

## Research Article

# Nonparametric Regression Model for Longitudinal Data with Mixed Truncated Spline and Fourier Series

Made Ayu Dwi Octavanny , I. Nyoman Budiantara , Heri Kuswanto ,  
and Dyah Putri Rahmawati

*Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia*

Correspondence should be addressed to I. Nyoman Budiantara; [nyomanbudiantara65@gmail.com](mailto:nyomanbudiantara65@gmail.com)

Received 30 July 2020; Revised 24 November 2020; Accepted 1 December 2020; Published 10 December 2020

Academic Editor: Victor Kovtunen

Copyright © 2020 Made Ayu Dwi Octavanny 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.

Existing literature in nonparametric regression has established a model that only applies one estimator to all predictors. This study is aimed at developing a mixed truncated spline and Fourier series model in nonparametric regression for longitudinal data. The mixed estimator is obtained by solving the two-stage estimation, consisting of a penalized weighted least square (PWLS) and weighted least square (WLS) optimization. To demonstrate the performance of the proposed method, simulation and real data are provided. The results of the simulated data and case study show a consistent finding.

## 1. Introduction

Regression analysis is aimed at modeling the association between the predictor and the response. If the data pattern shows an unknown regression curve, nonparametric regression is used [1]. However, if the form of the regression curve is known, parametric regression can be applied [2]. Additionally, nonparametric regression has high flexibility because the data is expected to find its regression curve estimation form without being influenced by the researcher's subjectivity [3]. In this study, we have analyzed several models such as kernel, spline [4–7], and Fourier series [8].

A spline estimator, which has an excellent ability to handle data with changes at subspecified intervals [9], was obtained using penalized least square optimization [10] and the Bayesian approach [11]. A spline estimator can be applied for cross-sectional data as well as longitudinal data. Additionally, several studies on nonparametric regression for longitudinal data have been addressed using kernel estimator [12, 13], generalized spline regression [14], and mixed-effects model [7]. Fourier series, which is useful to explain curves that show sine and cosine waves, is generally used if the data pattern is unknown and there is a tendency to iterate.

A considerable amount of research has used only one estimator for each predictor. However, because each predictor can have a different pattern, it was proposed to develop a mixed estimator. Recently, Sudiarsa et al. [15] discussed a study of the mixed estimator of the truncated spline and Fourier series. The study, which only discussed cross-sectional data, did not obtain a model for each subject as it did not include longitudinal data. Consequently, this study cannot be used to investigate response behavior based on the time change.

Although some research has been carried out on a mixed estimator, no studies have explored multisubject data so far. This paper proposes a new methodology for a mixed estimator of the truncated spline and Fourier series in the nonparametric regression for longitudinal data. This study addresses the gap in previous research by obtaining a mixed estimator of the truncated spline and Fourier series in the nonparametric regression for longitudinal data and applying it to simulated data and a case study.

This study is organized as follows. We briefly explain the materials and methods used in our study in Section 2. Section 3 consists of three subsections: the developed theory, simulation study, and case study. We present the developed nonparametric regression theory for longitudinal data with a

mixed estimator of the truncated spline and Fourier series with two-stage estimation in Section 3.1. In Section 3.2, we conduct a simulation study based on the developed theory to assess the proposed estimator’s behavior. To illustrate the applicability of the model, we use a dataset of patients with pulmonary tuberculosis in Section 3.3. Section 4 presents the conclusion.

## 2. Materials and Methods

Longitudinal data has  $n$  independent subjects and  $T$  observations for each subject. Given paired data  $(x_{1it}, \dots, x_{pit}, z_{1it}, \dots, z_{qit}, y_{it})$ , which consist of  $p$  and  $q$  predictors with  $n$  subjects, each subject has  $T$  observations. The relationship between  $x_{1it}, \dots, x_{pit}, z_{1it}, \dots, z_{qit}$  and  $y_{it}$ , which followed a nonparametric regression model for longitudinal data, is as follows:

$$y_{it} = \mu_{it}(x_{1it}, \dots, x_{pit}, z_{1it}, \dots, z_{qit}) + \varepsilon_{it}, i = 1, 2, \dots, n, t = 1, 2, \dots, T. \tag{1}$$

Each regression curve is additive so that the model can be expressed as

$$\mu_{it}(x_{1it}, \dots, x_{pit}, z_{1it}, \dots, z_{qit}) = \sum_{j=1}^p f_{ji}(x_{jit}) + \sum_{k=1}^q g_{ki}(z_{kit}), i = 1, 2, \dots, n, t = 1, 2, \dots, T. \tag{2}$$

$\sum_{j=1}^p f_{ji}(x_{jit})$  is the truncated spline component and  $\sum_{k=1}^q g_{ki}(z_{kit})$  is the Fourier series component.

This study’s first objective is to obtain the mixed estimator of the truncated spline and Fourier series in nonparametric regression for longitudinal data. To achieve this, we propose a two-stage estimation method. The first stage is estimating the components of the Fourier series using the penalized weighted least square (PWLS) method. The second stage is estimating the truncated spline component using the weighted least square (WLS) method. For the second goal, that is, a simulation study, we generate functions that meet the truncated spline and Fourier series characteristics. In the third step, we apply the developed theory to a dataset of patients with pulmonary tuberculosis.

## 3. Results and Discussion

**3.1. Mixed Model of Truncated Spline and Fourier Series with Two-Stage Estimation.** Lemmas and theorems are used to obtain a nonparametric regression model for longitudinal data with a mixed estimator. The regression curve component that is approximated by the Fourier series estimator is presented in Lemma 1 and the penalty component for the Fourier series function is presented in Lemma 2. Following the PWLS form in Lemma 3, we estimate the Fourier series component by using PWLS in Theorem 4. The regression curve component that is approximated by the truncated spline estimator is presented in Lemma 5, and we estimate the truncated spline component using the WLS method in Theorem 6. The results are summarized as follows:

**Lemma 1.** *If  $g_{ki}(z_{kit})$  is approximated by the Fourier series function, then the goodness of fit is*

$$N^{-1} \sum_{i=1}^n \sum_{t=1}^T w_{it} \left( y_{it}^* - \sum_{k=1}^q \sum_{t=1}^T g_{ki}(z_{kit}) \right)^2 = N^{-1} (\mathbf{y}^* - \mathbf{Zc})' \mathbf{W} (\mathbf{y}^* - \mathbf{Zc}), \tag{3}$$

where  $N = n \times T$  and  $\mathbf{W}$  is the  $nT \times nT$  weighting matrix.

*Proof.* The regression curve  $g_{ki}(z_{kit})$  is a regression curve of an unknown shape and is contained in continuous space  $C(0, \pi)$ . The component of  $g_{ki}(z_{kit}), k = 1, 2, \dots, q$ , in Equation (2) is approximated by the Fourier series function with the trend line as follows:

$$g_{ki}(z_{kit}) = d_{ki}z_{kit} + \frac{1}{2}c_{0ki} + \sum_{h=1}^H c_{hki} \cos hz_{kit} \tag{4}$$

If the regression curve  $\mathbf{g}$  in Equation (4) involves only one predictor, then it can be written as follows:

$$\mathbf{g} = (\mathbf{g}'_1 \mathbf{g}'_2 \dots \mathbf{g}'_n)', \tag{5}$$

$$\mathbf{g}_i = (g_{ki}(z_{ki1}) g_{ki}(z_{ki2}) \dots g_{ki}(z_{kiT}))', i = 1, 2, \dots, n. \tag{6}$$

By using Equation (4), Equation (6) can be written in the form of a matrix as follows:

$$\mathbf{g}_i = \begin{bmatrix} g_{ki}(z_{ki1}) \\ g_{ki}(z_{ki2}) \\ \vdots \\ g_{ki}(z_{kiT}) \end{bmatrix} = \begin{bmatrix} d_{ki}z_{ki1} + \frac{1}{2}c_{0ki} + c_{1ki} \cos 1z_{ki1} + c_{2ki} \cos 2z_{ki1} + \dots + c_{Hki} \cos Hz_{ki1} \\ d_{ki}z_{ki2} + \frac{1}{2}c_{0ki} + c_{1ki} \cos 1z_{ki2} + c_{2ki} \cos 2z_{ki2} + \dots + c_{Hki} \cos Hz_{ki2} \\ \vdots \\ d_{ki}z_{kiT} + \frac{1}{2}c_{0ki} + c_{1ki} \cos 1z_{kiT} + c_{2ki} \cos 2z_{kiT} + \dots + c_{Hki} \cos Hz_{kiT} \end{bmatrix} = \begin{bmatrix} z_{ki1} & \frac{1}{2} & \cos 1z_{ki1} & \cos 2z_{ki1} & \dots & \cos Hz_{ki1} \\ z_{ki2} & \frac{1}{2} & \cos 1z_{ki2} & \cos 2z_{ki2} & \dots & \cos Hz_{ki2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{kiT} & \frac{1}{2} & \cos 1z_{kiT} & \cos 2z_{kiT} & \dots & \cos Hz_{kiT} \end{bmatrix} \begin{bmatrix} d_{ki} \\ c_{0ki} \\ c_{1ki} \\ c_{2ki} \\ \vdots \\ c_{Hki} \end{bmatrix} = \mathbf{Z}_{ki} \mathbf{c}_{ki}. \tag{7}$$

The Fourier series function in the nonparametric regression component for longitudinal data with predictor  $q$  can be expressed in the following form:

$$\mathbf{g}_i = \sum_{k=1}^q \mathbf{Z}_{ki} \mathbf{c}_{ki} = \mathbf{Z}_{1i} \mathbf{c}_{1i} + \mathbf{Z}_{2i} \mathbf{c}_{2i} + \dots + \mathbf{Z}_{qi} \mathbf{c}_{qi} = \mathbf{Z}_i \mathbf{c}_i. \tag{8}$$

So,

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}'_1 \\ \mathbf{g}'_2 \\ \vdots \\ \mathbf{g}'_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \mathbf{c}_1 \\ \mathbf{Z}_2 \mathbf{c}_2 \\ \vdots \\ \mathbf{Z}_n \mathbf{c}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \end{bmatrix} = \mathbf{Zc}. \tag{9}$$

$\mathbf{Z}$  is a  $nT \times (2 + H) nq$  matrix as follows:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix}, \tag{10}$$

with

$$\mathbf{Z}_i = \begin{bmatrix} z_{1i1} & \frac{1}{2} & \cos z_{1i1} & \cos 2z_{1i1} & \cdots & \cos Hz_{1i1} & \cdots & z_{qi1} & \frac{1}{2} & \cos z_{qi1} & \cos 2z_{qi1} & \cdots & \cos Hz_{qi1} \\ z_{1i2} & \frac{1}{2} & \cos z_{1i2} & \cos 2z_{1i2} & \cdots & \cos Hz_{1i2} & \cdots & z_{qi2} & \frac{1}{2} & \cos z_{qi2} & \cos 2z_{qi2} & \cdots & \cos Hz_{qi2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{1iT} & \frac{1}{2} & \cos z_{1iT} & \cos 2z_{1iT} & \cdots & \cos Hz_{1iT} & \cdots & z_{qiT} & \frac{1}{2} & \cos z_{qiT} & \cos 2z_{qiT} & \cdots & \cos Hz_{qiT} \end{bmatrix}. \tag{11}$$

Whereas,  $\mathbf{c}$  is a  $(2 + H) nq \times 1$  vector given by

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \end{bmatrix}, \mathbf{c}_i = [d_{1i} \quad c_{01i} \quad c_{11i} \quad c_{21i} \quad \cdots \quad c_{H1i} \quad \cdots \quad d_{qi} \quad c_{0qi} \quad c_{1qi} \quad c_{2qi} \quad \cdots \quad c_{Hqi}]'. \tag{12}$$

To estimate the Fourier series component, the nonparametric regression model in Equation (1) can be written as

$$y_{it} - \sum_{j=1}^p f_{ji}(x_{jit}) = \sum_{k=1}^q g_{ki}(z_{kit}) + \varepsilon_{it}, \tag{13}$$

$$y_{it}^* = \sum_{k=1}^q g_{ki}(z_{kit}) + \varepsilon_{it}, i = 1, 2, \dots, n, t = 1, 2, \dots, T. \tag{14}$$

The model in Equation (14) can be written in matrix form:

$$\mathbf{y}^* = \mathbf{g} + \varepsilon. \tag{15}$$

Then, a goodness of fit for the model is formed as follows:

$$N^{-1} \sum_{i=1}^n \sum_{t=1}^T w_{it} \left( y_{it}^* - \sum_{k=1}^q g_{ki}(z_{kit}) \right)^2, \tag{16}$$

with  $N = n \times T$ . If the function  $g_{ki}(z_{kit})$  is approached by a Fourier series function as in Equation (4), then the goodness

of fit can be presented in the form

$$\begin{aligned} & N^{-1} \sum_{i=1}^n \sum_{t=1}^T w_{it} \left( y_{it}^* - \sum_{k=1}^q \sum_{t=1}^T g_{ki}(z_{kit}) \right)^2 \\ &= N^{-1} \sum_{i=1}^n \sum_{t=1}^T w_{it} \left( y_{it}^* - \sum_{k=1}^q \sum_{t=1}^T \left( d_{ki} z_{kit} + \frac{1}{2} c_{0ki} + \sum_{h=1}^H c_{hki} \cos hz_{kit} \right) \right)^2 \\ &= N^{-1} (\mathbf{y}^* - \mathbf{Zc})' \mathbf{W} (\mathbf{y}^* - \mathbf{Zc}), \end{aligned} \tag{17}$$

with  $\mathbf{W}$  as a weighting matrix for the regression of longitudinal data.

**Lemma 2.** *If the Fourier series is given, then the penalty component is*

$$\sum_{k=1}^q \lambda_k \int_0^{\pi/2} \frac{2}{\pi} \left( g_{ki}''(z_k) \right)^2 dz_k = \sum_{k=1}^q \lambda_k \int_0^{\pi/2} \frac{2}{\pi} \left( g_{ki}'(z_k) \right)^2 dz_k = \mathbf{c}' \mathbf{D}(\lambda) \mathbf{c}. \tag{18}$$

*Proof.* The penalty component  $\sum_{k=1}^q \lambda_k \int_0^{\pi/2} (2/\pi) (g_{ki}''(z_k))^2 dz_k$

in the PWLS optimization based on Equation (4) can be obtained as follows:

$$g_{ki}''(z_k) = \frac{d}{dz_k} \left[ \frac{d}{dz_k} \left( d_{ki}z_{kit} + \frac{1}{2}c_{0ki} + \sum_{h=1}^H c_{hki} \cos hz_{kit} \right) \right] = - \sum_{h=1}^H h^2 c_{hki} \cos hz_{kit}. \tag{19}$$

As a result,

$$P_k(g_k) = \int_0^\pi \frac{2}{\pi} \left( \sum_{h=1}^H h^2 c_{hki} \cos hz_{kit} \right)^2 dz_k = \frac{2}{\pi} \int_0^\pi \left[ \left( \sum_{h=1}^H h^2 c_{hki} \cos hz_{kit} \right)^2 + 2 \sum_{h < m} (h^2 c_{hki} \cos hz_{kit})(m^2 c_{mki} \cos mz_{kit}) \right] dz_k. \tag{20}$$

To simplify, we defined

$$A = \frac{2}{\pi} \sum_{h=1}^H \int_0^\pi (h^2 c_{hki} \cos hz_{kit})^2 dz_k, \tag{21}$$

$$B = \frac{2}{\pi} 2 \sum_{h < l} \int_0^\pi (h^2 c_{hki} \cos hz_{kit})(l^2 c_{lki} \cos lz_{kit}) dz_k.$$

The value of A will be obtained as follows.

$$A = \frac{2}{\pi} \sum_{h=1}^H \int_0^\pi (h^2 c_{hki} \cos hz_{kit})^2 dz_k = \frac{1}{\pi} \sum_{h=1}^H h^4 c_{hki}^2 \left[ z_{kit} + \frac{2}{h} \sin hz_{kit} \right]_0^\pi = \sum_{h=1}^H h^4 c_{hki}^2. \tag{22}$$

Furthermore, the value of B is given by

$$B = \frac{2}{\pi} 2 \sum_{h < l} \int_0^\pi (h^2 c_{hki} \cos hz_{kit})(l^2 c_{lki} \cos lz_{kit}) dz_k = \frac{4}{\pi} \sum_{h < l} (hl)^2 c_{hki} c_{lki} \int_0^\pi \cos hz_{kit} \cos lz_{kit} dz_k = 0. \tag{23}$$

Based on Equations (22) and (23), it can be written as follows:

$$P_k(g_k) = A + B = \sum_{h=1}^H h^4 c_{hki}^2. \tag{24}$$

For  $i = 1, 2, \dots, n$ , we obtained

$$\sum_{k=1}^q \lambda_k P_k(g_k) = \sum_{i=1}^n \mathbf{c}'_i \mathbf{D}(\lambda) \mathbf{c}_i = \mathbf{c}'_1 \mathbf{D}(\lambda) \mathbf{c}_1 + \mathbf{c}'_2 \mathbf{D}(\lambda) \mathbf{c}_2 + \dots + \mathbf{c}'_n \mathbf{D}(\lambda) \mathbf{c}_n = \mathbf{c}' \mathbf{D}(\lambda) \mathbf{c}, \tag{25}$$

where

$$\mathbf{D}(\lambda) = \begin{bmatrix} \mathbf{D}(\lambda_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}(\lambda_2) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{D}(\lambda_q) \end{bmatrix}, \tag{26}$$

$$\mathbf{D}(\lambda_i) = \begin{bmatrix} 0 & 0 & \lambda_i 1^4 & \lambda_i 2^4 & \dots & \lambda_i H^4 \\ 0 & 0 & \lambda_i 1^4 & \lambda_i 2^4 & \dots & \lambda_i H^4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_i 1^4 & \lambda_i 2^4 & \dots & \lambda_i H^4 \end{bmatrix} \dots$$

Thus, the penalty component can be expressed in a matrix form as follows:

$$\sum_{k=1}^q \lambda_k \int_0^\pi \frac{2}{\pi} (g_{ki}''(z_k))^2 dz_k = \mathbf{c}' \mathbf{D}(\lambda) \mathbf{c}. \tag{27}$$

**Lemma 3.** *If the goodness of fit component is presented in Lemma 1 and the penalty component is given by Lemma 2, then the PWLS is*

$$N^{-1}(\mathbf{y}^* - \mathbf{Zc})' \mathbf{W}(\mathbf{y}^* - \mathbf{Zc}) + \mathbf{c}' \mathbf{D}(\lambda) \mathbf{c}. \tag{28}$$

In general, PWLS is defined as follows:

$$N^{-1} \sum_{i=1}^n \sum_{t=1}^T w_{it} \left( y_{it}^* - \sum_{k=1}^q g_{ki}(z_{kit}) \right)^2 + \sum_{k=1}^q \lambda_k \int_0^\pi \frac{2}{\pi} (g_{ki}''(z_k))^2 dz_k, 0 < \lambda_k < \infty. \tag{29}$$

Besides, PWLS can be presented in the form of a matrix as follows:

$$N^{-1}(\mathbf{y}^* - \mathbf{Zc})' \mathbf{W}(\mathbf{y}^* - \mathbf{Zc}) + \mathbf{c}' \mathbf{D}(\lambda) \mathbf{c}. \tag{30}$$

The next step is to obtain a Fourier series estimator in nonparametric regression for longitudinal data derived in Theorem 4.

**Theorem 4.** *If paired data are given, which follows the non-parametric regression model for longitudinal data, then the*

mixed estimator that minimizes PWLS in Lemma 3,

$$\min_{g_k \in C(0, \pi)} \left\{ N^{-1} \sum_{i=1}^n \sum_{t=1}^T w_{it} \left( y_{it}^* - \sum_{k=1}^q \sum_{t=1}^T g_{ki}(z_{kit}) \right)^2 + \sum_{k=1}^q \lambda_k \int_0^\pi \frac{2}{\pi} \left( g_k''(z_k) \right)^2 dz_k \right\}, \tag{31}$$

is  $\widehat{\mathbf{g}}_{(k,h,\lambda)}(\mathbf{x}, \mathbf{z}) = \mathbf{L}\mathbf{y}^*$  with  $\mathbf{y}^* = \mathbf{y} - \mathbf{f}$  and  $\mathbf{L} = \mathbf{Z}[\mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda)]^{-1}\mathbf{Z}'\mathbf{W}$ .

*Proof.* The first estimation step in the mixed estimator of the truncated spline and Fourier series in the nonparametric regression model for longitudinal data is performed by estimating the form of the Fourier series estimator by using the PWLS method. The PWLS in Equation (29) can be written in the form of a matrix as follows:

$$Q(\mathbf{c}) = N^{-1}(\mathbf{y}^* - \mathbf{Z}\mathbf{c})'\mathbf{W}(\mathbf{y}^* - \mathbf{Z}\mathbf{c}) + \mathbf{c}'\mathbf{D}(\lambda)\mathbf{c} = N^{-1}\mathbf{y}^{*\prime}\mathbf{W}\mathbf{y}^* - 2N^{-1}\mathbf{c}'\mathbf{Z}'\mathbf{W}\mathbf{y}^* + N^{-1}\mathbf{c}'\mathbf{Z}'\mathbf{W}\mathbf{Z}\mathbf{c} + \mathbf{c}'\mathbf{D}(\lambda)\mathbf{c}. \tag{32}$$

Next, we complete PWLS optimization using the following steps:

$$\min_{\mathbf{c} \in \mathbb{R}^{(2+H)mq}} \{Q(\mathbf{c})\} = \min_{\mathbf{c} \in \mathbb{R}^{(2+H)mq}} \left\{ N^{-1}\mathbf{y}^{*\prime}\mathbf{W}\mathbf{y}^* - 2N^{-1}\mathbf{c}'\mathbf{Z}'\mathbf{W}\mathbf{y}^* + N^{-1}\mathbf{c}'\mathbf{Z}'\mathbf{W}\mathbf{Z}\mathbf{c} + \mathbf{c}'\mathbf{D}(\lambda)\mathbf{c} \right\}. \tag{33}$$

To complete the optimization, the estimators are obtained by performing a partial derivative of  $Q(\mathbf{c})$  concerning  $\mathbf{c}$  and the results are equaled to zero. The given results are

$$\widehat{\mathbf{c}} = [\mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda)]^{-1}\mathbf{Z}'\mathbf{W}\mathbf{y}^*. \tag{34}$$

By substituting  $\widehat{\mathbf{c}}$  into Equation (9), we get

$$\widehat{\mathbf{g}}_{(k,h,\lambda)}(\mathbf{x}, \mathbf{z}) = \mathbf{Z}\widehat{\mathbf{c}} = \mathbf{Z}[\mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda)]^{-1}\mathbf{Z}'\mathbf{W}\mathbf{y}^* = \mathbf{L}\mathbf{y}^*. \tag{35}$$

So, the model in Equation (15) can be written as

$$\mathbf{y}^* = \widehat{\mathbf{g}}_{(k,h,\lambda)}(\mathbf{x}, \mathbf{z}) = \mathbf{L}\mathbf{y}^*, \tag{36}$$

with  $\mathbf{y}^* = \mathbf{y} - \mathbf{f}$  and  $\mathbf{L} = \mathbf{Z}[\mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda)]^{-1}\mathbf{Z}'\mathbf{W}$ .

**Lemma 5.** *If  $f_{ji}(x_{jit})$  is approached with the truncated spline function, then the WLS is*

$$[(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\boldsymbol{\gamma}]'\mathbf{W}[(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\boldsymbol{\gamma}]. \tag{37}$$

$\mathbf{W}$  is a  $nT \times nT$  weighting matrix.

*Proof.*  $f_{ji}(x_{jit}), j = 1, 2, \dots, p$ , is a truncated spline estimator component. The component regression curve  $\mathbf{f}$  is a linear truncated spline function defined as follows:

$$f_{ji}(x_{jit}) = \alpha_{ji}x_{jit} + \sum_{u=1}^s \beta_{uji}(x_{jit} - K_{uji})_+, \tag{38}$$

with truncated function

$$(x_{jit} - K_{uji})_+ = \begin{cases} (x_{jit} - K_{uji}), & x_{jit} \geq K_{uji}, \\ 0, & x_{jit} < K_{uji}. \end{cases} \tag{39}$$

If the component of the regression curve  $\mathbf{f}$  involves one predictor, then it can be written as follows:

$$\mathbf{f} = (\mathbf{f}'_1 \mathbf{f}'_2 \dots \mathbf{f}'_n)'. \tag{40}$$

$$\mathbf{f}_i = \left( \sum_{j=1}^p f_{ji}(x_{ji1}) \sum_{j=1}^p f_{ji}(x_{ji2}) \dots \sum_{j=1}^p f_{ji}(x_{jiT}) \right)', i = 1, 2, \dots, n. \tag{41}$$

By using Equation (41), it can be described in the form of a matrix as follows:

$$\mathbf{f}_i = \begin{bmatrix} f_{ji}(x_{ji1}) \\ f_{ji}(x_{ji2}) \\ \vdots \\ f_{ji}(x_{jiT}) \end{bmatrix} = \begin{bmatrix} x_{ji1} \\ x_{ji2} \\ \vdots \\ x_{jiT} \end{bmatrix} \alpha_{ji} + \begin{bmatrix} (x_{ji1} - K_{1ji})_+^1 & (x_{ji1} - K_{2ji})_+^1 & \dots & (x_{ji1} - K_{sji})_+^1 \\ (x_{ji2} - K_{1ji})_+^1 & (x_{ji2} - K_{2ji})_+^1 & \dots & (x_{ji2} - K_{sji})_+^1 \\ \vdots & \vdots & \ddots & \vdots \\ (x_{jiT} - K_{1ji})_+^1 & (x_{jiT} - K_{2ji})_+^1 & \dots & (x_{jiT} - K_{sji})_+^1 \end{bmatrix} \begin{bmatrix} \beta_{1ji} \\ \beta_{2ji} \\ \vdots \\ \beta_{sji} \end{bmatrix} = \mathbf{X}_{ji}\alpha_{ji} + \mathbf{S}_{ji}\beta_{ji}. \tag{42}$$

The truncated spline function in the nonparametric regression component for longitudinal data with  $p$  predictors can be expressed in the following form:

$$\mathbf{f}_i = \sum_{j=1}^p [\mathbf{X}_{ji}\alpha_{ji} + \mathbf{S}_{ji}\beta_{ji}] = \mathbf{X}_i\alpha_i + \mathbf{S}_i\beta_i. \tag{43}$$

So,

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}'_1 \\ \mathbf{f}'_2 \\ \vdots \\ \mathbf{f}'_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \alpha_1 \\ \mathbf{X}_2 \alpha_2 \\ \vdots \\ \mathbf{X}_n \alpha_n \end{bmatrix} + \begin{bmatrix} \mathbf{S}_1 \beta_1 \\ \mathbf{S}_2 \beta_2 \\ \vdots \\ \mathbf{S}_n \beta_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \mathbf{X}\alpha + \mathbf{S}\beta. \quad (44)$$

Furthermore, it can be written as follows:

$$\mathbf{f} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = [\mathbf{X}|\mathbf{S}] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \mathbf{M}\gamma. \quad (45)$$

$\mathbf{M}$  is a  $nT \times (1+s)np$  matrix and  $\gamma$  is a  $(1+s)np \times 1$  vector.

The mixed model of nonparametric regression for longitudinal data in Equation (1) can be written in a matrix as follows:

$$\mathbf{y} = \mathbf{f} + \mathbf{g} + \varepsilon. \quad (46)$$

By substituting Equation (36) with Equation (46), we get

$$\mathbf{y} = \mathbf{f} + \mathbf{L}\mathbf{y}^* + \varepsilon. \quad (47)$$

The truncated spline component can be written as

$$\begin{aligned} \mathbf{y} - \mathbf{L}\mathbf{y}^* &= \mathbf{f} + \varepsilon, \\ \mathbf{y} - \mathbf{L}(\mathbf{y} - \mathbf{f}) &= \mathbf{f} + \varepsilon, \\ \mathbf{y} - \mathbf{L}\mathbf{y} + \mathbf{L}\mathbf{f} &= \mathbf{f} + \varepsilon, \\ \mathbf{y} - \mathbf{L}\mathbf{y} &= \mathbf{f} - \mathbf{L}\mathbf{f} + \varepsilon, \\ (\mathbf{I} - \mathbf{L})\mathbf{y} &= (\mathbf{I} - \mathbf{L})\mathbf{f} + \varepsilon. \end{aligned} \quad (48)$$

Substituting Equation (45) with Equation (48), we obtain

$$(\mathbf{I} - \mathbf{L})\mathbf{y} = (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma + \varepsilon. \quad (49)$$

If the function  $f_{ji}(x_{jit})$  is approximated by the truncated spline function as in Equation (38), then

$$\varepsilon = (\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma. \quad (50)$$

Therefore, we obtain the WLS by

$$\varepsilon' \varepsilon = [(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma]' \mathbf{W} [(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma], \quad (51)$$

where  $\mathbf{W}$  is a weighting matrix for the regression of longitudinal data. Next, the truncated spline estimator in the nonparametric regression for longitudinal data is derived in Theorem 6.

**Theorem 6.** *If paired data which follows the nonparametric regression model for longitudinal data is given, then the mixed estimator that minimizes WLS in Lemma 5,*

$$\min_{\gamma \in R^{(1+s)np}} \left\{ [(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma]' \mathbf{W} [(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma] \right\}, \quad (52)$$

is  $\hat{\mathbf{f}}_{(\mathbf{k}, \mathbf{h}, \lambda)}(\mathbf{x}, \mathbf{z}) = \mathbf{M}\mathbf{J}^{-1}\mathbf{K}\mathbf{y}$  with  $\mathbf{J} = (2 - \mathbf{L}')\mathbf{W}\mathbf{L}\mathbf{M} - \mathbf{W}\mathbf{M}$  and  $\mathbf{K} = [(\mathbf{L}' - \mathbf{I})\mathbf{W}(\mathbf{I} - \mathbf{L})]$ .

*Proof.* The second estimation stage of the mixed estimator of the truncated spline and Fourier series in the nonparametric regression model for longitudinal data is performed using the WLS method. The estimator can be obtained by completing the WLS optimization as follows:

$$\begin{aligned} \min_{\gamma \in R^{(1+s)np}} \left\{ [(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma]' \mathbf{W} [(\mathbf{I} - \mathbf{L})\mathbf{y} - (\mathbf{I} - \mathbf{L})\mathbf{M}\gamma] \right\} \\ = \min_{\gamma \in R^{(1+s)np}} \{Q(\gamma)\}. \end{aligned} \quad (53)$$

The estimators are obtained by performing a partial derivative of  $Q(\gamma)$  and the results are equaled to zero. The partial derivative results are as follows:

$$\hat{\gamma} = \left[ (2 - \mathbf{L}')\mathbf{W}\mathbf{L}\mathbf{M} - \mathbf{W}\mathbf{M} \right]^{-1} \left[ (\mathbf{L}' - \mathbf{I})\mathbf{W}(\mathbf{I} - \mathbf{L})\mathbf{y} \right] = \mathbf{J}^{-1}\mathbf{K}\mathbf{y}, \quad (54)$$

where  $\mathbf{J} = (2 - \mathbf{L}')\mathbf{W}\mathbf{L}\mathbf{M} - \mathbf{W}\mathbf{M}$  and  $\mathbf{K} = [(\mathbf{L}' - \mathbf{I})\mathbf{W}(\mathbf{I} - \mathbf{L})]$ .

By substituting  $\hat{\gamma}$  into the form of a truncated spline estimator component as in Equation (45), we obtain

$$\hat{\mathbf{f}}_{(\mathbf{k}, \mathbf{h}, \lambda)}(\mathbf{x}, \mathbf{z}) = \mathbf{M}\hat{\gamma} = \mathbf{M}\mathbf{J}^{-1}\mathbf{K}\mathbf{y}, \quad (55)$$

$$\hat{\mathbf{f}}_{(\mathbf{k}, \mathbf{h}, \lambda)}(\mathbf{x}, \mathbf{z}) = \mathbf{A}(\mathbf{k}, \mathbf{h}, \lambda)\mathbf{y}. \quad (56)$$

TABLE 1: Model for the simulation study.

Subject	Model
1	$y_{1t} = 8.25(x_{1t} - 1)(1 - x_{1t})^2 + \sin(2\pi z_{1t}) + \cos(2\pi z_{1t}) + \varepsilon_{1t}$
2	$y_{2t} = 9.55(x_{2t} - 1)(1 - x_{2t})^2 + \sin(2\pi z_{2t}) + \cos(2\pi z_{2t}) + \varepsilon_{2t}$
3	$y_{3t} = 6.75(x_{3t} - 1)(1 - x_{3t})^2 + \sin(2\pi z_{3t}) + \cos(2\pi z_{3t}) + \varepsilon_{3t}$
4	$y_{4t} = 3.25(x_{4t} - 1)(1 - x_{4t})^2 + \sin(2\pi z_{4t}) + \cos(2\pi z_{4t}) + \varepsilon_{4t}$
5	$y_{5t} = 5.55(x_{5t} - 1)(1 - x_{5t})^2 + \sin(2\pi z_{5t}) + \cos(2\pi z_{5t}) + \varepsilon_{5t}$
6	$y_{6t} = 7.75(x_{6t} - 1)(1 - x_{6t})^2 + \sin(2\pi z_{6t}) + \cos(2\pi z_{6t}) + \varepsilon_{6t}$
7	$y_{7t} = 4.25(x_{7t} - 1)(1 - x_{7t})^2 + \sin(2\pi z_{7t}) + \cos(2\pi z_{7t}) + \varepsilon_{7t}$
8	$y_{8t} = 8.55(x_{8t} - 1)(1 - x_{8t})^2 + \sin(2\pi z_{8t}) + \cos(2\pi z_{8t}) + \varepsilon_{8t}$
9	$y_{9t} = 9.75(x_{9t} - 1)(1 - x_{9t})^2 + \sin(2\pi z_{9t}) + \cos(2\pi z_{9t}) + \varepsilon_{9t}$
10	$y_{10t} = 4.75(x_{10t} - 1)(1 - x_{10t})^2 + \sin(2\pi z_{10t}) + \cos(2\pi z_{10t}) + \varepsilon_{10t}$

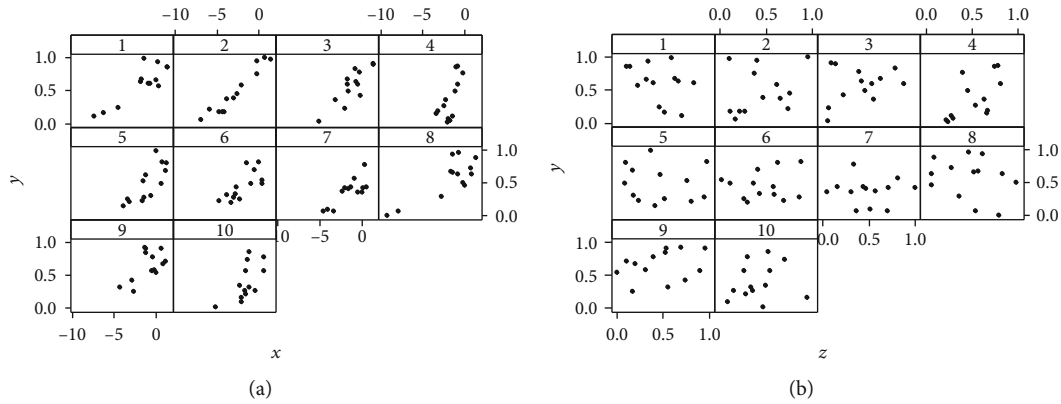


FIGURE 1: Scatterplot of simulated data for each subject and predictor: (a) truncated spline; (b) Fourier series.

After obtaining  $\hat{\gamma}$ , we obtain  $\hat{c}$  by substituting Equation (56) with Equation (34).

$$\begin{aligned}
 \hat{c} &= \left[ \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda) \right]^{-1} \mathbf{Z}'\mathbf{W}\mathbf{y}^* \\
 &= \left[ \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda) \right]^{-1} \mathbf{Z}'\mathbf{W}(\mathbf{y} - \mathbf{f}) \\
 &= \left[ \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda) \right]^{-1} \mathbf{Z}'\mathbf{W}(\mathbf{y} - \mathbf{M}\hat{\gamma}) \\
 &= \left[ \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda) \right]^{-1} \mathbf{Z}'\mathbf{W}(\mathbf{y} - \mathbf{M}\mathbf{J}^{-1}\mathbf{K}\mathbf{y}), \\
 \hat{c} &= \left[ \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda) \right]^{-1} \mathbf{Z}'\mathbf{W}(\mathbf{I} - \mathbf{M}\mathbf{J}^{-1}\mathbf{K})\mathbf{y}. \quad (57)
 \end{aligned}$$

So,

$$\begin{aligned}
 \hat{\mathbf{g}}_{(k,h,\lambda)}(\mathbf{x}, \mathbf{z}) &= \mathbf{Z}\hat{c} = \mathbf{Z} \left[ \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{N}\mathbf{D}(\lambda) \right]^{-1} \mathbf{Z}'\mathbf{W} \\
 &\quad \cdot (\mathbf{I} - \mathbf{M}\mathbf{J}^{-1}\mathbf{K})\mathbf{y} = \mathbf{L}(\mathbf{I} - \mathbf{M}\mathbf{J}^{-1}\mathbf{K})\mathbf{y}, \\
 \hat{\mathbf{g}}_{(k,h,\lambda)}(\mathbf{x}, \mathbf{z}) &= \mathbf{B}(\mathbf{k}, \mathbf{h}, \lambda)\mathbf{y}. \quad (58)
 \end{aligned}$$

By substituting  $\hat{\gamma}$  and  $\hat{c}$  into the mixed estimator of the

TABLE 2: Summary of simulation results.

Knots	Oscillation	$\lambda$	GCV
1	1	0.033	3.112
	2	1.715	3.142
	3	0.967	3.116
2	1	0.033	3.197
	2	1.800	3.209
	3	0.950	3.158
3	1	0.001	3.452
	2	0.100	3.471
	3	0.001	3.441

truncated spline and the Fourier series in the nonparametric regression for longitudinal data, the following estimation results are obtained.

$$\begin{aligned}
 \hat{\mu}_{(k,h,\lambda)} &= \mathbf{f}_{(k,h,\lambda)} + \hat{\mathbf{g}}_{(k,h,\lambda)} = \mathbf{M}\hat{\gamma} + \mathbf{Z}\hat{c} = \mathbf{M}\mathbf{J}^{-1}\mathbf{K}\mathbf{y} + \mathbf{L} \\
 &\quad \cdot (\mathbf{I} - \mathbf{M}\mathbf{J}^{-1}\mathbf{K})\mathbf{y} = [\mathbf{M}\mathbf{J}^{-1}\mathbf{K} + \mathbf{L}(\mathbf{I} - \mathbf{M}\mathbf{J}^{-1}\mathbf{K})]\mathbf{y}, \\
 \hat{\mu}_{(k,h,\lambda)} &= \mathbf{C}(\mathbf{k}, \mathbf{h}, \lambda)\mathbf{y}, \quad (59)
 \end{aligned}$$

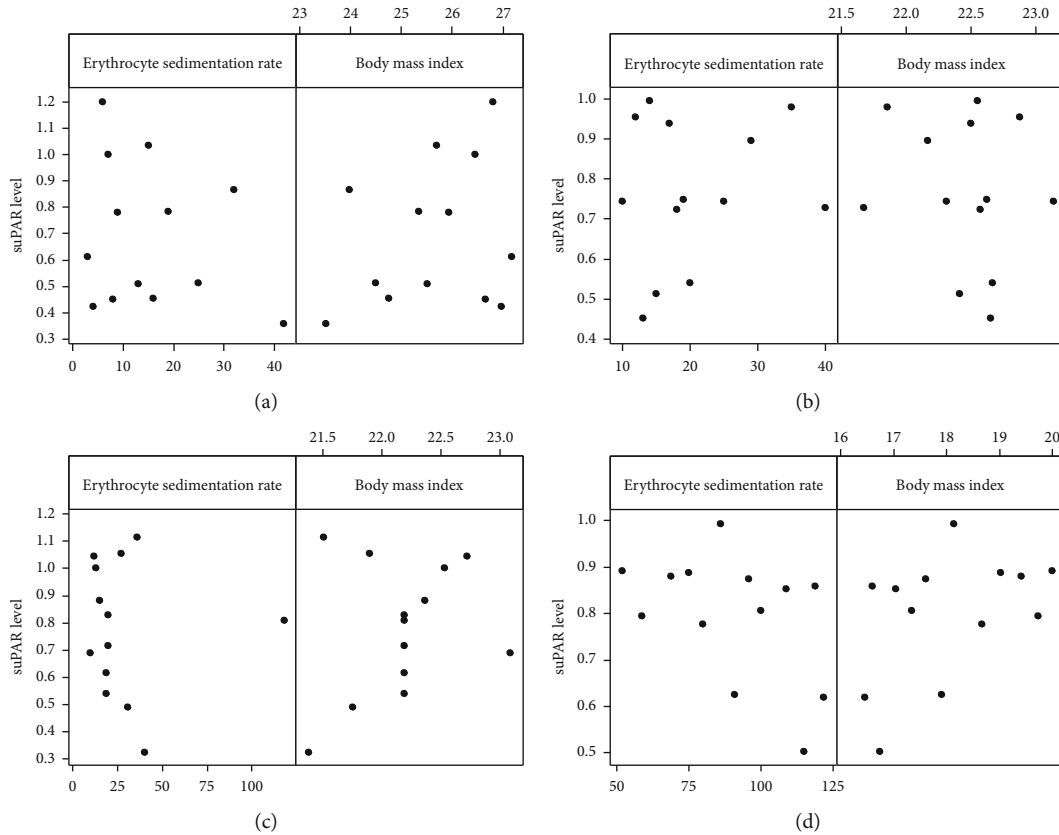


FIGURE 2: Scatterplot between suPAR levels with each predictor for the subject: (a) *minimal lesion*; (b) *mod advance*; (c) *far advance*; (d) *KP Miller*.

with  $L = Z[Z'WZ + ND(\lambda)]^{-1}Z'W$ ,  $J = (2 - L')WLM - WM$ , and  $K = [(L' - I)W(I - L)]$ .

The mixed estimator depends on the optimum knot point, oscillation parameter, and smoothing parameter. To obtain the best model, it is essential to select the optimum parameter. One of the criteria to select the optimum parameter is the generalized cross-validation (GCV) method [11]. The GCV function of the nonparametric regression model for longitudinal data is as follows:

$$GCV(K_{opt}, H_{opt}, \lambda_{opt}) = \frac{N^{-1} \|I - C(\mathbf{k}, \mathbf{h}, \lambda)\|^2}{[N^{-1} \text{trace}(I - C(\mathbf{k}, \mathbf{h}, \lambda))]^2}. \quad (60)$$

The optimum knot point, oscillation parameter, and smoothing parameter are obtained by solving the minimum optimization, as presented in Equation (60).

**3.2. Simulation Study.** To demonstrate the performance of the proposed method, we created one sample size  $n = 10$  with  $t = 13$ . For the simulation study, we considered ten models for each subject. The models are generated from the formula that contains two different functions to represent the truncated spline and Fourier series pattern. A polynomial function is used to present the truncated spline, while a trigonometry function is used to present the Fourier series. Additionally,  $x_{it}$  and  $z_{it}$  are generated from  $U(0, 1)$  distribu-

tion, and random errors  $\varepsilon_{it}$  are generated from a multivariate normal distribution.

Using ten subjects and two predictors, the formula for generated data is stated as follows:

$$y_{it} = f_{it}(x_{it}) + g_{it}(z_{it}) + \varepsilon_{it}, \quad i = 1, 2, \dots, 10, \quad t = 1, 2, \dots, 13. \quad (61)$$

The simulation study is applied based on these models, as shown in Table 1.

Figure 1 illustrates the partial relationship between the response and each predictor variable. It can be seen that the relationship between predictor  $x$  and the response for each subject tends to change at certain subintervals, which is suitable for the truncated spline estimator. The relationship between  $z$  and the response for each subject has a repetitive pattern with a particular trend line, which is suitable for the Fourier series estimator.

Wu and Zhang [7] stated that a regression's performance strongly depends on good knot locations and a good choice of the number of knots. In general, the number of knots is smaller than the sample size  $n$ . Considering the scatterplot of simulated data and computational convenience, our study uses three knots ( $K = 1, 2, 3$ ) and three oscillation parameters ( $H = 1, 2, 3$ ). To choose the optimum parameter, we use the minimum GCV criteria. Table 2 provides a summary of the GCV for varying knots and oscillations. What is remarkable



TABLE 3: Summary of case study results.

Knots	Oscillation	$\lambda$	GCV
1	1	1.250	0.239
	2	0.030	0.250
	3	0.100	0.321
2	1	0.065	0.259
	2	0.025	0.261
	3	0.975	0.266
3	1	0.025	0.259
	2	0.025	0.261
	3	1.150	0.266

is that using one knot and one oscillation with  $\lambda = 0.033$ , we obtain the best model with the lowest GCV, 3.112. This model yields satisfactory results with a root mean square error (RMSE) of 0.837.

3.3. *Case Study.* After conducting the simulation, we applied the proposed model to the case to confirm the results of the previous simulation. The data for this research was obtained from a study conducted by Fernandes and Solimun [16], that is, patients with pulmonary tuberculosis disease. Pulmonary tuberculosis is a contagious disease caused by *Mycobacterium tuberculosis*, which can attack various organs, particularly the lungs. This disease is typical among women in their productive years (ages 15-50 years). The World Health Organization (WHO) declared tuberculosis a global emergency in 1992 [17]. WHO report in 2013 stated that there were 8.6 million tuberculosis cases in 2012, of which 40% of the cases were in Southeast Asia. In a further report, Indonesia was noted as the country with the second-largest number of cases, 2.8 million, in 2015.

This study’s dataset consists of four patients ( $n = 4$ ) that represent radiological images of the thorax (stadium), which are *minimal lesion*, *mod advance*, *far advance*, and *KP Miller*. suPAR level as a response ( $y$ ) with mL units in several observation periods every two weeks for six months of treatment ( $t = 13$ ) is observed. The predictor variables are the erythrocyte sedimentation rate ( $x$ ) with mm/hour units and body mass index ( $z$ ) with  $\text{kg}/\text{m}^3$  units.

The partial relationship between the suPAR level and each predictor variable for each subject is presented in Figure 2. There were changes in data patterns in the four subjects observed for six months with measurements taken every two weeks. The plot in Figure 2 shows a different pattern for each predictor. For this reason, we propose a nonparametric regression approach based on a mixed estimator for longitudinal data. The erythrocyte sedimentation rate ( $x$ ) will be approached by a truncated spline estimator, while the body mass index ( $z$ ) will be approached by a Fourier series estimator.

In this case study, similar to the simulation study, only three knot points and three oscillation parameters were used. From the various knots and oscillation parameter results, we obtained the GCV values listed in Table 3. Interestingly, the data in this table is that the minimum GCV achieved one

TABLE 4: Knot point location for each subject.

Subject	Knot point location
<i>Minimal lesion</i>	6.184
<i>Mod advance</i>	12.449
<i>Far advance</i>	18.816
<i>KP Miller</i>	57.714

TABLE 5: The results of parameter estimation.

Subject	Parameter	Estimation
<i>Minimal lesion</i>	$\alpha_{01}$	-0.481
	$\alpha_{11}$	0.105
	$\beta_{111}$	-0.119
	$d_{11}$	0.004
	$c_{011}$	-0.849
	$c_{111}$	$5.6 \times 10^{-6}$
<i>Mod advance</i>	$\alpha_{02}$	0.441
	$\alpha_{12}$	-0.038
	$\beta_{112}$	0.041
	$d_{12}$	-0.069
	$c_{012}$	4.642
	$c_{112}$	$1.1 \times 10^{-7}$
<i>Far advance</i>	$\alpha_{03}$	-0.073
	$\alpha_{13}$	0.004
	$\beta_{113}$	-0.004
	$d_{13}$	0.008
	$c_{013}$	-1.903
	$c_{113}$	$1.4 \times 10^{-6}$
<i>KP Miller</i>	$\alpha_{04}$	-0.407
	$\alpha_{14}$	0.073
	$\beta_{114}$	-0.009
	$d_{14}$	0.058
	$c_{014}$	-0.513
	$c_{114}$	$1.6 \times 10^{-6}$

oscillation parameter under the same conditions as the simulation study, that is, one knot point. However, it has a different smoothing parameter,  $\lambda = 1.25$ . This model provides a GCV value of 0.239 with a RMSE of 0.197. The knot point location and the results of the parameter estimation for each subject of patients with pulmonary tuberculosis are presented in Tables 4 and 5, respectively.

Based on the optimal knot points in Table 3 and the parameter estimation for each subject in Table 4, the nonparametric regression model based on a mixed estimator for longitudinal data can be written as follows:

- (1) Model estimation for subject *minimal lesion*:

$$\begin{aligned}
\hat{y}_{1t} &= -0.481 + 0.105x_{11t} - 0.119(x_{11t} - 6.184)_+ \\
&\quad + 0.004z_{11t} - 0.849 + 5.6 \times 10^{-6} \cos z_{11t} \\
&= -1.33 + 0.105x_{11t} - 0.119(x_{11t} - 6.184)_+ \\
&\quad + 0.004z_{11t} + 5.6 \times 10^{-6} \cos z_{11t}
\end{aligned} \tag{62}$$

(2) Model estimation for subject *mod advance*:

$$\begin{aligned}
\hat{y}_{2t} &= 0.441 - 0.038x_{12t} + 0.041(x_{12t} - 12.449)_+ - 0.069z_{12t} \\
&\quad + 4.642 + 1.1 \times 10^{-7} \cos z_{12t} = 5.083 - 0.038x_{12t} \\
&\quad + 0.041(x_{12t} - 12.449)_+ - 0.069z_{12t} \\
&\quad + 1.1 \times 10^{-7} \cos z_{12t}
\end{aligned} \tag{63}$$

(3) Model estimation for subject *far advance*:

$$\begin{aligned}
\hat{y}_{3t} &= -0.073 + 0.004x_{13t} - 0.004(x_{13t} - 18.816)_+ \\
&\quad + 0.008z_{13t} - 1.903 + 1.4 \times 10^{-6} \cos z_{13t} \\
&= -1.976 + 0.004x_{13t} - 0.004(x_{13t} - 18.816)_+ \\
&\quad + 0.008z_{13t} + 1.4 \times 10^{-6} \cos z_{13t}
\end{aligned} \tag{64}$$

(4) Model estimation for subject *KP Miller*:

$$\begin{aligned}
\hat{y}_{4t} &= -0.407 + 0.073x_{14t} - 0.009(x_{14t} - 57.714)_+ \\
&\quad + 0.058z_{14t} - 0.513 + 1.6 \times 10^{-6} \cos z_{14t} \\
&= -0.92 + 0.073x_{14t} - 0.009(x_{14t} - 57.714)_+ \\
&\quad + 0.058z_{14t} + 1.6 \times 10^{-6} \cos z_{14t}
\end{aligned} \tag{65}$$

## 4. Conclusions

Based on the simulation study and the case study, we selected the best model by using the minimum GCV. The higher knot point or oscillation parameter does not produce a high GCV and vice versa. Therefore, we tried several combinations of knot points and the oscillation parameter to choose the best model. The result of the case study of patients suffering from pulmonary tuberculosis is similar to the simulation study. This study found that the best model uses a one knot point and one oscillation with different  $\lambda$ . It can be concluded that the simulation study supports the results of the case study.

A limitation of this study is that it does not investigate other sample sizes. Consequently, we cannot compare the performance of the developed theory for different sample sizes. Despite its limitations, the study certainly adds to our understanding of the mixed estimator's new theory in nonparametric regression for longitudinal data.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

The authors are thankful for financial support via the PMDSU Grant (Batch III) from the Ministry of Education and Culture of the Republic of Indonesia.

## References

- [1] I. N. Budiantara, V. Ratnasari, M. Ratna et al., "Modeling percentage of poor people in Indonesia using kernel and Fourier series mixed estimator in nonparametric regression," *Investigacion Operacional*, vol. 40, pp. 538–550, 2019.
- [2] W. Härdle, *Applied Nonparametric Regression (Econometric Society Monographs)*, Cambridge University Press, Cambridge, 2013.
- [3] M. Y. Cheng, R. L. Paige, S. Sun, and K. Yan, "Variance reduction for kernel estimators in clustered/longitudinal data analysis," *Journal of Statistical. Planning and Inference*, vol. 140, no. 6, pp. 1389–1397, 2010.
- [4] H. Becher, G. Kauermann, P. Khomski, and B. Kouyate, "Using penalized splines to model age- and season-of-birth-dependent effects of childhood mortality risk factors in rural Burkina Faso," *Biometrical Journal*, vol. 51, no. 1, pp. 110–122, 2009.
- [5] D. J. Best, P. J. Green, and B. W. Silverman, "Nonparametric regression and generalized linear models: a roughness penalty approach," *Biometrics*, vol. 50, no. 4, p. 1228, 1994.
- [6] B. Lestari, I. N. Budiantara, S. Sunaryo, and M. Mashuri, "Spline smoothing for multi-response nonparametric regression model in case of heteroscedasticity of variance," *Journal of Mathematics and Statistics*, vol. 8, no. 3, pp. 377–384, 2012.
- [7] H. Wu and J. T. Zhang, *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*, Wiley Series in Probability and Statistics, 2006.
- [8] M. Bilodeau, "Fourier smoother and additive models," *The Canadian Journal of Statistics*, vol. 3, pp. 257–269, 1992.
- [9] D. D. Cox and F. O'Sullivan, "Penalized likelihood-type estimators for generalized nonparametric regression," *Journal of Multivariate Analysis*, vol. 56, no. 2, pp. 185–206, 1996.
- [10] S. N. Wood, "On confidence intervals for generalized additive models based on penalized regression splines," *Australian and New Zealand Journal of Statistics*, vol. 48, no. 4, pp. 445–464, 2006.
- [11] G. Wahba, *Spline models for observational data*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1990.
- [12] N. Wang, "Marginal nonparametric kernel regression accounting for within-subject correlation," *Biometrika*, vol. 90, no. 1, pp. 43–52, 2003.

- [13] C. O. Wu and C. T. Chiang, "Kernel smoothing on varying coefficient models with longitudinal dependent variable," *Statistica Sinica*, vol. 10, no. 2, pp. 433–456, 2000.
- [14] H. Zhang, "Multivariate adaptive splines for analysis of longitudinal data," *Journal of Computational and Graphical Statistics*, vol. 6, no. 1, pp. 74–91, 1997.
- [15] I. W. Sudiarsa, I. N. Budiantara, and S. W. P. Suhartono, "Combined estimator Fourier series and spline truncated in multivariable nonparametric regression," *Applied Mathematical Sciences*, vol. 9, pp. 4997–5010, 2015.
- [16] A. A. R. Fernandes and Solimun, "Modeling of multi-response longitudinal data using linear mixed modeling patient with pulmonary tuberculosis in Syaiful Anwar hospital Malang," *International Journal of PharmTech Research*, vol. 9, pp. 478–486, 2016.
- [17] P. D. P. Indonesia, *Tuberkulosis: Pedoman Diagnosis dan Penatalaksanaan di Indonesia*, Indah Offset Citra Grafika, Jakarta, 2006.