Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 15(1) 2022: 12-23
http://ejournal.undip.ac.id/index.php/media_statistika

ESTIMATION OF SEMIPARAMETRIC REGRESSION CURVE WITH MIXED ESTIMATOR OF MULTIVARIABLE LINEAR TRUNCATED SPLINE AND MULTIVARIABLE KERNEL

Hesikumalasari[1], I Nyoman Budiantara[2], Vita Ratnasari[2], Khaerun Nisa[3]
[1] Universitas Islam Negeri Mataram, Indonesia
[2] Institut Teknologi Sepuluh Nopember, Indonesia
[3] Balai Penelitian dan Pengembangan Agama Makassar, Indonesia

e-mail: hesikumalasari@uinmataram.ac.id

DOI: 10.14710/medstat.15.1.12-23

Article Info:
Received: 5 March 2021
Accepted: 10 June 2022
Available Online: 27 July 2022

Keywords:
Semiparametric regression, mixed estimator, linear truncated spline, kernel, multivariable

Abstract: The response variable of the regression analysis has a linear relationship with one of the variable predictors, however the unknown relationship pattern with the other predictor variables. Consequently, it can be approached by using semiparametric regression model. The predictor variable that has a linear relationship with the response variable can be approached by using linear parametric curve called parametric component. Meanwhile, the unknown relationship between the response variable with another predictor variable can be approached by using nonparametric curve called nonparametric component. If the predictor variable in nonparametric component is more than one, then it can be approached by using a different nonparametric curve named combined or mixed estimator. In this research, nonparametric component is approached using mixed estimator of multivariable linear truncated spline and multivariable kernel. The objective of this research is to estimate the model of semiparametric regression curve with mixed estimator of multivariable truncated spline and multivariable kernel. Estimation of this mixed model using ordinary least square method.

1. INTRODUCTION

Regression analysis is the statistical method that is used to estimate the relationship pattern of the predictor variables and response variables. The main purpose of the regression analysis is to find out the estimated regression curve. The approaches which are generally applied to estimate the regression curve are parametric regression and nonparametric regression. Parametric regression is used when the regression curve is known whilst the nonparametric regression is used if the regression curve is unknown (Barry & Hardle, 1993). Meanwhile, there are some cases in regression analysis where there are both parametric and nonparametric components. Regression model which contains these two components is called semiparametric regression (Ruppert et al., 2003).

Semiparametric regression model which is developed by the scientists so far uses the same estimation for some or even all of its predictor variable. This is due to the assumption that the pattern of each predictor is considered to have the same pattern so researchers use only one form of estimator model for all predictor variables. Meanwhile, the reality is often encountered cases with different data patterns of each predictor variable. Therefore, to overcome these problems some researchers have developed mixed estimator of nonparametric regression curve such that each pattern of the data in nonparametric regression model is approached by the estimator curves corresponding to the data pattern.

Previous research that uses mixed estimator have been conducted by (Wayan Sudiarsa et al., 2015), (Nurcahayani et al., 2021), (Nisa' & Budiantara, 2020), which used mixed estimator of truncated spline and Fourier series, and (Mariati et al., 2020) using mixed estimator of smoothing spline and Fourier series. The mixed estimator of kernel and Fourier series has been conducted by (Afifah et al., 2017) and (Nisa et al., 2017). Meanwhile, previous research that use mixed estimator of spline and kernel have been conducted by (Ratnasari et al., 2016), (Budiantara et al., 2015), (Rismal et al., 2016), and (Hidayat et al., 2020) only involve a mixed estimator in nonparametric regression. There has been no research involving the mixed estimator on the semiparametric regression. Therefore, this research uses a mixed model of multivariable linear truncated spline and multivariable kernel on semiparametric regression. So, the objective of this research is to estimate the model of semiparametric regression with mixed estimator of multivariable truncated spline and multivariable kernel.

2. LITERATURE REVIEW

2.1. Parametric, Nonparametric, and Semiparametric Regression

Regression Analysis is a statistical method that explains the relationship pattern between an explanatory variable X and a response variable Y. Given a paired data (x_i, y_i), $i = 1,2,\ldots,n$, which the relationship pattern can be expressed in the regression model as follows (Barry & Hardle, 1993).

$$y_i = f(x_i) + \varepsilon_i \tag{1}$$

with y_i is the response variable, $f(x_i)$ is regression function, and ε_i is an independent random error, distributed normally with zero mean and σ^2 variance.

Parametric regression is used if the shape of the $f(x_i)$ curve in Equation (1) is known, which means that the pattern of the relationship between the predictor variable and the response variable is known. Meanwhile, nonparametric regression is used if the shape of the $f(x_i)$ curve in Equation (1) is unknown (Barry & Hardle, 1993). If regression model consists both of parametric components and nonparametric components, then it is called semiparametric regression (Ruppert et al., 2003). Given paired data $(\mathrm{x_i}, \mathrm{u_i}, \mathrm{v_i}, \mathrm{y_i})$, $i = 1,2,\ldots,n$, that are assumed to follow the model of semiparametric regression in the Equation (2).

$$y_i = f(x_i) + g(u_i) + h(v_i) + \varepsilon_i \tag{2}$$

with y_i is the response variable, the $f(x_i)$ curve is a parametric component, the $\mathrm{g(u_i)}$ and $h(v_i)$ are nonparametric components, ε_i is an independent random error, distributed normally with zero mean and σ^2 variance.

2.2. Linear Parametric Regression

Given a paired data (x_i, y_i), $i = 1,2,\ldots,n$, that have a relationship pattern expressed in the regression model in the Equation (1). Regression curve $f(x_i)$ is assumed approached using linear parametric regression. (Ellis et al., 1968) defined the linear form of parametric regression as written in the Equation (3).

$$f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \quad (3)$$

with β_0 is a constant and $\beta_1, \beta_2, \dots, \beta_p$ are coefficient of the predictor variables, $x_{i1}, x_{i2}, \dots, x_{ip}$ are the predictor variables.

2.3. Truncated Spline Linear and Kernel

Given a paired data $(x_i, u_i, v_i y_i), i = 1,2,\dots,n,$ which the relationship pattern expressed in regression model in Equation (2). Regression curve $g(u_i)$ is assumed to be approached using linear truncated spline function. (Barry & Hardle, 1993) defined linear truncated spline function as a random function which can be expressed in the Equation (4).

$$g(u_i) = \theta_1 u_i + \sum_{k=1}^{m} \lambda_k (u_i - K_k)_+ \quad (4)$$

with

$$(u_i - K_k)_+ = \begin{cases} (u_i - K_k)_+ \,; u_i \geq K_k \\ 0 \,; u_i < K_k \end{cases}$$

where $\theta_1, \lambda_1, \lambda_2, \dots, \lambda_m$ are unknown parameters, K_k is Knot-k, $k = 1,2,\dots,m$, and $K_1 < K_2 < \cdots < K_m$.

Regression curve $h(v_i)$ is assumed to be approached using kernel function. To estimate the regression curve $h(v_i)$ in nonparametric regression model, Nadaraya and Watson, at (Barry & Hardle, 1993) define kernel regression estimator that called Nadaraya-Watson estimator as written in the Equation (5).

$$\hat{h}_\alpha(v_i) = n^{-1} \sum_{j=1}^{n} W_{\alpha j}(v_i) y_j \quad (5)$$

with

$$W_{\alpha j}(v_i) = \frac{\frac{1}{\alpha} K\left(\frac{v_i - v_j}{\alpha}\right)}{n^{-1} \sum_{j=1}^{n} \frac{1}{\alpha} K\left(\frac{v_i - v_j}{\alpha}\right)}$$

Kernel estimator depends on the kernel function K and bandwidth parameter α. One of the several types of kernel functions is a Gaussian kernel. (Barry & Hardle, 1993) define the form of Gaussian kernel as written in the Equation (6).

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty \quad (6)$$

3. METHODOLOGY

3.1. Research Method

Steps to get parameter estimator of semiparametric regression curve with mixed estimator of multivariable linear truncated spline and multivariable kernel as follows:

a. Determine the semiparametric model;

b. Determine whether the semiparametric model be combined or mixed model;
c. Determine the parametric component (linear parametric curve) and nonparametric component (linear truncated spline and kernel curve);
d. Find the matrix form of semiparametric model;
e. Find the estimator of semiparametric regression curve by ordinary least square method as follows:
   1. Determine the error equation of the model;
   2. Determine the sum square of error;
   3. Find the first partial derivative of sum square of error;
   4. Equalizing the first derivative of sum square of error to zero;
f. Getting the estimator of semiparametric regression curve with mixed estimator of multivariable linear truncated spline and multivariable kernel.

4. RESULT

4.1. Estimation of Semiparametric Regression Curve with Mixed Estimator of Multivariable Linear Truncated Spline and Multivariable Kernel

Given a paired data $(x_i, u_{1i}, u_{2i}, \dots, u_{pi}, v_{1i}, v_{2i}, \dots, v_{qi}, y_i)$, $i = 1,2,\dots,n$ which is assumed to follow the semiparametric regression model with multivariable nonparametric component which is given by Equation (7).

$$y_i = \mu(x_i, u_{1i}, u_{2i}, \dots, u_{pi}, v_{1i}, v_{2i}, \dots, v_{qi}) + \varepsilon_i \quad (7)$$

If y_i is assumed to follow the mixed model, then it can be written as Equation (8).

$$y_i = f(x_i) + \sum_{r=1}^{p} g_r(u_{ri}) + \sum_{s=1}^{q} h_s(v_{si}) + \varepsilon_i \quad (8)$$

with y_i is the response variable, $f(x_i)$ curve, $i = 1,2,\dots,n$ is assumed referring to linear patter (parametric component), $g_r(u_{ri})$ curves, $r = 1,2,\dots,p$ and $h_s(v_{si})$ curves, $s = 1,2,\dots,q$ are assumed as nonparametric components. $\mathrm{g_r(u_{ri})}$ curve is assumed to be approached by using linear truncated spline function, $\mathrm{h_s(v_{si})}$ curve is assumed to be approached by using Nadaraya-Watson Kernel function and ε_i is an independent random error, normally distributed with zero mean and σ^2 variance. To get the estimation of semiparametric regression curve with mixed estimator of multivariable linear truncated spline and multivariable kernel, given some of the following lemmas.

Lemma 4.1 *Model of multivariable semiparametric regression is given in Equation (8). If the $f(x_i)$ curve, $i = 1,2,\dots,n$ is assumed to refer to a linear regression pattern, then:*

$$\tilde{f}(x) = \boldsymbol{X}\tilde{\beta}$$

with

$$\tilde{f}(x) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Proof. The $f(x_i)$ curve is assumed to refer to a linear pattern/linear parametric function in Equation (3), so the model is obtained for $i = 1$ as follows:

$$f(x_1) = \beta_0 + \beta_1\beta_{i1} + \beta_2\beta_{12} + \cdots + \beta_p\beta_{1p}$$

A similar way can be obtained for $i = 2,3, \dots, n$, so the model for $i = 1,2, \dots, n$ is as follow:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} \end{bmatrix}$$

Then can be written as:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

So, we obtain the following model:

$$\tilde{f}(x) = \boldsymbol{X}\tilde{\beta}$$ ■

Lemma 4.2 *Model of multivariable semiparametric regression is given in Equation (8). if the $g_r(u_{ri})$ curve, $r = 1,2, \dots, p$ is assumed approached by using linear truncated spline function, then:*

$$\sum_{r=1}^{p} g_r(u_{ri}) = \boldsymbol{G}(\tilde{k})\tilde{\theta}$$

with

$$\textstyle\sum_{r=1}^{p} g_r(u_{ri}) = g_1(u_{1i}) + g_2(u_{2i}) + \cdots + g_p(u_{pi}),$$

$$\boldsymbol{G}(\tilde{k}) = [\boldsymbol{G}(k_1) \quad \boldsymbol{G}(k_2) \quad \dots \quad \boldsymbol{G}(k_p)],$$

$$\tilde{\theta} = \begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \vdots \\ \tilde{\theta}_p \end{bmatrix}$$

Proof. The $g_r(u_{ri})$ curve is assumed to be approached by using linear truncated spline function in Equation (4), then it can be written as:

$$g_r(u_{ri}) = \theta_{r1} u_{ri} + \lambda_{r1}(u_{ri} - K_{r1})_+ + \cdots + \lambda_{rm}(u_{ri} - K_{rm})_+$$

with $\theta_1, \lambda_1, \lambda_2, \dots, \lambda_m$ are unknown parameters.

Model is obtained for $r = 1$ as follows:

$$g_1(u_{1i}) = \theta_{11} u_{1i} + \lambda_{11}(u_{1i} - K_{11})_+ + \cdots + \lambda_{1m}(u_{1i} - K_{1m})_+$$

with *i=1, 2,…, n*, the model is:

$$\begin{bmatrix} g_1(u_{11}) \\ g_1(u_{12}) \\ \vdots \\ g_1(u_{1n}) \end{bmatrix} = \begin{bmatrix} \theta_{11} u_{11} + \lambda_{11}(u_{11} - K_{11})_+ + \cdots + \lambda_{1m}(u_{11} - K_{1m})_+ \\ \theta_{11} u_{12} + \lambda_{11}(u_{12} - K_{11})_+ + \cdots + \lambda_{1m}(u_{12} - K_{1m})_+ \\ \vdots \\ \theta_{11} u_{1n} + \lambda_{11}(u_{1n} - K_{11})_+ + \cdots + \lambda_{1m}(u_{1n} - K_{1m})_+ \end{bmatrix}$$

Then it can be written as:

$$\begin{bmatrix} g_1(u_{11}) \\ g_1(u_{12}) \\ \vdots \\ g_1(u_{1n}) \end{bmatrix} = \begin{bmatrix} u_{11} & (u_{11}-K_{11}) & \dots & (u_{11}-K_{1m})_+ \\ u_{12} & (u_{12}-K_{11}) & \dots & (u_{12}-K_{1m})_+ \\ \vdots & \vdots & \ddots & \vdots \\ u_{1n} & (u_{1n}-K_{11}) & \dots & (u_{1n}-K_{1m})_+ \end{bmatrix} \begin{bmatrix} \theta_{11} \\ \lambda_{11} \\ \vdots \\ \lambda_{1m} \end{bmatrix}$$

Then we obtain the following model:

$$\tilde{g}_1(u_1) = \boldsymbol{G}(k_1)\tilde{\theta}_1$$

Similarly, it can be obtained for $r = 2$ until for $r = p$, so for multivariabel linear truncated spline with $r = 1,2,\dots,p$ we obtain the model as follow:

$$\sum_{r=1}^{p} g_r(u_{ri}) = \begin{bmatrix} \boldsymbol{G}(k_1) & \boldsymbol{G}(k_2) & \dots & \boldsymbol{G}(k_p) \end{bmatrix} \begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \vdots \\ \tilde{\theta}_p \end{bmatrix}$$

$$\sum_{r=1}^{p} g_r(u_{ri}) = \boldsymbol{G}(\tilde{k})\tilde{\theta}$$ ■

Lemma 4.3 *Model of multivariable semiparametric regression is given in Equation (8). If the $h_s(v_{si})$ curve, $s = 1,2,\dots q$ is assumed to be approached by using kernel, then:*

$$\sum_{s=1}^{q} \hat{h}_s(v_s) = \boldsymbol{D}(\tilde{\alpha})\tilde{y}$$

with

$$\boldsymbol{D}(\tilde{\alpha}) = \begin{bmatrix} n^{-1}\sum_{k=1}^{q} W_{\alpha_k 1}(v_1) & n^{-1}\sum_{k=1}^{q} W_{\alpha_k 2}(v_1) & \dots & n^{-1}\sum_{k=1}^{q} W_{\alpha_k n}(v_1) \\ n^{-1}\sum_{k=1}^{q} W_{\alpha_k 1}(v_2) & n^{-1}\sum_{k=1}^{q} W_{\alpha_k 2}(v_2) & \dots & n^{-1}\sum_{k=1}^{q} W_{\alpha_k n}(v_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}\sum_{k=1}^{q} W_{\alpha_k 1}(v_n) & n^{-1}\sum_{k=1}^{q} W_{\alpha_k 2}(v_n) & \dots & n^{-1}\sum_{k=1}^{q} W_{\alpha_k n}(v_n) \end{bmatrix}$$

and

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

Proof. The $h_s(v_{si})$ curve is assumed to be approached by using kernel function in the Equation (5), then can be written as follow:

$$\hat{h}_{\alpha_s}(v_i) = n^{-1}\sum_{j=1}^{n} W_{\alpha_{sj}}(v_i)y_j$$

Model is obtained for $s = 1$ as follows:

$$\hat{h}_{\alpha_1}(v_i) = n^{-1}\sum_{j=1}^{n} W_{\alpha_{1j}}(v_i)y_j$$

with *i=1,2,...,n*, model is obtained:

$$\begin{bmatrix}\hat{h}_{\alpha_1}(v_1)\\ \hat{h}_{\alpha_1}(v_2)\\ \vdots\\ \hat{h}_{\alpha_1}(v_n)\end{bmatrix} = \begin{bmatrix} n^{-1}W_{\alpha_1 1}(v_1)y_1 + n^{-1}W_{\alpha_1 2}(v_1)y_2 + \cdots + n^{-1}W_{\alpha_1 n}(v_1)y_n\\ n^{-1}W_{\alpha_1 1}(v_2)y_1 + n^{-1}W_{\alpha_1 2}(v_2)y_2 + \cdots + n^{-1}W_{\alpha_1 n}(v_2)y_n\\ \vdots\\ n^{-1}W_{\alpha_1 1}(v_n)y_1 + n^{-1}W_{\alpha_1 2}(v_n)y_2 + \cdots + n^{-1}W_{\alpha_1 n}(v_n)y_n\end{bmatrix}$$

Then can be written as:

$$\begin{bmatrix}\hat{h}_{\alpha_1}(v_1)\\ \hat{h}_{\alpha_1}(v_2)\\ \vdots\\ \hat{h}_{\alpha_1}(v_n)\end{bmatrix} = \begin{bmatrix} n^{-1}W_{\alpha_1 1}(v_1) & n^{-1}W_{\alpha_1 2}(v_1) & \dots & n^{-1}W_{\alpha_1 n}(v_1)\\ n^{-1}W_{\alpha_1 1}(v_2) & n^{-1}W_{\alpha_1 2}(v_2) & \dots & n^{-1}W_{\alpha_1 n}(v_2)\\ \vdots & \vdots & \ddots & \vdots\\ n^{-1}W_{\alpha_1 1}(v_n) & n^{-1}W_{\alpha_1 2}(v_n) & \dots & n^{-1}W_{\alpha_1 n}(v_n)\end{bmatrix}\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}$$

Then obtained the following model:

$$\underset{\sim}{\hat{h}}_{\alpha_1}(v) = \boldsymbol{D}(\alpha_1)\tilde{y}$$

Similarly, can be obtained for $s = 2$ until for $s = q$, so for multivariabel kernel with $s = 1,2,\dots,q$ obtained the model as follows:

$$\sum_{s=1}^{q}\hat{h}_s(v_s) = \boldsymbol{D}(\alpha_1)\tilde{y} + \boldsymbol{D}(\alpha_2)\tilde{y} + \cdots + \boldsymbol{D}(\alpha_q)\tilde{y}$$

$$\sum_{s=1}^{q}\hat{h}_s(v_s) = \boldsymbol{D}(\tilde{\alpha})\tilde{y}$$ ■

Based on Lemma 4.1, Lemma 4.2, and Lemma 4.3 we obtained the following theorem.

Theorem 4.1 *Given a paired data* $(x_i, u_{1i}, u_{2i}, \dots, u_{pi}, v_{1i}, v_{2i}, \dots, v_{qi}, y_i)$, $i = 1,2,\dots,n$ *which are assumed to follow the mixed model:*

$$y_i = \mu(x_i, u_{1i}, u_{2i}, \dots, u_{pi}, v_{1i}, v_{2i}, \dots, v_{qi}) + \varepsilon_i$$

$$= f(x_i) + \sum_{r=1}^{p} g_r(u_{ri}) + \sum_{s=1}^{q} h_s(v_{si}) + \varepsilon_i$$

If the $f(x_i)$ *curve,* $i = 1,2,\dots,n$ *is assumed to be approached by using linear parametric function, the* $g_r(u_{ri})$ *curve,* $r = 1,2,\dots,p$ *is assumed to be approached by using linear truncated spline function, and the* $h_s(v_{si})$ *curve,* $s = 1,2,\dots q$ *is assumed to be approached by using kernel, then:*

a. Semiparametric regression curve estimator for linear parametric component as follows:

$$\hat{\hat{f}}(x) = \boldsymbol{R}(\tilde{k}, \tilde{\alpha})\tilde{y}$$

with

$$\boldsymbol{R}(\tilde{k},\tilde{\alpha}) = \boldsymbol{X}\left[\boldsymbol{I} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}\right]^{-1}$$

$$\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha})) - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\right]$$

b. *Semiparametric regression curve estimator for multivariable linear truncated spline as follows:*

$$\hat{\tilde{g}}(u) = \boldsymbol{S}(\tilde{k},\tilde{\alpha})\tilde{y}$$

with $\boldsymbol{S}(\tilde{k},\tilde{\alpha})$ *as follow*

$$\boldsymbol{G}(\tilde{k})\left[\boldsymbol{I} - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\right]^{-1}$$

$$\left[\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha})) - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\right]$$

c. *Semiparametric regression curve estimator for the mixed of multivariable linear truncated spline and multivariable kernel as follows:*

$$\hat{\tilde{\mu}}(x,u,v) = \boldsymbol{T}(\tilde{k},\tilde{\alpha})\tilde{y}$$

with

$$\boldsymbol{T}(\tilde{k},\tilde{\alpha}) = \boldsymbol{R}(\tilde{k},\tilde{\alpha}) + \boldsymbol{S}(\tilde{k},\tilde{\alpha}) + \boldsymbol{D}(\tilde{\alpha})$$

Proof. The model of semiparametric regression which contains multivariable nonparametric components is given in Equation (8). Because the $f(x_i)$ curve, $i = 1,2,\ldots,n$ is assumed to be approached by using linear parametric function, the $g_r(u_{ri})$ curve, $r = 1,2,\ldots,p$ is assumed to be approached by using linear truncated spline function, and the $h_s(v_{si})$ curve, $s = 1,2,\ldots q$ is assumed approached by using kernel, then the model can be written as:

$$\tilde{y} = \boldsymbol{X}\tilde{\beta} + \boldsymbol{G}(\tilde{k})\tilde{\theta} + \boldsymbol{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon}$$

From model above, can be obtained:

$$\tilde{\varepsilon} = \tilde{y} - \boldsymbol{X}\tilde{\beta} + \boldsymbol{G}(\tilde{k})\tilde{\theta} + \boldsymbol{D}(\tilde{\alpha})\tilde{y}$$

Then the error can be written as follow:

$$\tilde{\varepsilon} = ((\boldsymbol{I} - \boldsymbol{D}(\tilde{\alpha}))\tilde{y} - \boldsymbol{X}\tilde{\beta} - \boldsymbol{G}(\tilde{k})\tilde{\theta}$$

The estimator of the model was obtained by using Ordinary Least Square (OLS) method. After completing optimization, then the error be:

$$\tilde{\varepsilon}^T\tilde{\varepsilon} = \tilde{y}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} - 2\tilde{\beta}^T\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y}$$
$$-2\tilde{\theta}^T\boldsymbol{G}(\tilde{k})^{\boldsymbol{T}}(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} + \tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta}$$
$$+2\tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\tilde{\theta} + \tilde{\theta}^T\boldsymbol{G}(\tilde{k})^{\boldsymbol{T}}\boldsymbol{G}(\tilde{k})\tilde{\theta}$$

For example, the error is $M(\tilde{\beta},\tilde{\theta})$, $M(\tilde{\beta},\tilde{\theta})$ can be written as follow:

$$M(\tilde{\beta},\tilde{\theta}) = \tilde{y}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} - 2\tilde{\beta}^T\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y}$$
$$-2\tilde{\theta}^T\boldsymbol{G}(\tilde{k})^{\boldsymbol{T}}(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} + \tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta}$$
$$+2\tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\tilde{\theta} + \tilde{\theta}^T\boldsymbol{G}(\tilde{k})^{\boldsymbol{T}}\boldsymbol{G}(\tilde{k})\tilde{\theta}$$

Then optimize $M(\tilde{\beta},\tilde{\theta})$ using the partial derivative to $\tilde{\beta}$ and $\tilde{\theta}$, partial derivative to $\tilde{\beta}$ is obtained:

$$\frac{\partial M(\tilde{\beta},\tilde{\theta})}{\partial\tilde{\beta}} = 0 - 2\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} + 2\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta} + 2\tilde{\beta}^T\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\tilde{\theta}$$

Meanwhile, a partial derivative to $\tilde{\theta}$ is obtained:

$$\frac{\partial M(\tilde{\beta},\tilde{\theta})}{\partial\tilde{\theta}} = 0 - 2\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} + 2\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}\tilde{\beta} + 2G(\tilde{k})^T\boldsymbol{G}(\tilde{k})\tilde{\theta}$$

Then equalizing the first derivative of sum square of error with zero, for $\tilde{\beta}$ obtained:

$$\hat{\tilde{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\tilde{\theta}$$

Meanwhile, equalizing the first derivative of sum square of error with zero for $\tilde{\theta}$ is obtained:

$$\hat{\tilde{\theta}} = \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\tilde{y} - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}\tilde{\beta}$$

Then substitute $\hat{\tilde{\theta}}$ into $\hat{\tilde{\beta}}$, followed by calculation the estimator of $\hat{\tilde{\beta}}$ is obtained**:**

$$\hat{\tilde{\beta}} = \left[\boldsymbol{I} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}\right]^{-1}$$
$$\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha})) - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\right]\tilde{y}$$

Based on $\hat{\tilde{\beta}}$, the estimator of semiparametric regression curve for linear parametric component is obtained:

$$\hat{\tilde{f}}(x) = \boldsymbol{X}\hat{\tilde{\beta}}$$
$$\hat{\tilde{f}}(x) = \boldsymbol{X}\left[\boldsymbol{I} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}\right]^{-1}$$
$$\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha})) - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\right]\tilde{y}$$
$$\hat{\tilde{f}}(x) = \boldsymbol{R}(\tilde{k},\tilde{\alpha})\tilde{y}$$

Then substitution of $\hat{\tilde{\beta}}$ into $\hat{\tilde{\theta}}$, followed by calculation so the estimator of $\hat{\tilde{\theta}}$ is obtained:

$$\hat{\tilde{\theta}} = \left[\boldsymbol{I} - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\right]^{-1}$$
$$\left[\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha})) - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{D}(\tilde{\alpha}))\right]\tilde{y}$$

Based on $\hat{\tilde{\theta}}$ the estimator of semiparametric regression curve for multivariable linear truncated spline component is obtained:

$$\hat{\hat{g}}(u) = \boldsymbol{G}(\tilde{k})\hat{\hat{\theta}}$$

$\hat{\hat{g}}$(u) can be written as

$$\boldsymbol{G}(\tilde{k})\left[\boldsymbol{I} - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}(\tilde{k})\right]^{-1}$$

$$\left[\left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T(\boldsymbol{I} - \boldsymbol{D}(\tilde{\alpha})) - \left(\boldsymbol{G}(\tilde{k})^T\boldsymbol{G}(\tilde{k})\right)^{-1}\boldsymbol{G}(\tilde{k})^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\left(\boldsymbol{I} - \boldsymbol{D}(\tilde{\alpha})\right)\right]\tilde{y}$$

$$\hat{\hat{g}}(u) = \boldsymbol{S}(\tilde{k}, \tilde{\alpha})\tilde{y}$$

So, the estimator of semiparametric regression curve for the mixed of multivariable linear truncated spline and multivariable kernel could be written in these three following terms:

$$\hat{\hat{\mu}}(x, u, v) = \hat{\hat{f}}(x) + \hat{\hat{g}}(u) + \hat{\hat{h}}(v)$$

$$\hat{\hat{\mu}}(x, u, v) = \left(\boldsymbol{R}(\tilde{k}, \tilde{\alpha}) + \boldsymbol{S}(\tilde{k}, \tilde{\alpha}) + \boldsymbol{D}(\tilde{\alpha})\right)\tilde{y}$$

$$\hat{\hat{\mu}}(x, u, v) = \boldsymbol{T}(\tilde{k}, \tilde{\alpha})\tilde{y} \qquad \blacksquare$$

5. CONCLUSION

The semiparametric model in this paper consists of two components which are parametric component and nonparametric component. Parametric component is approached using linear parametric function. Meanwhile, the nonparametric component is approached using mixed of multivariable linear truncated spline function and multivariable kernel function. Given a paired data $(x_i, u_{1i}, u_{2i}, \dots, u_{pi}, v_{1i}, v_{2i}, \dots, v_{qi}, y_i)$ which are assumed to follow the mixed model:

$$y_i = \mu(x_i, u_{1i}, u_{2i}, \dots, u_{pi}, v_{1i}, v_{2i}, \dots, v_{qi}) + \varepsilon_i$$
$$= f(x_i) + \sum_{r=1}^{p} g_r(u_{ri}) + \sum_{s=1}^{q} h_s(v_{si}) + \varepsilon_i$$

If the $f(x_i)$ curve, $i = 1,2, \dots, n$ is approached referring to linear pattern, the $g_r(u_{ri})$ curve, $r = 1,2, \dots, p$ is assumed to be approached by using linear truncated spline function, and the $h_s(v_{si})$ curve, $s = 1,2, \dots q$ is assumed to be approached by using kernel function. The following model was obtained:

$$\tilde{y} = \boldsymbol{X}\tilde{\beta} + \boldsymbol{G}(\tilde{k})\tilde{\theta} + \boldsymbol{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon}$$

The estimator is found by using *ordinary least square* method. Semiparametric regression curve estimator for linear parametric component as follows:

$$\hat{\hat{f}}(x) = \boldsymbol{R}(\tilde{k}, \tilde{\alpha})\tilde{y}$$

Semiparametric regression curve estimator for multivariable linear truncated spline as follow:

$$\hat{\hat{g}}(u) = \boldsymbol{G}(\tilde{k})\hat{\hat{\theta}}$$

Semiparametric regression curve estimator for the mixed of multivariable linear truncated spline and multivariable kernel as follow:

$$\hat{\hat{\mu}}(x,u,v) = \boldsymbol{T}(\tilde{k},\tilde{\alpha})\tilde{y}$$

with

$$\boldsymbol{T}(\tilde{k},\tilde{\alpha}) = \boldsymbol{R}(\tilde{k},\tilde{\alpha}) + \boldsymbol{S}(\tilde{k},\tilde{\alpha}) + \boldsymbol{D}(\tilde{\alpha})$$

REFERENCE

Afifah, N., Budiantara, I. N., & Latra, I. N. (2017). Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression. *Journal of Physics: Conference Series*, *855*(1). https://doi.org/10.1088/1742-6596/855/1/012002

Barry, D., & Hardle, W. (1993). Applied Nonparametric Regression. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *156*(1), 128. https://doi.org/10.2307/2982873

Budiantara, I. N., Ratnasari, V., Ratna, M., & Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties. *Applied Mathematical Sciences*, *9*(122), 6083–6094. https://doi.org/10.12988/ams.2015.58517

Ellis, D. M., Draper, N. P., & Smith, H. (1968). Applied Regression Analysis. *Applied Statistics*, *17*(1), 83. https://doi.org/10.2307/2985274

Hidayat, R., Budiantara, I. N., Otok, B. W., & Ratnasari, V. (2020). The Regression Curve Estimation by Using Mixed Smoothing Spline and Kernel (MsS-K) Model. *Communications in Statistics - Theory and Methods*. https://doi.org/10.1080/03610926.2019.1710201

Mariati, N. P. A. M., Budiantara, I. N., & Ratnasari, V. (2020). Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression. *Journal of Mathematics*, *2020*. https://doi.org/10.1155/2020/4712531

Nisa', K., & Budiantara, I. N. (2020). Modeling East Java Indonesia Life Expectancy Using Semiparametric Regression Mixed Spline Truncated and Fourier Series. *Media Statistika*, *13*(2), 149–160. https://doi.org/10.14710/medstat.13.2.149-160

Nisa, K., Budiantara, I. N., & Rumiati, A. T. (2017). Multivariable Semiparametric Regression Model with Combined Estimator of Fourier Series and Kernel. *IOP Conference Series: Earth and Environmental Science*, *58*(1), 12028. https://doi.org/10.1088/1755-1315/58/1/012028

Nurcahayani, H., Budiantara, I. N., & Zain, I. (2021). The Semiparametric Regression Curve Estimation by Using Mixed Truncated Spline and Fourier Series Model. *AIP Conference Proceedings*, *2329*(1), 060025. https://doi.org/10.1063/5.0042870

Ratnasari, V., Budiantara, I. N., Ratna, M., & Zain, I. (2016). Estimation of Nonparametric Regression Curve Using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel. *Global Journal of Pure and Applied Mathematics*, *12*(6), 5047–5057.

Rismal, Budiantara, I. N., & Prastyo, D. D. (2016). Mixture Model of Spline Truncated and Kernel in Multivariable Nonparametric Regression. *AIP Conference Proceedings*, *1739*(1), 020085. https://doi.org/10.1063/1.4952565

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Wayan Sudiarsa, I., Nyoman Budiantara, I., Suhartono, S., & Purnami, S. W. (2015). Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression. *Applied Mathematical Sciences*, *9*(97–100), 4997–5010. https://doi.org/10.12988/ams.2015.55394