weighted regression, nonparametric regression, maximum likelihood estimator, spatial data, truncated spline approach.

function approach can be used to solve the problems of regression curve that cannot be identified in the spatial data and the results of the model find the unbiased estimator of the parameter.

Introduction

The truncated spline nonparametric regression in geographically weighted regression model is the development of nonparametric regression that is calculated on the spatial factor. The method of geographically weighted regression is only able to overcome the problems of the spatial regression analysis whose regression curve is known and linear, but if the regression curve cannot be identified, then we need an approach of the nonparametric regression that will lead the data to find their own form of estimation and curve regression without affected by the subjectivity of the researcher. The approximation ability of a truncated spline which is a highly segmented and continuous absolute polynomial model provides a high flexibility to adapt more effectively to the local characteristics of the data. Spline truncated approach can be a tool to solve the problem of modeling spatial data analysis where the data patterns between the response variable and the predictor variable are unknown or regression curve is not known. Geographic weighting in the spatial data analysis is an important aspect in determining the different parameters at each point of observation location.

The GWR methods continue to be developed by the experts. Nakaya et al. [12] found a new model in the GWR known as geographically weighted Poisson regression model. Then Mei et al. [10] produced a new model that combined two methods: global and GWR regression model, the new method is mixed geographically weighted regression model. GWR models were also developed in the field of spatially related time series, known as spatio-temporal, spatio-temporal data exploration analyzed using GWR and geovisual analyticity. The work was conducted by Demšar et al. [3]. Spatio-temporal study was also developed by Huang et al. [7] in geographically and temporally weighted regression model. Furthermore, Yu [18] worked on spatial panel data, and established the methods of geographically weighted panel regression. Recently, Wrenn and Sam [17] found a model geographically and temporally weighted likelihood regression. The study using semiparametric regression in the GWR model conducted by Holan et al. [8] produced a new method which is semiparametric geographically weighted response curves where the approach function used was bivariate penalized spline regression. Subsequently Ribeiro et al. [15] conducted research in the semiparametric regression field, named semiparametric Poisson geographically weighted regression. Work on GWR and site-specific by using kernel was performed by Paez et al. [13, 14]. This study was about estimation and inference geographically weighted regression models with a specific location with the help of the kernel bandwidths.

Spline approach can be used as a tool to solve the problem of modeling a nonparametric analysis of spatial data. Spline truncated was developed by Budiantara [2]. Then Jiawei et al. [9] conducted work using spline truncated *B*-spline wavelet on the interval (BSWI) and Giannelli et al. [6] used a truncated spline with *B*-spline method. Truncated spline approach has been reviewed by Samsodin and Budiantara [16], Merdekawati and Budiantara [11] and Bintariningrum and Budiantara [1]. The statistical inference was used to reduce the estimator from the models, to investigate the properties of estimator, the shape of the distribution of the test statistic on the model and to find the method for selecting the optimal knots points for the model. Based on the description above, it encourages the researchers to examine the truncated spline nonparametric regression in geographically weighted regression models.

Truncated Spline Nonparametric Regression in Geographically Weighted Regression Models

Nonparametric regression model (Eubank [4]) in general can be presented in the following equation:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, ..., n \tag{1}$$

with $f(x_i)$ as a regression curve that is approached with the truncated spline function (Budiantara [2]) of order m and point knots $K_1, K_2, ..., K_r$ are given by the equation:

$$f(x_i) = \sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h}(x_i - K_h)_+^m, \tag{2}$$

where β_k, β_{m+h} are real constants with $k = 0, 1, 2, ..., m$; $h = 1, 2, ..., r$ and *truncated* function (Budiantara [2]) as:

$$(x_i - K_h)_+^m = \begin{cases} (x_i - K_h)^m, & x_i \geq K_h, \\ 0, & x_i < K_h. \end{cases} \tag{3}$$

When equation (2) is substituted in equation (1), we obtain the equation of truncated spline nonparametric regression (Budiantara [2]) as follows:

$$y_i = \sum_{k=0}^{m} \beta_k x_i^k + \sum_{h=1}^{r} \beta_{m+h}(x_i - K_h)_+^m + \varepsilon_i, \quad i = 1, 2, ..., n \tag{4}$$

with y_i as the response variable and x_i the predictor variable. If truncated spline nonparametric regression has more than one predictor variable, for example given x_p, where $p = 1, 2, ..., l$, then we have

$$y_i = \sum_{p=1}^{l} f_p(x_{pi}) + \varepsilon_i = \sum_{p=1}^{l} \left(\sum_{k=0}^{m} \beta_{pk} x_{pi}^k + \sum_{h=1}^{r} \beta_{p(m+h)}(x_{pi} - K_h)_+^m \right) + \varepsilon_i$$

so that the truncated spline nonparametric regression for more than one predictor variable is:

$$y_i = \sum_{p=1}^{l} \left(\sum_{k=0}^{m} \beta_{pk} x_{pi}^k + \sum_{h=1}^{r} \beta_{p(m+h)}(x_{pi} - K_h)_+^m \right) + \varepsilon_i$$

with truncated spline function of order m at the knot points $K_1, K_2, ..., K_r$.

The functional relationship between the response variable and the predictor variable of each regression coefficient depends on the location of forming a pattern of nonparametric relationships. Therefore, to resolve the problems, the solution that we want to develop is an approach of the truncated spline in the model geographically weighted regression.

Geographically weighted regression model (Fotheringham et al. [5]) of the relationship between the response variable Y and the predictor variables $x_1, x_2, ..., x_l$ at the location i is:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \beta_2(u_i, v_i)x_{2i} + \cdots + \beta_l(u_i, v_i)x_{li} + \varepsilon_i.$$

In this study, the first thing to do is to substitute the truncated spline function with one variable x in the model of geographically weighted regression which provides the following:

$$y_i = \sum_{k=0}^{m} \beta_k(u_i, v_i)x_i^k + \sum_{h=1}^{r} \beta_{m+h}(u_i, v_i)(x_i - K_h)_+^m + \varepsilon_i, \qquad (5)$$

where (u_i, v_i) are coordinates of the geographical location on the ith location, variable response y_i on the ith location, x_i is the predictor variable on the ith location, K_h is the point of knots and error ε_i is in normal distribution, independent with mean zero and variance σ^2, β_k and β_{m+h} are real constants.

Estimation Model

The assumption of truncated spline nonparametric regression in GWR model is normal distributed error with zero mean and variance σ^2. The approach used parameter estimators $\boldsymbol{\beta}(u_i, v_i)$ and $\sigma^2(u_i, v_i)$ with maximum likelihood weighting estimator.

On the jth location of observation, y_j is normal distributed with mean and variance as follows:

$$y_j \sim N\left(\sum_{k=0}^{m} \beta_k(u_j, v_j)x_j^k + \sum_{h=1}^{r} \beta_{m+h}(u_j, v_j)(x_j - K_h)_+^m, \sigma^2(u_j, v_j)\right).$$

The first step is to establish the likelihood function as follows:

$$f(y_1 | \boldsymbol{\beta}(u_1, v_1), \sigma^2(u_1, v_1)) \cdots f(y_n | \boldsymbol{\beta}(u_n, v_n), \sigma^2(u_n, v_n))$$

$$= \prod_{i=1}^{n} f(y_j | \boldsymbol{\beta}(u_j, v_j), \sigma^2(u_j, v_j))$$

described in the form

$$\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi\sigma^2(u_j, v_j)}}\exp\left(-\frac{1}{2\sigma^2(u_j, v_j)}\right.\right.$$

$$\left.\left.\times\left(Y_j - \left(\sum_{k=0}^{m}\beta_k(u_j, v_j)x_i^k + \sum_{h=1}^{r}\beta_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2\right)\right]$$

shortened as

$$(2\pi)^{-\frac{n}{2}}(\sigma^2(u_j, v_j))^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2(u_j, v_j)}G\right) \tag{6}$$

with

$$G = \sum\nolimits_{i=1}^{n}\left(y_j - \left(\sum\nolimits_{k=0}^{m}\beta_k(u_j, v_j)x_j^k\right.\right.$$

$$\left.\left.+\sum\nolimits_{h=1}^{r}\beta_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2.$$

After obtaining the joint density function of $Y_1, Y_2, ..., Y_n$, we estimate the model, given the geographical weighting on the jth location represented by $w_{i(j)}$. So the likelihood function for the jth location can be found as follows:

$$L(\boldsymbol{\beta}(u_j, v_j), \sigma^2(u_j, v_j)|Y)$$

$$= (2\pi)^{-\frac{n}{2}}(\sigma^2(u_j, v_j))^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2(u_j, v_j)}\sum_{i=1}^{n}w_{i(j)}\right.$$

$$\left.\times\left(y_j - \left(\sum_{k=0}^{m}\beta_k(u_j, v_j)x_j^k + \sum_{h=1}^{r}\beta_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2\right].$$

$w_{i(j)} =$ the value of weighting on the ith location to the jth location.

The form of the weighted likelihood function then performed the operation to facilitate the natural logarithm mathematical operations in order to obtain the parameter estimator as:

$$\ln L(\boldsymbol{\beta}(u_j, v_j), \sigma^2(u_j, v_j)|Y)$$

$$= \ln((2\pi)^{-\frac{n}{2}}(\sigma^2(u_j, v_j))^{-\frac{n}{2}})\ln\left[\exp\left(-\frac{1}{2\sigma^2(u_j, v_j)}\sum_{i=1}^{n} w_{ik}\right.\right.$$

$$\left.\left.\times\left(y_j - \left(\sum_{k=0}^{m}\beta_k(u_j, v_j)x_j^k + \sum_{h=1}^{r}\beta_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2\right)\right]$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2(u_j, v_j)) - \frac{1}{2\sigma^2(u_j, v_j)}H \tag{7}$$

and

$$H = \sum_{i=1}^{n} w_{ik}\left(\left(y_j - \left(\sum_{k=0}^{m}\beta_k(u_j, v_j)x_j^k + \sum_{h=1}^{r}\beta_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2\right)$$

$$= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(u_j, v_j))^T \boldsymbol{W}(u_j, v_j)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(u_j, v_j))$$

with

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^m & (x_1 - K_1)_+^m & (x_1 - K_2)_+^m & \cdots & (x_1 - K_r)_+^m \\ 1 & x_2^1 & x_2^2 & \cdots & x_2^m & (x_2 - K_1)_+^m & (x_2 - K_2)_+^m & \cdots & (x_2 - K_r)_+^m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & x_n^2 & \cdots & x_n^m & (x_n - K_1)_+^m & (x_n - K_2)_+^m & \cdots & (x_n - K_r)_+^m \end{bmatrix},$$

$$\boldsymbol{\beta}(u_j, v_j) = \begin{bmatrix} \beta_0(u_j, v_j) \\ \beta_1(u_j, v_j) \\ \beta_2(u_j, v_j) \\ \vdots \\ \beta_m(u_j, v_j) \\ \beta_{m+1}(u_j, v_j) \\ \beta_{m+2}(u_j, v_j) \\ \vdots \\ \beta_{m+r}(u_j, v_j) \end{bmatrix},$$

$$\boldsymbol{W}(u_j, v_j) = \mathrm{diag}(w_1(u_j, v_j), w_2(u_j, v_j), \ldots, w_n(u_j, v_j)).$$

Parameter estimators $\boldsymbol{\beta}(u_j, v_j)$ and $\sigma^2(u_j, v_j)$ are obtained by maximizing $\ln L$ shape in equation (7) and the next stage of the process of differential against each parameter $\boldsymbol{\beta}(u_j, v_j)$:

$$\frac{\partial(\ln L(\boldsymbol{\beta}(u_j, v_j), \sigma^2(u_j, v_j)|Y))}{\partial(\boldsymbol{\beta}(u_j, v_j))^T} = \frac{\partial(H)}{\partial(\boldsymbol{\beta}(u_j, v_j))^T} \tag{8}$$

so that

$$\frac{\partial(H)}{\partial(\boldsymbol{\beta}(u_j, v_j))^T} = \frac{\partial((\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(u_j, v_j))^T \boldsymbol{W}(u_j, v_j)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}(u_j, v_j)))}{\partial(\boldsymbol{\beta}(u_j, v_j))^T}$$

$$= \frac{\begin{array}{c}\partial(\boldsymbol{Y}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y} - 2(\boldsymbol{\beta}(u_j, v_j))^T \boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y} \\ + (\boldsymbol{\beta}(u_j, v_j))^T \boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X}\boldsymbol{\beta}(u_j, v_j))\end{array}}{\partial(\boldsymbol{\beta}(u_j, v_j))^T}$$

$$= 2\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X}\boldsymbol{\beta}(u_j, v_j),$$

thus obtained $\boldsymbol{\beta}(u_j, v_j)$ is

$$\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X}\boldsymbol{\beta}(u_j, v_j) = \boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y},$$

$$\hat{\boldsymbol{\beta}}(u_j, v_j) = (\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}.$$

Lemma 1. *If* $\boldsymbol{B}(u_j, v_j)$ *is a matrix of parameter* $\beta(u_j, v_j)$ *of the spline nonparametric regression model with geographic weighting following equation* (5), *then the parameter estimator* $\hat{\boldsymbol{B}}(u_j, v_j)$ *on the jth location is* $\hat{\boldsymbol{B}}(u_j, v_j) = (\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}$.

Proof. It is known that $\hat{\boldsymbol{\beta}}(u_j, v_j) = (\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}$ is the estimator $\boldsymbol{\beta}(u_j, v_j)$ on the *i*th location. It can be seen that $\hat{\boldsymbol{B}}(u_i, v_i)$ is a matrix of parameter estimator on the *j*th location, that is:

$$\hat{\boldsymbol{B}}(u_j, v_j) = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1(u_j, v_j) \\ \hat{\boldsymbol{\beta}}_2(u_j, v_j) \\ \vdots \\ \hat{\boldsymbol{\beta}}_p(u_j, v_j) \end{bmatrix} = \begin{bmatrix} (\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}_1 \\ (\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}_2 \\ \vdots \\ (\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}_p \end{bmatrix}. \quad \square$$

Lemma 2. *If* $\hat{\sigma}^2(u_j, v_j)$ *is a parameter estimator* $\sigma^2(u_j, v_j)$ *of the spline nonparametric regression model with geographic weighting following equation* (5), *then* $\hat{\sigma}^2(u_j, v_j)$ *is a parameter estimator on the jth location given by*

$$\hat{\sigma}^2(u_j, v_j)$$
$$= \frac{\sum_{i=1}^{n} w_{i(j)}\left(\left(y_j - \left(\sum_{k=0}^{m}\hat{\beta}_k(u_j, v_j)x_j^k + \sum_{h=1}^{r}\hat{\beta}_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2\right)}{n}.$$

Proof. Known parameter estimator $\sigma^2(u_j, v_j)$ is obtained from the differential equation (7) as follows:

$$\frac{\partial(\ln L(\boldsymbol{\beta}(u_j, v_j), \sigma^2(u_j, v_j)|Y))}{\partial(\sigma^2(u_j, v_j))} = \frac{\partial\left(-\frac{n}{2}\ln(\sigma^2(u_j, v_j)) - \frac{1}{2\sigma^2(u_j, v_j)}H\right)}{\partial(\sigma^2(u_j, v_j))}$$

$$-\frac{n}{2\sigma^2(u_j, v_j)} + \frac{1}{2(\sigma^2(u_j, v_j))^2} H = 0.$$

Parameter estimator is given by

$$\hat{\sigma}^2(u_j, v_j) = \frac{H}{n}.$$

The estimator is calculated based on the characteristic of the locality of spatial data models is expressed as:

$$\hat{\sigma}^2(u_j, v_j)$$
$$= \frac{\sum_{i=1}^{n} w_{i(j)}\left(\left(y_j - \left(\sum_{k=0}^{m}\hat{\beta}_k(u_j, v_j)x_j^k + \sum_{h=1}^{r}\hat{\beta}_{h+m}(u_j, v_j)(x_j - K_h)_+^m\right)\right)^2\right)}{n}.$$

$\square$

Lemma 3. *If* $\hat{\boldsymbol{\beta}}(u_j, v_j)$ *is a parameter estimator of spline nonparametric regression model with a geographic weighting following equation* (5), *then* $\hat{\boldsymbol{\beta}}(u_j, v_j)$ *is an unbiased estimator for* $\boldsymbol{\beta}(u_j, v_j)$.

Proof. After obtaining a parameter estimator $\boldsymbol{\beta}(u_j, v_j)$, the unbiased characteristic of the estimator $\boldsymbol{\beta}(u_j, v_j)$ can be described as:

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}(u_j, v_j)) &= E((\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_j, v_j)\boldsymbol{Y}) \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_j, v_j)E(\boldsymbol{Y}) \\
&= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_j, v_j)\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_j, v_j)\boldsymbol{X}\boldsymbol{\beta}(u_j, v_j) \\
&= \boldsymbol{\beta}(u_j, v_j).
\end{aligned}$$

The above result proves that $\hat{\boldsymbol{\beta}}(u_j, v_j)$ is an unbiased estimator of $\boldsymbol{\beta}(u_j, v_j)$. $\square$

In the model estimator for each observation at the ith location can be obtained in the following way:

$$\hat{\boldsymbol{Y}}_i = \boldsymbol{X}_i^T \hat{\boldsymbol{\beta}}(u_i, v_i) = \boldsymbol{X}_i^T ((\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}(u_i, v_i) \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}(u_i, v_i) \boldsymbol{Y}). \tag{9}$$

Equation (9) can be re-written in the form:

$$\hat{\boldsymbol{Y}} = \boldsymbol{SY},$$

where $\boldsymbol{S}$ is an $n \times n$ matrix given below and $\boldsymbol{I}$ is the identity matrix of order n. Note that

$$\boldsymbol{S}_{n \times n} = \begin{bmatrix} \boldsymbol{X}_1^T (\boldsymbol{X}^T \boldsymbol{W}(u_1, v_1) \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_1, v_1) \\ \boldsymbol{X}_2^T (\boldsymbol{X}^T \boldsymbol{W}(u_2, v_2) \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{X}_n^T (\boldsymbol{X}^T \boldsymbol{W}(u_n, v_n) \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_n, v_n) \end{bmatrix}.$$

The error vector estimator is

$$\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}. \tag{10}$$

To find an unbiased estimator for σ^2, it is necessary to look for the estimator value of sum square error (SSE) of the model, which is obtained by squaring equation (10), as follows:

$$SSE = \boldsymbol{e}^T \boldsymbol{e} = ((\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y})^T ((\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}) = \boldsymbol{Y}^{\boldsymbol{T}} (\boldsymbol{I} - \boldsymbol{S})^T (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y} \tag{11}$$

with

$$\begin{aligned} E(\boldsymbol{e}) &= E(\boldsymbol{Y} - \hat{\boldsymbol{Y}}) \\ &= E(\boldsymbol{Y}) - E(\hat{\boldsymbol{Y}}) \\ &= \boldsymbol{X}^T \boldsymbol{\beta}(u_i, v_i) - \boldsymbol{X}^T \hat{\boldsymbol{\beta}}(u_i, v_i) = 0 \end{aligned}$$

and *error* variance as:

$$Var(\boldsymbol{e}) = E[(\boldsymbol{e} - E(\boldsymbol{e}))(\boldsymbol{e} - E(\boldsymbol{e}))^T] = \sigma^2 \boldsymbol{I}. \tag{12}$$

Based on equation (12), equation (11) can be described as follows:

$$SSE = \boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{e} - E(\boldsymbol{e}))^T(\boldsymbol{e} - E(\boldsymbol{e})) = \boldsymbol{e}^T(\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{e}.$$

Because $(\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})$ is a symmetric matrix and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, the value estimator SSE is:

$$E(SSE) = E(\boldsymbol{e}^T(\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{e}) = tr((\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S}))E(\boldsymbol{e}\boldsymbol{e}^T)$$

$$= tr((\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S}))\sigma^2\boldsymbol{I}$$

and

$$E\left(\frac{SSE}{tr((\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S}))}\right) = \sigma^2.$$

Unbiased estimator is given by

$$\hat{\sigma}^2 = \frac{SSE}{tr((\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S}))}.$$

Conclusion

In this paper, we described that the nonparametric regression with truncated spline approach in the GWR models can be a solution for relationship predictor variables and the response variables that have a spatial aspect explaining the relationship between the two which is nonlinear identified as nonparametric. Results of the model parameter estimation are proved unbiased.

The shape and distribution of the test statistic models in truncated spline GWR model and application model in empirical data can be considered for future study.

References

[1] M. F. Bintariningrum and I. N. Budiantara, Pemodelan regresi nonparametrik spline truncated dan aplikasinya pada Angka Kelahiran Kasar di Surabaya, Journal Sains 3(1) (2014).

[2] I. N. Budiantara, Model keluarga spline polinomial truncated dalam regresi semiparametrik, Berkala Ilmiah MIPA 15 (2005).

[3] U. Demšar, A. Stewart Fotheringham and M. Charlton, Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics, Information Visualization 7 (2008), 181-197.

[4] R. L. Eubank, Nonparametric Regression and Spline Smoothing, Marcel Dekker, New York, 1999.

[5] A. S. Fotheringham, C. Brundson and M. Charlton, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, John Wiley & Sons Ltd, England, 2002.

[6] C. Giannelli, B. Juttler and H. Speleers, THB-spline: the truncated basis for hierarchical spline, Comput. Aided Geom. Design 29 (2012), 485-498.

[7] B. Huang, B. Wu and M. Barry, Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices, International Journal of Geographical Information Science 24 (2010), 383-401.

[8] S. Holan, S. Wang, A. Arab, E. John Sadler and Kenneth Stone, Semiparametric geographically weighted response curves with application to site-specific agriculture, Journal of Agricultural, Biological, and Environmental Statistics 13(4) (2008), 424-439.

[9] X. Jiawei, H. Zhengjia and C. Xuefeng, The construction of wavelet-based truncated conical shell element using *B*-spline wavelet on the interval, Acta Mechanica Solida Sinica 19 (2006). DOI: 10.1007/s10338-006-0641-5.

[10] C.-L. Mei, N. Wang and W.-X. Zhang, Testing the importance of the explanatory variables in a mixed geographically weighted regression model, Environment and Planning A 38(3) (2006), 587-598.

[11] I. P. Merdekawati and I. N. Budiantara, Pemodelan regresi spline truncated multivariabel pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah, Jurnal Sains dan Seni Pomits 2(1) (2013).

[12] T. Nakaya, A. S. Fotheringham, C. Brunsdon and M. Charltron, Geographically weighted Poisson regression for disease association mapping, Stat. Med. 24(17) (2005), 2695-2717.

[13] A. Paez, T. Uchida and K. Miyamoto, A general framework for estimation and inference of geographically weighted regression models: location-specific kernel bandwidths and a test for locational heterogeneity, Environment Planning A 34(4) (2002a), 733-754.

[14] A. Paez, T. Uchida and K. Miyamoto, A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests, Environment and Planning A 34(5) (2002b), 883-904.

[15] M. C. Ribeiro, A. J. Sousa and M. J. Pereira, A coregionalization model to assist the selection process of local and global variables in semi-parametric geographically weighted Poisson regression, Procedia Environmental Sciences 26 (2015), 53-56.

[16] M. Samsodin and I. N. Budiantara, Truncated polynomial spline regression multiresponse to modelling indicators of poverty in East Java, 2012.

[17] D. H. Wrenn and A. G. Sam, Geographically and temporally weighted likelihood regression: exploring the spatiotemporal determinants of land use change, Regional Science and Urban Economics 44 (2014), 60-74.

[18] D. Yu, Exploring spatiotemporally varying regressed relationships: the geographically weighted panel regression analysis, Proceedings of the Joint International Conference on Theory, Data Handling and Modeling in Geospatial Information Science, 2010, pp. 134-139.

[19] W. Wibowo, S. Haryatmi and I. N. Budiantara, On multiresponse semiparametric regression model, J. Math. Stat. 8(4) (2012), 489-499.

[20] I. N. Budiantara, V. Ratnasari, M. Ratna and I. Zain, The combination of spline and kernel estimator for nonparametric regression and its properties, Journal of Applied Sciences 9(122) (2015), 6083-6094.

[21] R. Pane, I. N. Budiantara, I. Zain and B. W. O. Otok, Parametric and nonparametric estimators in Fourier series semiparametric regression and their characteristics, Journal of Applied Mathematical Sciences 8(102) (2014), 5053-5064.