

## Research Article

# A New Method of Hypothesis Test for Truncated Spline Nonparametric Regression Influenced by Spatial Heterogeneity and Application

Sifriyani <sup>1</sup>, I. N. Budiantara <sup>2</sup>, S. H. Kartiko,<sup>3</sup> and Gunardi<sup>3</sup>

<sup>1</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia

<sup>2</sup>Department of Statistics, Faculty of Mathematics, Computing and Data Sciences, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

<sup>3</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Yogyakarta, Indonesia

Correspondence should be addressed to I. N. Budiantara; nyomanbudiantara65@gmail.com

Received 19 April 2018; Revised 17 July 2018; Accepted 5 August 2018; Published 12 September 2018

Academic Editor: Giovanni P. Galdi

Copyright © 2018 Sifriyani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study developed a new method of hypothesis testing of model conformity between truncated spline nonparametric regression influenced by spatial heterogeneity and truncated spline nonparametric regression. This hypothesis test aims to determine the most appropriate model used in the analysis of spatial data. The test statistic for model conformity hypothesis testing was constructed based on the likelihood ratio of the parameter set under  $H_0$  whose components consisted of parameters that were not influenced by the geographical factor and the set under the population parameter whose components consisted of parameters influenced by the geographical factor. We have proven the distribution of test statistics  $V$  and verified that each of the numerators and denominators in the statistic test  $V$  followed a distribution of  $\chi^2$ . Since there was a symmetric and idempotent matrix  $S$ , it could be proved that  $\bar{Y}^T S \bar{Y} / \sigma^2 \sim \chi^2_{(n-lm-1)}$ . Matrix  $D(u_i, v_i)$  was positive semidefinite and contained weighting matrix  $W(u_i, v_i)$  which had different values in every location; therefore matrix  $D(u_i, v_i)$  was not idempotent. If  $\bar{Y}^T D(u_i, v_i) \bar{Y} \geq 0$  and  $D(u_i, v_i)$  was not idempotent and also  $\bar{Y}$  was a  $N(0, I)$  distributed random vector, then there were constants  $k$  and  $r$ ; hence  $\bar{Y}^T D(u_i, v_i) \bar{Y} \sim k\chi_r^2$ ; therefore it was concluded that test statistic  $V$  followed an F distribution. The modeling is implemented to find factors that influence the unemployment rate in 38 areas in Java in Indonesia.

## 1. Introduction

This study examines theoretically the multivariate nonparametric regression influenced by spatial heterogeneity with truncated spline approach. The model is the development of truncated spline nonparametric regression that takes into account geographic or spatial factors. Truncated spline is a function constructed on the basis of polynomial components and truncated components; i.e., polynomial pieces that have knot points, which can overcome the pattern of changes in data behavior. Truncated spline approach is used as a solution to solve the problem of spatial data analysis modeling; that is, the relationship between the response variable and the predictor variable does not follow a certain pattern and there is a changing pattern in certain subintervals. The response

variable in the model contains the predictor variables whose respective regression coefficients depend on the location where the data is observed, due to differences in environmental and geographic characteristics between the observation sites; therefore each observation has different variations (spatial heterogeneity). Spatial is one type of dependent data, where data at a location is influenced by the measurement of data at another location (spatial dependency).

This study determines the model conformity hypothesis test between multivariable nonparametric regression that is influenced by spatial heterogeneity with truncated spline approach and multivariable nonparametric regression in general. This hypothesis test aims to determine the model that is most suitable for spatial data analysis. The test statistic was derived using the maximum likelihood ratio test (MLRT)

method. The first step in this study was formulating the hypothesis to be tested and then defining the set of parameters under  $H_0$  whose components consist of parameters that are not influenced by geographical factors and the set under population parameters whose components consist of parameters influenced by geographical factors. Likelihood ratio  $\Lambda$  was constructed based on the maximum ratio of the likelihood function under  $H_0$  as the numerator and set under the population as a denominator. Based on the likelihood ratio test statistic  $V$  was obtained. Furthermore, the distribution of test statistic  $V$  was determined. To prove the distribution of test statistic  $V$ , we first proved that each numerator and denominator are chi square distributed.

The purpose of this study is to obtain a new method for the determination of hypothesis test of model conformity between multivariate nonparametric truncated spline regression influenced by spatial heterogeneity versus multivariate nonparametric truncated spline regression in general. This hypothesis test aims to determine what model is most suitable for spatial data analysis.

## 2. Truncated Spline Nonparametric Regression Influenced by Heterogeneity Spatial

Truncated spline nonparametric regression influenced by spatial heterogeneity is the development of nonparametric regression for spatial data with parameter estimators local to each location of observation. Truncated spline approach is used to solve spatial analysis problems whose regression curve is unknown [1]. The assumption of the regression model used is the normal distributed error with mean zero and variance  $\sigma^2(u_i, v_i)$  at each location  $(u_i, v_i)$ . Location coordinates  $(u_i, v_i)$  are an important factor in determining the weights used to estimate the parameters of the model. Given data  $(x_{1i}, x_{2i}, \dots, x_{li}, y_i)$  and relationship between  $(x_{1i}, x_{2i}, \dots, x_{li})$  and  $y_i$ , it is assumed to follow multivariate nonparametric regression model as follows:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{li}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

$y_i$  is response variable and  $f(x_{1i}, x_{2i}, \dots, x_{li})$  is unknown regression curve and assumed to be additive. If  $f(x_{1i}, x_{2i}, \dots, x_{li})$  is approached with a truncated spline function. Mathematically, the relation between response variable  $y_i$  and the predictor variable  $(x_{1i}, x_{2i}, \dots, x_{li})$  at  $i$ -th location for the multivariate nonparametric truncated spline regression model can be expressed as follows [2]:

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pi}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pi} - K_{ph})_+^m + \varepsilon_i \quad (2)$$

with truncated function:

$$(x_{pi} - K_{ph})_+^m = \begin{cases} (x_{pi} - K_{ph}), & x_{pi} \geq K_{ph} \\ 0, & x_{pi} < K_{ph} \end{cases} \quad (3)$$

Equation (2) is a multivariate nonparametric truncated spline regression model of degree  $m$  with  $n$  area. The components in (2) are described as follows:

$y_i$  is a response variable at  $i$ -th location, where  $i = 1, 2, \dots, n$ .

$x_{pi}$  is a  $p$ -th predictor variable at  $i$ -th location with  $p = 1, 2, \dots, l$ .

$K_{ph}$  is an  $h$ -th knot point in  $p$ -th predictor variable component with  $h = 1, 2, \dots, r$ .

$\beta_{pk}(u_i, v_i)$  is a polynomial component parameter of a multivariate nonparametric truncated spline regression.  $\beta_{pk}(u_i, v_i)$  is a  $k$ -th parameter from  $p$ -th predictor variable at  $i$ -th location.  $\delta_{p,m+h}(u_i, v_i)$  is a truncated component from multivariate nonparametric truncated spline regression.  $\delta_{p,m+h}(u_i, v_i)$  is an  $l + h$ -th parameter in  $h$ -th knot point and  $p$ -th predictor variable at  $i$ -th location.

Multivariate nonparametric truncated spline regression in (2) is described as follows:

$$y_i = (\beta_0(u_i, v_i) + \beta_{11}(u_i, v_i) x_{1i} + \beta_{12}(u_i, v_i) x_{1i}^2 + \dots + \beta_{1m}(u_i, v_i) x_{1i}^m + \beta_{21}(u_i, v_i) x_{2i} + \beta_{22}(u_i, v_i) x_{2i}^2 + \beta_{23}(u_i, v_i) x_{2i}^3 + \dots + \beta_{2m}(u_i, v_i) x_{2i}^m + \dots + \beta_{l1}(u_i, v_i) x_{li} + \beta_{l2}(u_i, v_i) x_{li}^2 + \beta_{l3}(u_i, v_i) x_{li}^3 + \dots + \beta_{lm}(u_i, v_i) x_{li}^m) + \delta_{1,m+1}(u_i, v_i) (x_{1i} - K_{11})_+^m + \dots + \delta_{1,m+r}(u_i, v_i) (x_{1i} - K_{1r})_+^m + \delta_{2,m+1}(u_i, v_i) (x_{2i} - K_{21})_+^m + \dots + \delta_{2,m+r}(u_i, v_i) (x_{2i} - K_{2r})_+^m + \dots + \delta_{l,m+1}(u_i, v_i) (x_{li} - K_{l1})_+^m + \dots + \delta_{l,m+r}(u_i, v_i) (x_{li} - K_{lr})_+^m + \varepsilon_i \quad (4)$$

Equation (4) can also be expressed as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0(u_1, v_1) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_1, v_1) x_{p1}^k \\ \beta_0(u_2, v_2) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_2, v_2) x_{p2}^k \\ \vdots \\ \beta_0(u_n, v_n) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_n, v_n) x_{pn}^k \end{bmatrix} + \begin{bmatrix} \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_1, v_1) (x_{p1} - K_{ph})_+^m \\ \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_2, v_2) (x_{p2} - K_{ph})_+^m \\ \vdots \\ \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_n, v_n) (x_{pn} - K_{ph})_+^m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (5)$$

Thus (5) can be expressed by

$$\tilde{Y} = \tilde{f} + \tilde{\varepsilon} = \mathbf{X}\tilde{\beta}(u_i, v_i) + \mathbf{P}\tilde{\delta}(u_i, v_i) + \tilde{\varepsilon} \quad (6)$$

whose vector  $\tilde{f}$  contains of truncated spline function with geographical weighting sized  $n \times 1$ ; response variable and error, respectively, are given by vectors as follows:

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

$$\tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

(7)

Vectors  $\tilde{Y}$  and  $\tilde{\varepsilon}$  are, respectively, sized  $n \times 1$ .

Meanwhile matrices  $\mathbf{X}$  and  $\mathbf{P}$  are, respectively, given by

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & x_{11}^3 & \cdots & x_{11}^m & x_{21} & x_{21}^2 & \cdots & x_{21}^m & \cdots & x_{l1} & x_{l1}^2 & \cdots & x_{l1}^m \\ 1 & x_{12} & x_{12}^2 & x_{12}^3 & \cdots & x_{12}^m & x_{22} & x_{22}^2 & \cdots & x_{22}^m & \cdots & x_{l2} & x_{l2}^2 & \cdots & x_{l2}^m \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{1n}^2 & x_{1n}^3 & \cdots & x_{1n}^m & x_{2n} & x_{2n}^2 & \cdots & x_{2n}^m & \cdots & x_{ln} & x_{ln}^2 & \cdots & x_{ln}^m \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} (x_{11} - K_{11})_+^m & \cdots & (x_{11} - K_{1r})_+^m & (x_{21} - K_{21})_+^m & \cdots & (x_{l1} - K_{l1})_+^m & \cdots & (x_{l1} - K_{lr})_+^m \\ (x_{12} - K_{11})_+^m & \cdots & (x_{12} - K_{1r})_+^m & (x_{22} - K_{21})_+^m & \cdots & (x_{l2} - K_{l1})_+^m & \cdots & (x_{l2} - K_{lr})_+^m \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ (x_{1n} - K_{11})_+^m & \cdots & (x_{1n} - K_{1r})_+^m & (x_{2n} - K_{21})_+^m & \cdots & (x_{ln} - K_{l1})_+^m & \cdots & (x_{ln} - K_{lr})_+^m \end{bmatrix} \quad (8)$$

Vectors  $\tilde{\beta}(u_i, v_i)$  and  $\tilde{\delta}(u_i, v_i)$  are, respectively, given by

$$\tilde{\beta}(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_{11}(u_i, v_i) \\ \beta_{12}(u_i, v_i) \\ \beta_{13}(u_i, v_i) \\ \vdots \\ \beta_{1m}(u_i, v_i) \\ \beta_{21}(u_i, v_i) \\ \beta_{22}(u_i, v_i) \\ \vdots \\ \beta_{2m}(u_i, v_i) \\ \vdots \\ \beta_{l1}(u_i, v_i) \\ \beta_{l2}(u_i, v_i) \\ \vdots \\ \beta_{lm}(u_i, v_i) \end{bmatrix},$$

$$\tilde{\delta}(u_i, v_i) = \begin{bmatrix} \delta_{1,m+1}(u_i, v_i) \\ \delta_{1,m+2}(u_i, v_i) \\ \vdots \\ \delta_{1,m+r}(u_i, v_i) \\ \delta_{2,m+1}(u_i, v_i) \\ \delta_{2,m+2}(u_i, v_i) \\ \vdots \\ \delta_{2,m+r}(u_i, v_i) \\ \vdots \\ \delta_{l,m+1}(u_i, v_i) \\ \delta_{l,m+2}(u_i, v_i) \\ \vdots \\ \delta_{l,m+r}(u_i, v_i) \end{bmatrix}$$

(9)

Matrix  $\mathbf{X}$  is sized  $n \times (1 + lm)$ ; matrix  $\mathbf{P}$  contains predictor variable of truncated function sized  $n \times lr$ . Vector  $\tilde{\beta}(u_i, v_i)$  is a parameter vector sized  $(1 + lm) \times 1$ . Vector  $\tilde{\delta}(u_i, v_i)$  is a parameter vector containing truncated function sized  $lr \times 1$ .

The estimator forms  $\widehat{\beta}(u_i, v_i)$ ,  $\widehat{\delta}(u_i, v_i)$ , and  $\widehat{f}$  are complete in Theorem 1 and Corollary 2 [2].

**Theorem 1.** *If the regression model (2) with an error  $\varepsilon_i$  normally distributed with zero mean and variance  $\sigma^2(u_i, v_i)$  was given Maximum Likelihood Estimator (MLE), it is used to obtain estimator  $\widehat{\beta}(u_i, v_i)$  and  $\widehat{\delta}(u_i, v_i)$  as follows.*

$$\begin{aligned}\widehat{\beta}(u_i, v_i) &= \mathbf{A}(\mathbf{K}) \widetilde{\mathbf{Y}} \\ \widehat{\delta}(u_i, v_i) &= \mathbf{B}(\mathbf{K}) \widetilde{\mathbf{Y}}\end{aligned}\quad (10)$$

where

$$\begin{aligned}\mathbf{A}(\mathbf{K}) &= \mathbf{S}(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \left[ \mathbf{X}^T \right. \\ &\quad \left. - \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{P}(\mathbf{P}^T \mathbf{W}(u_i, v_i) \mathbf{P})^{-1} \mathbf{P}^T \right] \mathbf{W}(u_i, v_i) \\ \mathbf{B}(\mathbf{K}) &= \mathbf{R}(\mathbf{P}^T \mathbf{W}(u_i, v_i) \mathbf{P})^{-1} \left[ \mathbf{P}^T \right. \\ &\quad \left. - \mathbf{P}^T \mathbf{W}(u_i, v_i) \mathbf{X}(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \right] \mathbf{W}(u_i, v_i)\end{aligned}\quad (11)$$

**Corollary 2.** *If  $\widehat{\beta}(u_i, v_i)$  and  $\widehat{\delta}(u_i, v_i)$  are given by Theorem 1, then the estimator for the regression curve  $\widehat{f}$  is given by*

$$\widehat{f} = \mathbf{X} \widehat{\beta}(u_i, v_i) + \mathbf{P} \widehat{\delta}(u_i, v_i) = \mathbf{C}(\mathbf{K}) \widetilde{\mathbf{Y}}, \quad (12)$$

where

$$\mathbf{C}(\mathbf{K}) = \mathbf{X} \mathbf{A}(\mathbf{K}) + \mathbf{P} \mathbf{B}(\mathbf{K}) \quad (13)$$

Estimator of regression curve  $\widehat{f}$  contains the polynomial components represented by matrix  $\mathbf{X}$  and truncated components represented by matrix  $\mathbf{P}$  [3]. If the matrix  $\mathbf{P} = \mathbf{0}$ , then the estimator multivariable of nonparametric regression curve in the Geographically Weighted Regression (GWR) models with truncated spline approach,  $\widehat{f}$ , will change to estimator polynomial parametric regression curve in the GWR model. Furthermore, if  $\mathbf{P} = \mathbf{0}$  and matrix  $\mathbf{X}$  contains a linear function, the estimator of the multivariable spline nonparametric regression curve in the GWR model,  $\widehat{f}$ , will change to estimator of linear parametric regression curves in the GWR model or multiple linear regression in the GWR model developed by many researchers such as Brusdon and Fotheringham [4], Fotheringham, Brunson, and Charlton (2003), Demsar, Fotheringham, and Charlton [5], Yan Li, Yan Jiao, and Joan A. Browder [6], Shan-shan Wu, Hao Yang, Fei Guo, and Rui -Ming Han [7], and Benassi and Naccarato [8].

This study continued the previous research [2]; in this study the test statistics that will be used in the truncated spline nonparametric regression influenced by spatial heterogeneity modeling will be found; further research continued the distribution of test statistics and rejection areas.

### 3. Method

The hypothesis test for model conformity between multivariate nonparametric truncated spline regression influenced by spatial heterogeneity with nonparametric truncated spline regression is derived.

*Step 1.* Formulating hypothetical model:

$$H_0: \beta_{pk}(u_i, v_i) = \beta_{pk} \text{ and } \delta_{p,m+h}(u_i, v_i) = \delta_{p,m+h},$$

$$p = 1, 2, \dots, l; k = 1, 2, \dots, m; h = 1, 2, \dots, r; \\ i = 1, 2, \dots, n$$

$$H_1: \text{at least, there is one of } \beta_{pk}(u_i, v_i) \neq \beta_{pk} \text{ or } \\ \delta_{p,m+h}(u_i, v_i) \neq \delta_{p,m+h}$$

$$p = 1, 2, \dots, l; k = 1, 2, \dots, m; h = 1, 2, \dots, r; \\ i = 1, 2, \dots, n.$$

*Step 2.* Defining the set of parameters under population  $\Omega$ .

*Step 3.* Determining estimators  $\widehat{\eta}_\Omega(u_i, v_i)$  and  $\widehat{\sigma}_\Omega^2(u_i, v_i)$  which are parameters in the space under population ( $\Omega$ ).

*Step 4.* Obtaining maximum likelihood function under population ( $\Omega$ ).

*Step 5.* Defining parameter space under  $H_0$ , i.e.,  $\omega$ .

*Step 6.* Determining estimators  $\widetilde{\eta}_\omega$  and  $\sigma_\omega^2$  which are parameters under  $H_0$ .

*Step 7.* Obtaining maximum likelihood function under space  $H_0$ .

*Step 8.* Obtaining likelihood ratio  $\Lambda$ .

*Step 9.* Obtaining test statistic  $V$  from model conformity testing.

*Step 10.* Specifying the distribution of numerator  $\tau$  from test statistic  $V$ .

*Step 11.* Specifying the distribution of denominator  $\tau^*$  from test statistic  $V$ .

*Step 12.* Specifying the distribution of test statistic  $V$ .

*Step 13.* Deciding the rejection area of  $H_0$  and writing the conclusion.

## 4. Parameter Estimation under Space $H_0$ and Space Population in the Model

A hypothesis testing of model conformity for nonparametric spline regression with spatial heterogeneity was designed by using hypothesis formulation:

$$H_0: \beta_{pk}(u_i, v_i) = \beta_{pk} \text{ and } \delta_{p,m+h}(u_i, v_i) = \delta_{p,m+h},$$

$$H_1: \text{at least, there is one of } \beta_{pk}(u_i, v_i) \neq \beta_{pk} \text{ or } \\ \delta_{p,m+h}(u_i, v_i) \neq \delta_{p,m+h}.$$

This hypothesis test was derived using maximum likelihood ratio test method by defining the parameter spaces under  $H_0(\omega)$  and under population ( $\Omega$ ). The parameter space under  $H_0(\omega)$  is given by

$$\omega = \{ \beta_{11}, \beta_{21}, \dots, \beta_{lm}, \delta_{1,m+1}, \delta_{1,m+2}, \dots, \delta_{1,m+r}, \dots, \delta_{l,m+1}, \\ \dots, \delta_{l,m+r}, \sigma_\omega^2 \} = \{ \widetilde{\beta}, \widetilde{\delta}, \sigma_\omega^2 \} = \{ \widetilde{\eta}_\omega, \sigma_\omega^2 \} \quad (14)$$

where  $\widetilde{\eta}_\omega = \begin{bmatrix} \widetilde{\beta} \\ \widetilde{\delta} \end{bmatrix}$ .

$$\begin{aligned} \tilde{\beta} &= [\beta_{11} \ \beta_{12} \ \dots \ \beta_{1m} \ \beta_{21} \ \dots \ \beta_{2m} \ \dots \ \beta_{lm}]^T \\ \tilde{\delta} &= [\delta_{1,m+1} \ \delta_{1,m+2} \ \dots \ \delta_{1,m+r} \ \delta_{2,m+1} \ \delta_{2,m+2} \ \dots \ \delta_{2,m+r} \ \dots \ \delta_{l,m+r}]^T. \end{aligned} \tag{15}$$

While the parameter space under the population ( $\Omega$ ) is given by

$$\begin{aligned} \Omega &= \{ \beta_0(u_i, v_i), \beta_{11}(u_i, v_i), \beta_{21}(u_i, v_i), \dots, \beta_{lm}(u_i, v_i), \\ &\delta_{1,m+1}(u_i, v_i), \dots, \delta_{l,m+r}(u_i, v_i), \sigma_\Omega^2(u_i, v_i) \} \\ &= \{ \tilde{\beta}(u_i, v_i), \tilde{\delta}(u_i, v_i); i = 1, 2, \dots, n, \sigma_\Omega^2(u_i, v_i) \} \\ &= \{ \tilde{\eta}_\Omega(u_i, v_i); i = 1, 2, \dots, n, \sigma_\Omega^2(u_i, v_i) \} \end{aligned} \tag{16}$$

Obtaining the test statistic of hypothesis above required some lemmas as follows.

**Lemma 3.** *If  $\tilde{\eta}_\Omega(u_i, v_i)$  is a parameter under population ( $\Omega$ ) from nonparametric spline regression with spatial heterogeneity (2), then estimator  $\hat{\tilde{\eta}}_\Omega(u_i, v_i)$  is given by*

$$\hat{\tilde{\eta}}_\Omega(u_i, v_i) = (Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i) \tilde{Y} \tag{17}$$

*Proof.* To obtain estimator  $\hat{\tilde{\eta}}_\Omega(u_i, v_i)$  we form likelihood function under population parameter space  $L(\Omega)$ . Therefore  $y_i$  has normal distribution with mean

$$\begin{aligned} &\beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pi}^k \\ &+ \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pi} - K_{ph})_+^m \end{aligned} \tag{18}$$

and variance  $\sigma_\Omega^2(u_i, v_i)$ ; then probability functions  $y_1, y_2, \dots, y_n$  are given by

$$\begin{aligned} f(y_1, y_2, \dots, y_n) &= \prod_{j=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma_\Omega^2(u_i, v_i)}} \right. \\ &\cdot \exp \left( -\frac{1}{2\sigma_\Omega^2(u_i, v_i)} w_{j(i)} \left[ y_j - \left( \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pj}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pj} - K_{ph})_+^m \right) \right]^2 \right) \left. \right\} \end{aligned} \tag{19}$$

Obtained likelihood function is as follows:

$$\begin{aligned} L(\Omega) &= (2\pi)^{-n/2} (\sigma_\Omega^2(u_i, v_i))^{-n/2} \\ &\cdot \exp \left( -\frac{1}{2\sigma_\Omega^2(u_i, v_i)} \sum_{j=1}^n w_{j(i)} \left[ y_j - \left( \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pj}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pj} - K_{ph})_+^m \right) \right]^2 \right) \end{aligned} \tag{20}$$

Equation (20) in matrix form is

$$\begin{aligned} L(\Omega) &= (2\pi)^{-n/2} (\sigma_\Omega^2(u_i, v_i))^{-n/2} \\ &\cdot \exp \left( -\frac{1}{2\sigma_\Omega^2(u_i, v_i)} (\tilde{Y} - Q\tilde{\eta}_\Omega(u_i, v_i))^T \mathbf{W}(u_i, v_i) \right. \\ &\cdot (\tilde{Y} - Q\tilde{\eta}_\Omega(u_i, v_i)) \left. \right). \end{aligned} \tag{21}$$

Estimator  $\hat{\tilde{\eta}}_\Omega(u_i, v_i)$  is obtained on the basis of the following derivative results:

$$\frac{\partial \ln L(\Omega)}{\partial \tilde{\eta}_\Omega(u_i, v_i)} = 0 \tag{22}$$

Then the following is obtained:

$$\begin{aligned} &\frac{\partial \ln L(\Omega)}{\partial \tilde{\eta}_\Omega(u_i, v_i)} \\ &= \frac{\partial (-2\tilde{\eta}_\Omega(u_i, v_i)^T Q^T \mathbf{W}(u_i, v_i) \tilde{Y})}{\partial \tilde{\eta}_\Omega(u_i, v_i)} \\ &+ \frac{\partial (\tilde{\eta}_\Omega(u_i, v_i)^T Q^T \mathbf{W}(u_i, v_i) Q \tilde{\eta}_\Omega(u_i, v_i))}{\partial \tilde{\eta}_\Omega(u_i, v_i)} \\ &= -2Q^T \mathbf{W}(u_i, v_i) \tilde{Y} + 2Q^T \mathbf{W}(u_i, v_i) Q \hat{\tilde{\eta}}_\Omega(u_i, v_i) \end{aligned} \tag{23}$$

Therefore,

$$\widehat{\eta}_{\Omega}(u_i, v_i) = (Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i) \bar{Y} \quad (24)$$

Furthermore, estimator  $\widehat{\sigma}_{\Omega}^2(u_i, v_i)$  is shown in Lemma 4.

$$L(\Omega) = (2\pi)^{-n/2} (\sigma_{\Omega}^2(u_i, v_i))^{-n/2} \cdot \exp\left(-\frac{1}{2\sigma_{\Omega}^2(u_i, v_i)} \sum_{j=1}^n \left[ y_j - \left( \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pi}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pi} - K_{ph})_+^m \right) \right]^2 \right) \quad (25)$$

is given by

$$\widehat{\sigma}_{\Omega}^2(u_i, v_i) = \frac{(\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))}{n} \quad (26)$$

$$L(\Omega) = (2\pi)^{-n/2} (\sigma_{\Omega}^2(u_i, v_i))^{-n/2} \cdot \exp\left(-\frac{1}{2\sigma_{\Omega}^2(u_i, v_i)} \sum_{j=1}^n \left[ y_j - \left( \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pj}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pj} - K_{ph})_+^m \right) \right]^2 \right) \quad (27)$$

The ln likelihood function is given by

$$\begin{aligned} \ln L(\Omega) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \sigma_{\Omega}^2(u_i, v_i) \\ &+ -\frac{1}{2\sigma_{\Omega}^2(u_i, v_i)} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T \\ &\cdot (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i)) \end{aligned} \quad (28)$$

**Lemma 4.** If  $\sigma_{\Omega}^2(u_i, v_i)$  is a parameter in space under population  $(\Omega)$  from nonparametric spline regression with spatial heterogeneity spatial (2), then estimator  $\widehat{\sigma}_{\Omega}^2(u_i, v_i)$  which is obtained from likelihood function:

*Proof.* Estimator  $\widehat{\sigma}_{\Omega}^2(u_i, v_i)$  is obtained using likelihood function:

Furthermore, estimator  $\widehat{\sigma}_{\Omega}^2(u_i, v_i)$  is obtained on the basis of the following derivative results:

$$\frac{\partial \ln L(\Omega)}{\partial \sigma_{\Omega}^2(u_i, v_i)} = 0 \quad (29)$$

Then the following was obtained:

$$\begin{aligned} \frac{\partial \ln L(\Omega)}{\partial \sigma_{\Omega}^2(u_i, v_i)} &= \frac{\partial \left( -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \sigma_{\Omega}^2(u_i, v_i) - \frac{1}{2\sigma_{\Omega}^2(u_i, v_i)} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i)) \right)}{\partial \sigma_{\Omega}^2(u_i, v_i)} \\ &= -\frac{n}{2\widehat{\sigma}_{\Omega}^2(u_i, v_i)} + \frac{1}{2(\widehat{\sigma}_{\Omega}^2(u_i, v_i))^2} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i)) \end{aligned} \quad (30)$$

Therefore,

$$\widehat{\sigma}_{\Omega}^2(u_i, v_i) = \frac{(\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))}{n} \quad (31)$$

Based on estimators  $\widehat{\eta}_{\Omega}(u_i, v_i)$  and  $\widehat{\sigma}_{\Omega}^2(u_i, v_i)$  which are given by Lemmas 3 and 4 the following is obtained:

$$L(\widehat{\Omega}) = L(\widehat{\eta}_{\Omega}(u_i, v_i), \widehat{\sigma}_{\Omega}^2(u_i, v_i)) = (2\pi)^{-n/2} (\widehat{\sigma}_{\Omega}^2(u_i, v_i))^{-n/2} \exp\left(-\frac{1}{2\widehat{\sigma}_{\Omega}^2(u_i, v_i)} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))\right)$$

$$\begin{aligned} &= (2\pi)^{-n/2} (\widehat{\sigma}_{\Omega}^2(u_i, v_i))^{-n/2} \exp\left(-\frac{1}{2\widehat{\sigma}_{\Omega}^2(u_i, v_i)} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))\right) \\ &= (2\pi)^{-n/2} (\widehat{\sigma}_{\Omega}^2(u_i, v_i))^{-n/2} \exp\left(-\frac{1}{2\widehat{\sigma}_{\Omega}^2(u_i, v_i)} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))\right) \\ &= (2\pi)^{-n/2} (\widehat{\sigma}_{\Omega}^2(u_i, v_i))^{-n/2} \exp\left(-\frac{1}{2\widehat{\sigma}_{\Omega}^2(u_i, v_i)} (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))^T (\bar{Y} - Q\widehat{\eta}_{\Omega}(u_i, v_i))\right) \end{aligned} \quad (32)$$

**Lemma 5.** If  $\tilde{\eta}_\omega$  and  $\sigma_\omega^2$  are parameters under  $H_0$  from multivariate nonparametric truncated spline influenced by spatial heterogeneity model (2), then estimator for  $\hat{\tilde{\eta}}_\omega$  is given by

$$\hat{\tilde{\eta}}_\omega = (Q^T Q)^{-1} Q^T \tilde{Y} \tag{33}$$

and estimator for  $\sigma_\omega^2$  is given by

$$\hat{\sigma}_\omega^2 = \frac{(\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)^T (\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)}{n} \tag{34}$$

*Proof.* To obtain estimators  $\hat{\tilde{\eta}}_\omega$  and  $\hat{\sigma}_\omega^2$  we form likelihood function under parameter space  $H_0 L(\omega)$ . Therefore  $y_i$  has normal distribution with mean

$$\beta_0 + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk} x_{pi}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h} (x_{pi} - K_{ph}) \tag{35}$$

and variance  $\sigma_\omega^2$ ; then probability functions  $y_1, y_2, \dots, y_n$  are given by

$$f(y_1, y_2, \dots, y_n) = \prod_{j=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma_\omega^2}} \exp \left( -\frac{1}{2\sigma_\omega^2} \left[ y_j - \left( \beta_0 + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk} x_{pi}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h} (x_{pi} - K_{ph}) \right) \right]^2 \right) \right\} \tag{36}$$

The following likelihood functions were obtained:

$$L(\omega) = (2\pi)^{-n/2} (\sigma_\omega^2)^{-n/2} \exp \left( -\frac{1}{2\sigma_\omega^2} \sum_{j=1}^n \left[ y_j - \left( \beta_0 + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk} x_{pi}^k + \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h} (x_{pi} - K_{ph}) \right) \right]^2 \right) \tag{37}$$

The equation is in the form of a matrix

$$L(\omega) = (2\pi)^{-n/2} (\sigma_\omega^2)^{-n/2} \cdot \exp \left( -\frac{1}{2\sigma_\omega^2} (\tilde{Y} - Q\tilde{\eta}_\omega)^T (\tilde{Y} - Q\tilde{\eta}_\omega) \right). \tag{38}$$

Estimators  $\hat{\tilde{\eta}}_\omega$  and  $\hat{\sigma}_\omega^2$  are obtained:

$$\begin{aligned} \hat{\tilde{\eta}}_\omega &= (Q^T Q)^{-1} Q^T \tilde{Y} \\ \hat{\sigma}_\omega^2 &= \frac{(\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)^T (\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)}{n} \end{aligned} \tag{39}$$

Based on Lemma 5, maximum likelihood function is obtained as follows:

$$L(\hat{\omega}) = L(\hat{\tilde{\eta}}_\omega, \hat{\sigma}_\omega^2) = (2\pi)^{-n/2} (\hat{\sigma}_\omega^2)^{-n/2} \cdot \exp \left( -\frac{1}{2\hat{\sigma}_\omega^2} (\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)^T (\tilde{Y} - Q\hat{\tilde{\eta}}_\omega) \right) \tag{40}$$

$\hat{\tilde{\eta}}_\omega$  and  $\hat{\sigma}_\omega^2$  are parameters estimator of under  $H_0$  from multivariate nonparametric regression with truncated spline approach.

### 5. Statistics Test for Truncated Spline Nonparametric Regression with Spatial Heterogeneity

The test statistic for the model conformity hypothesis test can be obtained by using Lemmas 3, 4, and 5. In the next step,

we show the likelihood ratio for test statistic presented in Lemma 6.

**Lemma 6.** If  $L(\hat{\Omega})$  and  $L(\hat{\omega})$ , respectively, are given by (32) and (40), then the likelihood ratio  $\Lambda$  is given by

$$\Lambda = \frac{\tilde{Y}^T S \tilde{Y}}{\tilde{Y}^T D(u_i, v_i) \tilde{Y}} \tag{41}$$

where

$$\begin{aligned} S &= (I - Q(Q^T Q)^{-1} Q^T) \\ D(u_i, v_i) &= (I - W(u_i, v_i) Q(Q^T W(u_i, v_i) Q)^{-1} Q^T) \\ &\cdot (I - Q(Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i)) \end{aligned} \tag{42}$$

*Proof.* Based on Lemmas 3, 4, and 5, and also (32) and (40), the likelihood ratio is obtained:

$$\begin{aligned} \Lambda &= \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{(2\pi)^{-n/2} (\hat{\sigma}_\omega^2)^{-n/2} \exp(-n/2)}{(2\pi)^{-n/2} (\hat{\sigma}_\Omega^2(u_i, v_i))^{-n/2} \exp(-n/2)} \\ &= \left( \frac{\hat{\sigma}_\omega^2}{\hat{\sigma}_\Omega^2(u_i, v_i)} \right)^{-n/2} \\ &= \left( \frac{(\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)^T (\tilde{Y} - Q\hat{\tilde{\eta}}_\omega)}{(\tilde{Y} - Q\hat{\tilde{\eta}}_\Omega(u_i, v_i))^T (\tilde{Y} - Q\hat{\tilde{\eta}}_\Omega(u_i, v_i))} \right)^{-n/2} \end{aligned} \tag{43}$$

Based on (3) and (5), the likelihood ratio

$$\begin{aligned}
 &= \left( \frac{(\tilde{Y} - Q(Q^T Q)^{-1} Q^T \tilde{Y})^T (\tilde{Y} - Q(Q^T Q)^{-1} Q^T \tilde{Y})}{(\tilde{Y} - Q(Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i) \tilde{Y})^T (\tilde{Y} - Q(Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i) \tilde{Y})} \right)^{-n/2} \\
 &= \left( \frac{\tilde{Y}^T (I - Q(Q^T Q)^{-1} Q^T)^T (I - Q(Q^T Q)^{-1} Q^T) \tilde{Y}}{\tilde{Y}^T (I - Q(Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i))^T (I - Q(Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i)) \tilde{Y}} \right)^{-n/2} \tag{44} \\
 &= \left( \frac{\tilde{Y}^T (I - Q(Q^T Q)^{-1} Q^T) \tilde{Y}}{\tilde{Y}^T (I - W(u_i, v_i) Q(Q^T W(u_i, v_i) Q)^{-1} Q^T) (I - Q(Q^T W(u_i, v_i) Q)^{-1} Q^T W(u_i, v_i)) \tilde{Y}} \right)^{-n/2} \\
 &= \left( \frac{\tilde{Y}^T S \tilde{Y}}{\tilde{Y}^T D(u_i, v_i) \tilde{Y}} \right)^{-n/2}
 \end{aligned}$$

□

Given test statistic for model conformity hypothesis is presented by Theorem 7.

**Theorem 7.** *If likelihood ratio  $\Lambda$  is given by Lemma 6, then test statistic for  $H_0$  versus  $H_1$  in (2) is given by*

$$V = \frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T (I - \xi))} \tag{45}$$

*Proof.* Based on Lemma 6, the likelihood ratio is as follows:

$$\Lambda = \left( \frac{\tilde{Y}^T S \tilde{Y}}{\tilde{Y}^T D(u_i, v_i) \tilde{Y}} \right)^{-n/2} \tag{46}$$

Based on MLRT method,  $H_0$  is rejected if

$$\Lambda = \left( \frac{\tilde{Y}^T S \tilde{Y}}{\tilde{Y}^T D(u_i, v_i) \tilde{Y}} \right)^{-n/2} < c \tag{47}$$

For a constant  $c$ , (43) is equivalent to

$$\left( \frac{\tilde{Y}^T S \tilde{Y}}{\tilde{Y}^T D(u_i, v_i) \tilde{Y}} \right) > c^{-2/n} \tag{48}$$

In the two sections of the inequality above, each numerator is divided by  $n - lm - 1$  and each denominator is divided by  $\text{tr}((I - \xi)^T (I - \xi))$ ; then the following inequality is obtained:

$$\begin{aligned}
 &\frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T (I - \xi))} \\
 &> c^{-2/n} \left( \frac{\text{tr}((I - \xi)^T (I - \xi))}{n - lm - 1} \right) = c^* \tag{49}
 \end{aligned}$$

Based on (44), the test statistic for  $H_0$  versus  $H_1$  is given by

$$V = \frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T (I - \xi))} \tag{50}$$

□

Furthermore, the distribution of statistics test  $V$  will be found.

The statistics test given in Theorem 7 is test statistics developed from the spline truncated approach in the GWR model, different from the one developed by Leung, Mei, and Zhang [9], Leung, Mei, and Zhang [10], and Mennis and Jordan [11] using GWR without using the Truncated Spine approach.

### 6. Distribution of Test Statistic and Critical Area of Hypothesis

To prove the distribution of test statistic  $V$ , we first prove  $\tau = \tilde{Y}^T S \tilde{Y} / \sigma^2 \sim \chi^2_{(n-lm-1)}$  and  $\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \sigma^2 (\sum_{i=1}^n \lambda_i^2 / \text{tr}((I - \xi)^T (I - \xi))) \sim \chi^2_{(\text{tr}((I - \xi)^T (I - \xi))^2 / \sum_{i=1}^n \lambda_i^2)}$ . The proofs are presented in Theorems 8 and 9 as follows.

**Theorem 8.** *If  $S$  is a matrix given by Lemma 6 then statistic is*

$$\tau = \frac{\tilde{Y}^T S \tilde{Y}}{\sigma^2} \sim \chi^2_{(n-lm-1)}. \tag{51}$$

*Proof.* To prove this Lemma, the following steps are taken.

Matrix  $S$  is shown which is a symmetric and idempotent matrix as follows:

$$S^T = (I - Q(Q^T Q)^{-1} Q^T)^T = (I - Q(Q^T Q)^{-1} Q^T) \tag{52}$$

Based on equation above, it is proved that matrix  $S$  is symmetric.

$$\begin{aligned}
 S^T S &= (I - Q(Q^T Q)^{-1} Q^T)^T (I - Q(Q^T Q)^{-1} Q^T) \\
 &= I - 2Q(Q^T Q)^{-1} Q^T \\
 &\quad + Q(Q^T Q)^{-1} Q^T Q(Q^T Q)^{-1} Q^T
 \end{aligned}$$

$$\begin{aligned}
 &= I - 2Q(Q^T Q)^{-1} Q^T + Q(Q^T Q)^{-1} Q^T \\
 &= I - Q(Q^T Q)^{-1} Q^T
 \end{aligned} \tag{53}$$

It is proved that matrix  $S$  is idempotent. Furthermore,  $\text{tr}(S)$  is calculated as follows:

$$\text{tr}(S) = \text{tr}\left(I - Q(Q^T Q)^{-1} Q^T\right) = n - lm - 1 \tag{54}$$

Therefore, it is proved that

$$\tau = \frac{\tilde{Y}^T S \tilde{Y}}{\sigma^2} \sim \chi_{(n-lm-1)}^2 \tag{55}$$

**Theorem 9.** If  $D(u_i, v_i)$  is a matrix given by Lemma 6 then statistic is

$$\begin{aligned}
 \tau^* &= \frac{\tilde{Y}^T D(u_i, v_i) \tilde{Y}}{\sigma^2 \left( \left( \sum_{i=1}^n \lambda_i^2 \right) / \text{tr} \left( (I - \xi)^T (I - \xi) \right) \right)} \\
 &\sim \chi_{\left( \frac{\text{tr} \left( (I - \xi)^T (I - \xi) \right)}{\sum_{i=1}^n \lambda_i^2} \right)}^2
 \end{aligned} \tag{56}$$

*Proof.* Based on (24), we obtain

$$\hat{\eta}_{\Omega}(u_i, v_i) = \left( Q^T W(u_i, v_i) Q \right)^{-1} Q^T W(u_i, v_i) \tilde{Y} \tag{57}$$

And estimator is obtained:

$$\begin{aligned}
 \hat{Y} &= Q \hat{\eta}_{\Omega}(u_i, v_i) \\
 &= Q \left( Q^T W(u_i, v_i) Q \right)^{-1} Q^T W(u_i, v_i) \tilde{Y} = \xi \tilde{Y}
 \end{aligned} \tag{58}$$

in which

$$\begin{aligned}
 \xi &= Q \left( Q^T W(u_i, v_i) Q \right)^{-1} Q^T W(u_i, v_i) \\
 &= \begin{bmatrix} q_1^T \left( Q^T W(u_1, v_1) Q \right)^{-1} Q^T W(u_1, v_1) \\ q_2^T \left( Q^T W(u_2, v_2) Q \right)^{-1} Q^T W(u_2, v_2) \\ \vdots \\ q_n^T \left( Q^T W(u_n, v_n) Q \right)^{-1} Q^T W(u_n, v_n) \end{bmatrix}
 \end{aligned} \tag{59}$$

$$q_i^T = [1 \quad x_{1i} \quad x_{1i}^2 \quad \cdots \quad x_{1i}^m \quad x_{2i} \quad \cdots \quad (x_{li} - K_{lr})_+^m]$$

Further, error vector is given as follows:

$$\hat{\varepsilon} = \tilde{Y} - \hat{Y} = (I - \xi) \tilde{Y} \tag{60}$$

Sum Square of Error (SSE) of model is obtained by squaring the following error vectors:

$$\begin{aligned}
 \text{SSE} &= \tilde{Y}^T (I - \xi)^T (I - \xi) \tilde{Y} = (\tilde{Y} - \xi \tilde{Y})^T (\tilde{Y} - \xi \tilde{Y}) \\
 &= \left( (I - \xi) (\tilde{Y} - E(\tilde{Y})) \right)^T \left( (I - \xi) (\tilde{Y} - E(\tilde{Y})) \right) \\
 &= (\tilde{Y} - E(\tilde{Y}))^T (I - \xi)^T (I - \xi) (\tilde{Y} - E(\tilde{Y})) \\
 &= \tilde{\varepsilon}^T (I - \xi)^T (I - \xi) \tilde{\varepsilon}
 \end{aligned} \tag{61}$$

Furthermore,

$$\begin{aligned}
 E(\text{SSE}) &= E\left(\tilde{\varepsilon}^T (I - \xi)^T (I - \xi) \tilde{\varepsilon}\right) \\
 &= \sigma^2 \text{tr}\left((I - \xi)^T (I - \xi)\right)
 \end{aligned} \tag{62}$$

Since SSE is a quadratic form from random variable:

$$\tilde{Y}^T (I - \xi)^T (I - \xi) \tilde{Y} \geq 0, \tag{63}$$

hence, matrix  $(I - \xi)^T (I - \xi)$  is positive semidefinite but not idempotent. Next, we obtain

$$\frac{\text{SSE}}{\sigma^2} = \tilde{\varepsilon}^{*T} (I - \xi)^T (I - \xi) \tilde{\varepsilon}^*, \quad \tilde{\varepsilon}^* = \frac{\tilde{\varepsilon}}{\sigma} \tag{64}$$

Since  $\tilde{\varepsilon} \sim N(\bar{0}, \sigma^2 I)$ , hence  $\tilde{\varepsilon}^* \sim N(\bar{0}, I)$ , and since matrix  $(I - \xi)^T (I - \xi)$  is not idempotent, the distribution of statistic is

$$\frac{\text{SSE}}{\sigma^2} \sim k \chi_r^2 \tag{65}$$

For constants  $k$  and  $r$ , based on (55), we obtain

$$E(\text{SSE}) = \sigma^2 \text{tr}\left((I - \xi)^T (I - \xi)\right) \tag{66}$$

$$E\left(\frac{\text{SSE}}{\sigma^2}\right) = \text{tr}\left((I - \xi)^T (I - \xi)\right)$$

Since  $(I - \xi)^T (I - \xi)$  is symmetric and positive semidefinite, hence there is an orthogonal matrix  $\Theta$ ; therefore,

$$\Theta (I - \xi)^T (I - \xi) \Theta^T = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \tag{67}$$

$\Lambda$  is a diagonal matrix in which  $\lambda_1, \lambda_2, \dots, \lambda_n$  are eigenvalues from matrix  $(I - \xi)^T (I - \xi)$ . Hence

$$\begin{aligned}
 \frac{\text{SSE}}{\sigma^2} &= \tilde{\varepsilon}^{*T} (I - \xi)^T (I - \xi) \tilde{\varepsilon}^* \\
 &= \tilde{\varepsilon}^{*T} \Theta^T (I - \xi)^T (I - \xi) \Theta \tilde{\varepsilon}^* \\
 &= \tilde{\varepsilon}^{*T} \Lambda \tilde{\varepsilon}^* \\
 &= \sum_{i=1}^n \lambda_i \varepsilon_i^{*2}
 \end{aligned} \tag{68}$$

in which  $\tilde{\varepsilon}^* = [\varepsilon_1^* \quad \varepsilon_2^* \quad \cdots \quad \varepsilon_n^*]^T$ . Random variables  $\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_n^*$  are independent, identical, and normal distributed; therefore

$$\varepsilon_i^{*2} \sim \chi_1^2 \tag{69}$$

with mean 1 and variance 2; therefore

$$\begin{aligned}
 \text{var}\left(\frac{\text{SSE}}{\sigma^2}\right) &= \text{var}\left(\sum_{i=1}^n \lambda_i \varepsilon_i^{*2}\right) \\
 &= \sum_{i=1}^n \lambda_i^2 \text{var}\left(\varepsilon_i^{*2}\right) \\
 &= 2 \sum_{i=1}^n \lambda_i^2
 \end{aligned} \tag{70}$$

Since  $SSE/\sigma^2 \sim k\chi_r^2$ ,

$$E\left(\frac{SSE}{\sigma^2}\right) = \text{tr}\left((I - \xi)^T(I - \xi)\right) = kr \tag{71}$$

$$\text{var}\left(\frac{SSE}{\sigma^2}\right) = 2\sum_{i=1}^n \lambda_i^2 = 2k^2r$$

Hence the values of  $k$  and  $r$  are as follows:  $r = (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2$  and  $k = (\sum_{i=1}^n \lambda_i^2) / \text{tr}((I - \xi)^T(I - \xi))$ ; as a result,

$$\frac{\tilde{Y}^T D(u_i, v_i) \tilde{Y}}{\sigma^2} \sim \left( \frac{\sum_{i=1}^n \lambda_i^2}{\text{tr}((I - \xi)^T(I - \xi))} \right) \chi_{(\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2}^2 \tag{72}$$

Hence

$$\tau^* = \frac{\tilde{Y}^T D(u_i, v_i) \tilde{Y}}{\sigma^2 \left( (\sum_{i=1}^n \lambda_i^2) / \text{tr}((I - \xi)^T(I - \xi)) \right)} \sim \chi_{(\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2}^2 \tag{73}$$

□ Hence

$$V = \frac{(\tilde{Y}^T S \tilde{Y} / (n - lm - 1))}{\left( \tilde{Y}^T D(u_i, v_i) \tilde{Y} / \left( (\sum_{i=1}^n \lambda_i^2) / \text{tr}((I - \xi)^T(I - \xi)) \right) \right) / \left( (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2 \right)}$$

$$= \frac{(\tilde{Y}^T S \tilde{Y} / (n - lm - 1))}{\left( \tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T(I - \xi)) \right)} \tag{77}$$

$$= \frac{(\tilde{Y}^T S \tilde{Y} / \sigma^2) / (n - lm - 1)}{\left( \tilde{Y}^T D(u_i, v_i) \tilde{Y} / \sigma^2 \left( (\sum_{i=1}^n \lambda_i^2) / \text{tr}((I - \xi)^T(I - \xi)) \right) \right) / \left( (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2 \right)}$$

$$\sim F_{((n - lm - 1), (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2)}$$

The critical area for the model conformity hypothesis is derived which is given by Lemma 11.

**Lemma 11.** *If given test statistic  $V$  is as in Theorem 7, then the critical area for  $H_0$  is given by*

$$C(y, x_1, x_2, \dots, x_l) = \{(y, x_1, x_2, \dots, x_l); V > c\} \tag{78}$$

A constant  $c$  is obtained according to

$$P(V > c) = \alpha, \quad 0 < \alpha < 1 \tag{79}$$

in which  $\alpha$  is a determined level of significance and

$$V \sim F_{(n - lm - 1, (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2)}. \tag{80}$$

*Proof.* Based on Theorem 7, the following relationship is obtained:

**Corollary 10.** *If statistic  $V$  is given by Theorem 7, then*

$$V = \frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T(I - \xi))} \sim F_{((n - lm - 1), (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2)} \tag{74}$$

*Proof.* Based on Theorem 8, statistic is obtained:

$$\tau = \frac{\tilde{Y}^T S \tilde{Y}}{\sigma^2} \sim \chi_{(n - lm - 1)}^2 \tag{75}$$

Based on Theorem 9, statistic is obtained:

$$\tau^* = \frac{\tilde{Y}^T D(u_i, v_i) \tilde{Y}}{\sigma^2 \left( (\sum_{i=1}^n \lambda_i^2) / \text{tr}((I - \xi)^T(I - \xi)) \right)} \sim \chi_{(\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2}^2 \tag{76}$$

$$\frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \left( (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2 \right)} > c^{-2/n} \frac{n - lm - 1}{\left( (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2 \right)} = c^* \tag{81}$$

for a constant  $c^*$ . According to Corollary 10, statistic is obtained:

$$V = \frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \left( (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2 \right)} \sim F_{((n - lm - 1), (\text{tr}((I - \xi)^T(I - \xi)))^2 / \sum_{i=1}^n \lambda_i^2)} \tag{82}$$

TABLE 1: Description of our research data and the predictor variables.

| Variable       | Data | Minimum   | Maximum    | Rata-rata   | Standard Deviation |
|----------------|------|-----------|------------|-------------|--------------------|
| Y              | 38   | 0,61      | 8,59       | 4,3939      | 1,81385            |
| X <sub>1</sub> | 38   | 4,59      | 25,80      | 12,0963     | 4,99263            |
| X <sub>2</sub> | 38   | 0,2105    | 7,6445     | 2,631579    | 1,6160512          |
| X <sub>3</sub> | 38   | 0,07      | 7,19       | 5,5747      | 1,22772            |
| X <sub>4</sub> | 38   | 35,10     | 33.548,70  | 3.378,7658  | 6.497,23435        |
| X <sub>5</sub> | 38   | 851582    | 2507632    | 1516816.21  | 420415.074         |
| X <sub>6</sub> | 38   | 0,0074281 | 0,1640892  | 0,052024430 | 0,0474488599       |
| X <sub>7</sub> | 38   | 0,0405726 | 12,4704588 | 2,631578947 | 2,8535682531       |
| X <sub>8</sub> | 38   | 474       | 85122      | 28730,32    | 22372,865          |

Source: BPS (2017a, 2017b, and 2017c).

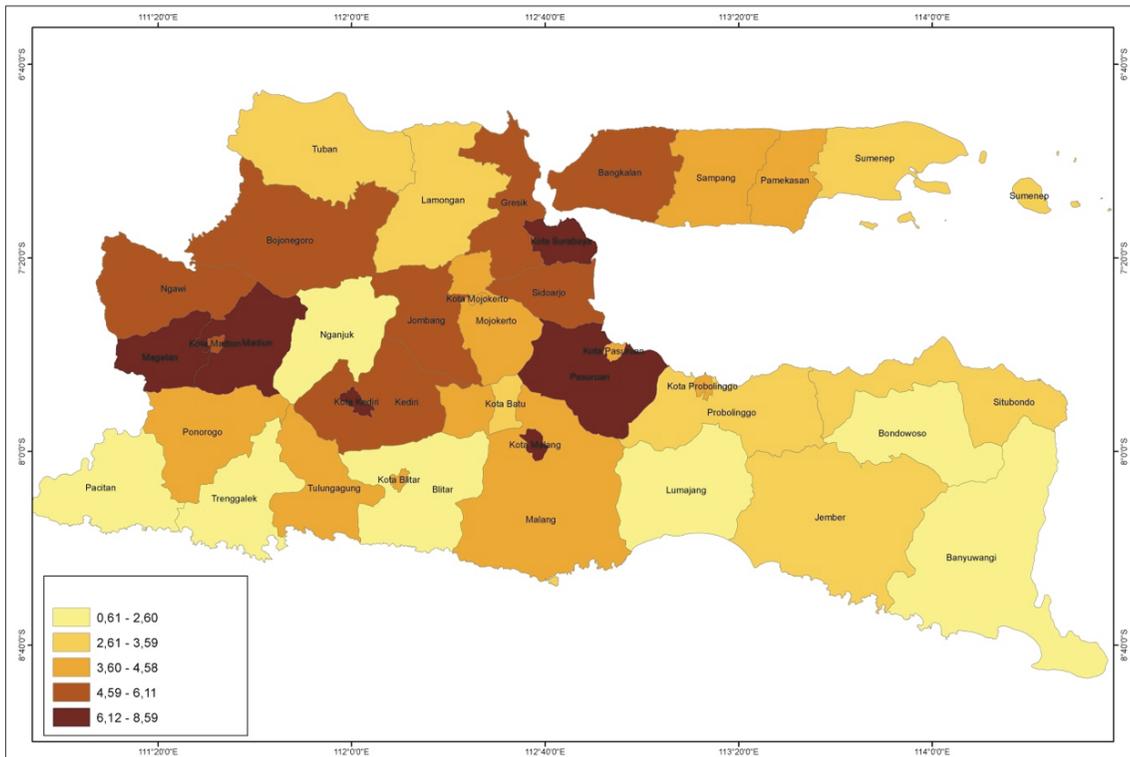


FIGURE 1: Open Unemployment Mapping in province of Java, Indonesia.

for a level of significance  $\alpha$  given by  $H_0$  which is rejected if

$$V > F_{\alpha((n-lm-1), (tr((I-\xi)^T(I-\xi)))^2 / \sum_{i=1}^n \lambda_i^2)} \quad (83)$$

After finding the hypothesis test formulation, the suitability of the model between the truncated spline nonparametric regression model which is influenced by spatial heterogeneity and nonparametric regression (global) will then be implemented on unemployment rate data in 38 regions in Java Indonesia.

### 7. Empirical Study on Unemployment Rate in Java Indonesia

7.1. Description of Research Data. In this study, the nonparametric truncated spline regression model influenced

by spatial heterogeneity was applied to Open Unemployment Rate (OUR) data in province of Java, Indonesia, and some predictor variables that were suspected to affect it, i.e., population density ( $X_1$ ), percentage of the poor ( $X_2$ ), percentage of population with low education ( $X_3$ ), percentage of population working in agriculture sector ( $X_4$ ), area of agricultural land ( $X_5$ ), economic growth rate ( $X_6$ ), regional minimum wage ( $X_7$ ), and ratio number of large industries being number of labor force ( $X_8$ ). The amount of data used is 382 from 38 provinces and 8 predictor variables. Table 1 shows the description of our research data and the predictor variables.

The spread of Open Unemployment Rate in East Java is shown by Figure 1. It shows the percentage of East Java unemployment rate in 2015.

TABLE 2: Breusch-Pagan test.

| Test          | Significance Value | Decision     |
|---------------|--------------------|--------------|
| Breusch-Pagan | 0.002414           | Reject $H_0$ |

7.2. *Spatial Heterogeneity Test.* Each region has different characteristics and different parameters, as well as different functional forms; this is what proves spatial aspect. Breusch-Pagan testing is used to see the spatial heterogeneity of each location. Table 2 shows the Breusch-Pagan test.

Since spatial effect testing is fulfilled, i.e., there are effects of spatial heterogeneity, then the case can be solved by using the point approach. Furthermore, an analysis was performed using nonparametric truncated spline regression model influenced by spatial heterogeneity.

7.3. *Model Conformity Test.* Hypotheses for model conformity test between multivariate nonparametric truncated spline regression model influenced by heterogeneity spatial and multivariate nonparametric truncated spline regression model (global) are as follows:

$$H_0: \beta_{pk}(u_i, v_i) = \beta_{pk} \text{ and } \delta_{p,m+h}(u_i, v_i) = \delta_{p,m+h},$$

$$p = 1, 2, \dots, 8; k = 1; h = 1, 2, 3; i = 1, 2, \dots, 38$$

$$H_1: \text{at least, there is one of } \beta_{pk}(u_i, v_i) \neq \beta_{pk} \text{ or } \delta_{p,m+h}(u_i, v_i) \neq \delta_{p,m+h},$$

$$p = 1, 2, \dots, 8; k = 1; h = 1, 2, 3; i = 1, 2, \dots, 38$$

The test statistic is given by Theorem 7 as follows:

$$V = \frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T (I - \xi))} \quad (84)$$

Matrix  $S$  was constructed by multivariate nonparametric truncated spline. Matrix  $D(u_i, v_i)$  was constructed by multivariate nonparametric truncated spline regression. Hence, the numerator is obtained:

$$\frac{\tilde{Y}^T S \tilde{Y}}{n - lm - 1} = 3,7814 \quad (85)$$

with degree of freedom  $n - lm - 1 = 29$ . Meanwhile the denominator is obtained:

$$\frac{\tilde{Y}^T D(u_i, v_i) \tilde{Y}}{\text{tr}((I - \xi)^T (I - \xi))} = 1,8356 \quad (86)$$

with degree of freedom  $\text{tr}((I - \xi)^T (I - \xi)) = 27,0762$ . Test statistic  $V = 2,06$  with level of significance  $\alpha = 0,05$  was obtained and concluded to reject  $H_0$  since  $V > F_{(0,05;29,27)} = 1,88$ . Therefore, there is a significant difference between multivariate nonparametric truncated spline regression influenced by spatial heterogeneity and nonparametric truncated spline regression. Due to the influence of geographical factors on the model, the appropriate model used is multivariate

nonparametric truncated spline regression influenced by spatial heterogeneity.

The modeling application used Open Unemployment Rate (OUR) data in 38 districts/cities in East Java. The results of the empirical study showed that the OUR data has a geographical influence, namely, spatial heterogeneity, and based on the results of the model conformity hypothesis test, the appropriate model used is a multivariable nonparametric truncated spline regression model influenced by spatial heterogeneity with the weighted Gaussian kernel function. The modeling produced a coefficient of determination of 80.42%.

## 8. Conclusion

Multivariate nonparametric regression with truncated spline approach influenced by spatial heterogeneity is given as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^l \sum_{k=1}^m \beta_{pk}(u_i, v_i) x_{pi}^k$$

$$+ \sum_{p=1}^l \sum_{h=1}^r \delta_{p,m+h}(u_i, v_i) (x_{pi} - K_{ph})_+^m + \varepsilon_i \quad (87)$$

Based on the results of the discussion and data analysis, some conclusions can be drawn as follows:

- (1) The hypotheses for model conformity between multivariate nonparametric truncated spline regression model influenced by spatial heterogeneity and nonparametric truncated spline regression (global) are as follows:

$$H_0: \beta_{pk}(u_i, v_i) = \beta_{pk} \text{ and } \delta_{p,m+h}(u_i, v_i) = \delta_{p,m+h},$$

$$p = 1, 2, \dots, l; k = 1, 2, \dots, m; h = 1, 2, \dots, r; i = 1, 2, \dots, n$$

$$H_1: \text{at least, there is one of } \beta_{pk}(u_i, v_i) \neq \beta_{pk} \text{ or } \delta_{p,m+h}(u_i, v_i) \neq \delta_{p,m+h},$$

$$p = 1, 2, \dots, l; k = 1, 2, \dots, m; h = 1, 2, \dots, r; i = 1, 2, \dots, n$$

Test statistic derived using Maximum Likelihood Ratio Test (MLRT) is obtained as follows:

$$V = \frac{\tilde{Y}^T S \tilde{Y} / (n - lm - 1)}{\tilde{Y}^T D(u_i, v_i) \tilde{Y} / \text{tr}((I - \xi)^T (I - \xi))} \quad (88)$$

- (2) The distribution of multivariate nonparametric truncated spline regression model influenced by spatial heterogeneity is as follows:

$$V \sim F_{((n-lm-1), (\text{tr}((I-\xi)^T(I-\xi)))^2 / \sum_{i=1}^n \lambda_i^2)} \quad (89)$$

with level of significance  $\alpha$ ; therefore  $H_0$  is rejected if

$$V > F_{\alpha((n-lm-1), (\text{tr}((I-\xi)^T(I-\xi)))^2 / \sum_{i=1}^n \lambda_i^2)} \quad (90)$$

## Data Availability

The data of 38 districts/cities in East Java used to support the findings of this study have been deposited in <https://www.bps.go.id/>.

## Conflicts of Interest

The authors of this work declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors thank The Ministry of Research, Technology and Higher Education, Republic of Indonesia/Kementerian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia (Kemenristekdikti RI) for funding this work.

## References

- [1] Sifriyani, Haryatmi, I. N. Budiantara, and Gunardi, "Geographically weighted regression with spline approach," *Far East Journal of Mathematical Sciences*, vol. 101, no. 6, pp. 1183–1196, 2017.
- [2] Sifriyani, Sri Haryatmi Kartiko, I. Nyoman Budiantara, and Gunardi, "Development of nonparametric geographically weighted regression using truncated spline approach," *Songklanakarın Journal of Science and Technology*, vol. 40, no. 4, pp. 909–920, 2018.
- [3] Sifriyani, *Multivariable Nonparametric Regression Truncated Spline in the Geographically Weighted Regression Models, [Dissertation]*, Universitas Gadjah Mada, 2018.
- [4] C. Brunson, A. S. Fotheringham, and M. Charlton, "Some notes on parametric significance tests for geographically weighted regression," *Journal of Regional Science*, vol. 39, no. 3, pp. 497–524, 1999.
- [5] U. Demšar, A. S. Fotheringham, and M. Charlton, "Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics," *Information Visualization*, vol. 7, no. 3-4, pp. 181–197, 2008.
- [6] Y. Li, Y. Jiao, and J. A. Browder, "Modeling spatially-varying ecological relationships using geographically weighted generalized linear model: A simulation study based on longline seabird bycatch," *Fisheries Research*, vol. 181, pp. 14–24, 2016.
- [7] S.-S. Wu, H. Yang, F. Guo, and R.-M. Han, "Spatial patterns and origins of heavy metals in Sheyang River catchment in Jiangsu, China based on geographically weighted regression," *Science of the Total Environment*, vol. 580, pp. 1518–1529, 2017.
- [8] F. Benassi and A. Naccarato, "Households in potential economic distress. A geographically weighted regression model for Italy, 2001–2011," *Spatial Statistics*, vol. 21, no. part B, pp. 362–376, 2017.
- [9] Y. Leung, C.-L. Mei, and W.-X. Zhang, "Statistical tests for spatial nonstationarity based on the geographically weighted regression model," *Environment and Planning A*, vol. 32, no. 1, pp. 9–32, 2000.
- [10] Y. Leung, C.-L. Mei, and W.-X. Zhang, "Testing for spatial autocorrelation among the errors of the geographically weighted regression," *Environment and Planning A*, vol. 32, no. 1, pp. 871–890, 2000.

- [11] J. L. Mennis and L. Jordan, "The distribution of environmental equity: Exploring spatial nonstationarity in multivariate models of air toxic releases," *Annals of the Association of American Geographers*, vol. 95, no. 2, pp. 249–268, 2005.

