Mixture Spline Smoothing and Kernel Estimator in Multi-Response Nonparametric Regression

Dyah Putri Rahmawati, I Nyoman Budiantara, Dedy Dwi Prastyo, and Made Ayu Dwi Octavanny

Abstract—**In previous research about multi-response nonparametric regression models, each predictor variable is considered to have the same pattern concerning each response variable. In contrast, multi-response cases are often encountered with different patterns among the predictor variables. Therefore, a mixture estimator in multi-response nonparametric regression needs to be developed. This study proposes an additive mixture of Spline Smoothing and Kernel estimator in multi-response nonparametric regression. Our approach can handle the previously mentioned issue in a multi-response nonparametric regression problem, i.e., some predictors showing changing patterns in certain sub-intervals, such as Spline Smoothing patterns, and other predictors exhibiting random patterns, commonly modeled using Kernel regression. A two-stage estimation procedure, i.e., Penalized Weighted Least Square followed by Weighted Least Square, was used to obtain this mixture estimator. Furthermore, a simulation study and real data analysis were conducted to illustrate the performance of the proposed multi-response mixture estimator. The results indicate that the proposed multi-response mixture estimator can be applied appropriately and gives satisfactory results with a coefficient of determination (R^2) close to 1 and a Mean Absolute Percentage Error (MAPE) of less than 5%.**

Index Terms—**kernel, mixture estimator, multi-response, nonparametric regression, spline smoothing**

I. Introduction

REGRESSION analysis is a popular statistical method for predicting. Regression analysis is often used in many fields to determine the functional relationship between the predictor variables and the response variables [1]. Nonparametric regression is one type of regression analysis that is suitable for cases in which the shape of the regression curve is unknown or when there is no information about the shape of the curve between the response variable and the predictor variable. This regression has high flexibility where the data can drive to estimate of the regression curve without subjectivity from the researcher. Among several nonparametric regression estimators, Spline Smoothing is an estimator that has an excellent statistical and visual interpretation. It has a remarkable ability to handle data with patterns that change at certain sub-intervals using smooth functions [2]. There are several studies on the Spline Smoothing estimator in nonparametric regression, including Eubank [2] , Wang et al. [3], Lestari et al. [4], Aydin et al. [5], Fernandes et al. [6], and Diana [7]. In addition, the Kernel estimator in nonparametric regression has been extensively developed by many researchers, including Gasser and Muller [8], Hall and Huang [9], Okumura and Naito [10], Dharmasena et al. [11], Du et al. [12], Chamidah and Saifudin [13], and Erçelik and Nadar [14]. Data with random patterns are commonly modeled using Kernel regression [15].

In the development of nonparametric regression research, modeling using multiple response variables (multi-response) has been proposed by several researchers, including Wang et al. [3], who proposed bi-response Spline Smoothing; Lestari et al. [4], who developed a multi-response Spline Smoothing estimator on cross-section data; and Fernandes et al. [6], who proposed bi-response Spline Smoothing estimator on longitudinal data. All of these use the Penalized Weighted Least Square (PWLS) estimation method. Chamidah and Saifudin [13] developed a multi-response Kernel estimator using the Weighted Least Square (WLS) estimation method. However, these studies only proposed one type of nonparametric regression estimator. In several multi-response cases, there are often different patterns among the predictor variables. Thus, modeling with a mixture estimator in multi-response nonparametric regression is needed to handle these issues.

Until now, modeling with mixture estimators has only been proposed for models with one response (see [15]–[19]). One of them is mixture Spline Smoothing and Kernel estimators by Hidayat et al. [17]. Therefore, in this study, we proposed a new theory about the mixture of Spline Smoothing and Kernel estimator for multi-response nonparametric regression. This proposed multi-response mixture estimator is the development of a mixture estimator by Hidayat et al. [17]. In addition to developing into the multi-response model, our mixture estimator also estimates the Kernel component, whereas the mixture estimator by Hidayat et al. [17] still considers Kernel components to be fixed. This mixture estimator is obtained through two stages of estimation, i.e., the first stage using the PWLS estimation method and the second stage employing the WLS estimation

Manuscript received December 25, 2020 revised May 18, 2021. This work was supported by Ministry of Research, Technology and Higher Education (Ristekdikti), Republic of Indonesia through PMDSU scholarship.

D. P. Rahmawati is a Doctoral student of the Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia (e-mail: dyah.18062@mhs.its.ac.id).

I. N. Budiantara is a Professor of the Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia (corresponding author to provide e-mail: i_nyoman_b@statistika.its.ac.id).

D. D. Prastyo is a Lecturer of the Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia (e-mail: dedy-dp@statistika.its.ac.id).

M. A. D. Octavanny is a Doctoral student of the Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia (e-mail: octavanny@gmail.com).

method.

Furthermore, a simulation study and real data analysis are conducted to illustrate the performance of the proposed multi-response mixture estimator. The simulation data were generated from a formula that contains two different functions (polynomial and exponential) to represent two different patterns between the predictors and responses. This mixture estimator is also applied to the three dimensions of the Human Development Index (HDI), namely Life Expectancy Index (LEI), Education Index (EI), and Gross National Income Index (GNII) in the Papua Province. The predictor variables are population growth rate, dependency ratio, percentage of the population working in agriculture, and economic growth rate. This multi-response case has several predictor patterns that tend to change in certain sub-intervals and other predictor patterns tend to be random.

The rest of this paper is organized as follows. In Section II.A, we review the spline estimator and the Kernel estimator in nonparametric regression. In Section II.B, we introduced the two-stage estimation method, i.e., the PWLS and followed by the WLS. In Section III.A, we present our proposed mixture Spline Smoothing and Kernel estimator in multi-response nonparametric regression. Section III.B describes the selection of Smoothing and Bandwidth Parameters to obtain the best model from this proposed multi-response mixture estimator. In Section III.C and III.D, we conduct a simulation study and real data analysis to illustrate the performance of the proposed multi-response mixture estimator. The last section presents conclusion and further research.

II. Materials and Methods

A. Spline Smoothing and Kernel Estimator in Nonparametric Regression

A regression model that states the relationship between one predictor and one response is as follows:

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

where y_i is the response variable, $\mu(x_i)$ is the regression curve, and ε_i is the random error that is assumed to be normally distributed, identical, and independent with mean 0 and variance σ^2. If the form of the regression function $\mu(x_i)$ in (1) is unknown, the appropriate type of regression approach is nonparametric regression [15]. Several nonparametric regression models can be used to estimate the regression function, including Spline Smoothing and Kernel.

The Spline estimator is applied appropriately if the pattern of the data tends to change at certain sub-intervals. In addition, the spline estimator can produce flexible and smooth curves. The smoothness of the spline curve is determined by its smoothing parameter [20]. The Spline Smoothing regression function is assumed to be smooth and contained in the Sobolev space or stated as

$$\mu \in W_2^m[a,b] \text{ with } W_2^m[a,b] = \left\{ \mu : \int_{a_j}^{b_j} \left[\mu^{(m)}(x) \right]^2 dx < \infty \right\},$$

where m is the order of the polynomial Spline. The Spline Smoothing is obtained from minimizing the Penalized Least Square (PLS). The PLS formula is an estimation criterion that combines the goodness-of-fit and penalty functions [2]:

$$\underset{\mu \in W_2^m(a,b)}{Min} \left\{ n^{-1} \sum_{i=1}^{n} \left(y_i - \mu(x_i) \right)^2 + \lambda \int_a^b \left(\mu^{(m)}(x_i) \right)^2 dx \right\}. \tag{2}$$

The first component in (2) is the function that measures data suitability (goodness of fit) like the Ordinary Least Square, and the second component is the penalty component for a function $\mu(x_i)$. The penalty component is the determination of the roughness of the curve with λ as the smoothing parameter. The smoothing parameter has a vital role in determining the smoothness of the function. Accordingly, this approach is known as PLS.

Unlike the Spline Smoothing estimator, the Kernel estimator is commonly used to model data with a seemingly random pattern. The Kernel approach depends on the bandwidth value, which functions to control the smoothness of the estimated function. The selection of the optimal bandwidth is crucial for the Kernel estimator [21]. If the regression function $\mu(x_i)$ in (1) is unknown and is approached by the Kernel estimator, then the function $\mu(x_i)$ can be approximated by the Taylor series with order m as follows [13], [22]:

$$\mu(x) \approx \mu(x_0) + \frac{(x - x_0)^1 \mu^1(x_0)}{1!} + \frac{(x - x_0)^2 \mu^2(x_0)}{2!} + ... + \frac{(x - x_0)^m \mu^m(x_0)}{m!}. \tag{3}$$

If $\tau_v(x_0) = \dfrac{\mu^v(x_0)}{v!}, v = 1, 2, ..., m$, then (3) can be stated as

$$\mu(x) \approx \tau_0(x_0) + (x - x_0)^1 \tau_1(x_0) + (x - x_0)^2 \tau_2(x_0) + ... + (x - x_0)^m \tau_m(x_0). \tag{4}$$

The Kernel estimator is obtained when polynomial order $m = 0$ [22]; then, (4) with $m = 0$ and involving all observations can be expressed as

$$\boldsymbol{\mu}(x) \approx \boldsymbol{\tau}_0(x_0). \tag{5}$$

The Kernel estimator can be obtained through the WLS estimation method as follows:

$$\underset{\alpha}{Min} \left\{ (\boldsymbol{y} - \boldsymbol{\mu})^T \mathbf{K}_\alpha(x_0) (\boldsymbol{y} - \boldsymbol{\mu}) \right\}, \tag{6}$$

where $\mathbf{K}_\alpha(x_0)$ is the weighted matrix from the Kernel function, with α is the bandwidth parameter.

B. The Two-Stage Estimation Method

This study proposes a new theory about the mixture of Spline Smoothing and Kernel estimator for multi-response nonparametric regression. This mixture estimator is obtained through two stages of estimation, with the first stage using the PWLS estimation method and the second stage employing the WLS estimation method. The two-stage estimation method to obtain our proposed mixture estimator is described as follows.

If given the data pairs $(y_{1i}, y_{2i}, ..., y_{ri}, x_{1i}, x_{2i}, ..., x_{pi}, t_{1i}, t_{2i}, ..., t_{qi})$, which, following additive multi-response nonparametric regression, model as follows:

$$y_{hi} = \omega_{hi}(x_{ji}, t_{ki}) + \varepsilon_{hi},\ \ h = 1, 2;\ i = 1, 2, ..., n; \quad j = 1, 2, ..., p; k = 1, 2, ..., q, \tag{7}$$

where $\omega_{hi}(x_{ji}, t_{ki})$ is a regression curve in which the function form is assumed to be unknown, with some predictor variables having patterns that change at certain sub-intervals and the remaining predictors typically having random patterns. This regression curve is an additive function, so it can be written as follows [23]:

$$\omega_{hi} = \sum_{j=1}^{p} g_{hj}(x_{ji}) + \sum_{k=1}^{q} f_{hk}(t_{ki}),\ h = 1, 2, ..., r,\ i = 1, 2, ..., n, \quad (8)$$

where $\sum_{j=1}^{p} g_{hj}(x_{ji})$ is a regression curve component with p predictor variables and $\sum_{k=1}^{q} f_{hk}(t_{ki})$ is the regression curve component with q predictor variables. The ε_{hi} in (7) is a random error that is assumed to be a normal distribution of N-variate ($N = nr$) with zero mean and variance–covariance matrix $\mathbf{\Sigma}$. In the multi-response cases where y_{hi} and $y_{h'i}$ are paired, the h-response and the h'-response error $(h \neq h')$ correlate. The error correlation between responses can be stated as follows [3]:

$$\text{corr}(\varepsilon_{hi}, \varepsilon_{h'i}) = \rho;\ \rho = \frac{\text{cov}(\varepsilon_{hi}, \varepsilon_{h'i})}{\sigma_h \sigma_{h'}};\ i = 1, 2, ..., n;$$
$$h = 1, 2, ..., r;\ h \neq h'.$$

In vector notation, (8) can be expressed as

$$\boldsymbol{\omega} = \boldsymbol{g} + \boldsymbol{f}. \quad (9)$$

Therefore, (7) can be written in vector form as follows:

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{f} + \boldsymbol{\varepsilon}. \quad (10)$$

The regression curve $\boldsymbol{g}$ is approximated by the Spline Smoothing function. This regression curve is assumed to be smooth and contained in the Sobolev space $\boldsymbol{g} \in W_2^m[a,b]$ [24]. Meanwhile, the regression curve $\boldsymbol{f}$ is approximated by the Kernel function. This multi-response mixture Spline Smoothing and Kernel estimator can be obtained by a two-stage estimation method. The Spline Smoothing component $(\boldsymbol{g})$ is estimated by PWLS in the first stage of estimation, and the Kernel component $(\boldsymbol{f})$ is estimated by the WLS method in the second stage of the estimation procedure. Before estimating the Spline Smoothing component, we need to modify (10) into the following form:

$$\boldsymbol{h} = \boldsymbol{g} + \boldsymbol{\varepsilon}, \quad (11)$$

with $\boldsymbol{h} = \boldsymbol{y} - \boldsymbol{f}$.

The estimation of Spline Smoothing components $\hat{\boldsymbol{g}}$ can be obtained using the PWLS optimization formula of (11) as follows:

$$\underset{g_h \in W_2^m(a_h, b_h)}{Min} \Bigg\{ N^{-1} (\boldsymbol{h} - \boldsymbol{g})^T \mathbf{\Sigma}^{-1} (\boldsymbol{h} - \boldsymbol{g}) +$$
$$\sum_{h=1}^{r} \lambda_h \int_{a_h}^{b_h} \left[g_h^{(m)}(x) \right]^2 dx \Bigg\} 0 < \lambda_h < \infty, \quad (12)$$

where λ_h is the smoothing parameter of the Spline Smoothing estimator for each response. The estimation results from the first stage are substituted into (10).

In the second stage of estimation, the estimation for the Kernel component $\hat{\boldsymbol{f}}$ can be obtained using the WLS optimization as follows:

$$\underset{\alpha_h}{Min} \left\{ (\boldsymbol{y} - \hat{\boldsymbol{g}} - \boldsymbol{f})^T \mathbf{\Sigma}^{-1} \mathbf{K}_{\boldsymbol{\alpha}}(t_0) (\boldsymbol{y} - \hat{\boldsymbol{g}} - \boldsymbol{f}) \right\}, \quad (13)$$

where α_h is the bandwidth parameter of the Kernel estimator for each response and $\mathbf{K}_{\boldsymbol{\alpha}}(t_0)$ is the weighted matrix for the Kernel estimator, whereas $\mathbf{\Sigma}^{-1}$ in (12) and (13) is the weighted matrix in multi-response nonparametric regression formed from the variance–covariance matrix of error [3]. By substituting the results from the two-stage estimation into (9), the mixture of Spline Smoothing and Kernel estimator in multi-response nonparametric regression can be obtained.

III. Results and Discussion

A. Estimation of Mixture Spline Smoothing and Kernel Estimator in Multi-Response Nonparametric Regression

The function form of the regression curves $\boldsymbol{g}$ and $\boldsymbol{f}$ in (9) is unknown. Therefore, before making the two-stage estimation, we must obtain the function form of these regression curves. The function form of the regression curve $\boldsymbol{g}$ is presented in Lemma 1, whereas the function form of the regression curve $\boldsymbol{f}$ is shown in Lemma 2.

Lemma 1: Suppose a given function $\sum_{j=1}^{p} g_{hj}(x_{ji})$ is expressed in vector notation $\boldsymbol{g}$. This function is assumed to be smooth and contained in the Sobolev space $\boldsymbol{g} \in W_2^m[a,b]$; then, this function form is as follows:

$$\boldsymbol{g} = \mathbf{U}\mathbf{d} + \mathbf{T}\mathbf{c}, \quad (14)$$

where

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{U}_r \end{pmatrix},\ \mathbf{d} = \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_r \end{pmatrix},$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_r \end{pmatrix},\ \mathbf{c} = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_r \end{pmatrix},$$

with $\mathbf{U}_h = \mathbf{U}_{h1}^* + \mathbf{U}_{h2}^* + ... + \mathbf{U}_{hp}^*$,

$\mathbf{T}_h = \mathbf{T}_{h1}^* + \mathbf{T}_{h2}^* + ... + \mathbf{T}_{hp}^*;\ h = 1, 2, ..., r.$

Proof: Suppose $g_h^*; h = 1, 2, ..., r$ is a function with one predictor variable. This function is contained in the Hilbert space W. The Hilbert space can be decomposed into a direct sum of two spaces W_0 and W_1 as $W = W_0 \oplus W_1$, with $W_0 \perp W_1$. If $\{\theta_{h1}, \theta_{h2}, ..., \theta_{hm}\}$ is the basis in W_0 and $\{\psi_{h1}, \psi_{h2}, ..., \psi_{hn}\}$ is the basis in W_1, then for each function $g_h^* \in W$, $u_h \in W_0$ and $v_h \in W_1$ can be described as follows:

$$g_h^* = u_h + v_h$$
$$= \sum_{l=1}^{m} d_{hl}\theta_{hl} + \sum_{i=1}^{n} c_{hi}\psi_{hi}$$

$$=\boldsymbol{\theta}_h^T\mathbf{d}_h+\boldsymbol{\psi}_h^T\mathbf{c}_h\,. \qquad (15)$$

where $\boldsymbol{\theta}_h=\begin{pmatrix}\theta_{h1}\\ \theta_{h2}\\ \vdots\\ \theta_{hm}\end{pmatrix}, \mathbf{d}_h=\begin{pmatrix}d_{h1}\\ d_{h2}\\ \vdots\\ d_{hm}\end{pmatrix}, \boldsymbol{\psi}_h=\begin{pmatrix}\psi_{h1}\\ \psi_{h2}\\ \vdots\\ \psi_{hn}\end{pmatrix}, \mathbf{c}_h=\begin{pmatrix}c_{h1}\\ c_{h2}\\ \vdots\\ c_{hn}\end{pmatrix}$

with d_{hl} and c_{hi} are any constants.

The equation (15) is a limited linear function in W; therefore, it can be stated in the inner product as follows:

$$L_x g_h^*=\langle \eta_{hi}, g_h^*\rangle$$
$$=g_h^*(x_i),\ g_h^*\in W\,. \qquad (16)$$

Based on (15), (16) can be written as

$$g_h^*(x_i)=\langle \eta_{hi}, g_h^*\rangle$$
$$=\langle \eta_{hi}, \boldsymbol{\theta}_h^T\mathbf{d}_h+\boldsymbol{\psi}_h^T\mathbf{c}_h\rangle$$
$$=\langle \eta_{hi}, \boldsymbol{\theta}_h^T\mathbf{d}_h\rangle+\langle \eta_{hi}, \boldsymbol{\psi}_h^T\mathbf{c}_h\rangle$$
$$=\left\langle \eta_{hi}, \begin{pmatrix}\theta_{h1} & \theta_{h2} & \cdots & \theta_{hm}\end{pmatrix}\begin{pmatrix}d_{h1}\\ d_{h2}\\ \vdots\\ d_{hm}\end{pmatrix}\right\rangle+$$
$$\left\langle \eta_{hi}, \begin{pmatrix}\psi_{h1} & \psi_{h2} & \cdots & \psi_{hn}\end{pmatrix}\begin{pmatrix}c_{h1}\\ c_{h2}\\ \vdots\\ c_{hn}\end{pmatrix}\right\rangle$$
$$=d_{h1}\langle \eta_{hi},\theta_{h1}\rangle+\ldots+d_{hm}\langle \eta_{hi},\theta_{hm}\rangle+$$
$$c_{h1}\langle \eta_{hi},\psi_{h1}\rangle+\ldots+c_{hn}\langle \eta_{hi},\psi_{hn}\rangle, \qquad (17)$$

for each response $(h=1,2,\ldots,r)$ with all observations $(i=1,2,\ldots,n)$ (17) can be stated as follows:

$$\boldsymbol{g}_h^*=\begin{pmatrix} d_{h1}\langle \eta_{h1},\theta_{h1}\rangle+\ldots+d_{hm}\langle \eta_{h1},\theta_{hm}\rangle+ \\ d_{h1}\langle \eta_{h2},\theta_{h1}\rangle+\ldots+d_{hm}\langle \eta_{h2},\theta_{hm}\rangle+ \\ \vdots \\ d_{h1}\langle \eta_{hn},\theta_{h1}\rangle+\ldots+d_{hm}\langle \eta_{hn},\theta_{hm}\rangle+ \\ c_{h1}\langle \eta_{h1},\psi_{h1}\rangle+\ldots+c_{hn}\langle \eta_{h1},\psi_{hn}\rangle \\ c_{h1}\langle \eta_{h2},\psi_{h1}\rangle+\ldots+c_{hn}\langle \eta_{h2},\psi_{hn}\rangle \\ \vdots \\ c_{h1}\langle \eta_{hn},\psi_{h1}\rangle+\ldots+c_{hn}\langle \eta_{hn},\psi_{hn}\rangle \end{pmatrix},$$

$$\boldsymbol{g}_h^*=\mathbf{U}_h^*\mathbf{d}_h+\mathbf{T}_h^*\mathbf{c}_h\,, \qquad (18)$$

with

$$\mathbf{U}_h^*=\begin{pmatrix}\langle \eta_{h1},\theta_{h1}\rangle & \langle \eta_{h1},\theta_{h2}\rangle & \cdots & \langle \eta_{h1},\theta_{hm}\rangle\\ \langle \eta_{h2},\theta_{h1}\rangle & \langle \eta_{h2},\theta_{h2}\rangle & \cdots & \langle \eta_{h2},\theta_{hm}\rangle\\ \vdots & \vdots & \ddots & \vdots\\ \langle \eta_{hn},\theta_{h1}\rangle & \langle \eta_{hn},\theta_{h2}\rangle & \cdots & \langle \eta_{hn},\theta_{hm}\rangle\end{pmatrix}, \mathbf{d}_h=\begin{pmatrix}d_{h1}\\ d_{h2}\\ \vdots\\ d_{hm}\end{pmatrix},$$

$$\mathbf{T}_h^*=\begin{pmatrix}\langle \eta_{h1},\psi_{h1}\rangle & \langle \eta_{h1},\psi_{h2}\rangle & \cdots & \langle \eta_{h1},\psi_{hn}\rangle\\ \langle \eta_{h2},\psi_{h1}\rangle & \langle \eta_{h2},\psi_{h2}\rangle & \cdots & \langle \eta_{h2},\psi_{hn}\rangle\\ \vdots & \vdots & \ddots & \vdots\\ \langle \eta_{hn},\psi_{h1}\rangle & \langle \eta_{hn},\psi_{h2}\rangle & \cdots & \langle \eta_{hn},\psi_{hn}\rangle\end{pmatrix}, \mathbf{c}_h=\begin{pmatrix}c_{h1}\\ c_{h2}\\ \vdots\\ c_{hn}\end{pmatrix}.$$

The components of matrix $\mathbf{U}_h^*$ can be described as $\langle \eta_{hi},\theta_{hs}\rangle=\dfrac{x_i^{s-1}}{(s-1)!},\ i=1,2,\ldots,n;\ h=1,2,\ldots,r;\ s=1,2,\ldots,m,$ whereas $\langle \eta_{hi},\psi_{hi}\rangle$ as components of the matrix $\mathbf{T}_h^*$ are equal to $\langle \psi_{hi},\psi_{hi^*}\rangle$, where

$$\langle \psi_{hi},\psi_{hi^*}\rangle=\int_a^b\frac{(x_i-u)_+^{m-1}(x_{i^*}-u)_+^{m-1}}{((m-1)!)^2}\,du,\quad i=1,2,\ldots,n;$$

$i^*=1,2,\ldots,n;\ \ h=1,2,\ldots,r$.

If the regression curve $\boldsymbol{g}$ in (9) is a function with r responses and p predictors, then it can be stated as follows:

$$\boldsymbol{g}=\begin{pmatrix}\boldsymbol{g}_1\\ \boldsymbol{g}_2\\ \vdots\\ \boldsymbol{g}_r\end{pmatrix}. \qquad (19)$$

The function for each response with p predictors $(\boldsymbol{g}_h; h=1,2,\ldots,r)$ is the modification from (18) as follows:

$$\boldsymbol{g}_h=\sum_{j=1}^{p}\boldsymbol{g}_{hj}^*$$
$$=\sum_{j=1}^{p}\left(\mathbf{U}_{hj}^*\mathbf{d}_h+\mathbf{T}_{hj}^*\mathbf{c}_h\right)$$
$$=\mathbf{U}_{h1}^*\mathbf{d}_h+\mathbf{U}_{h2}^*\mathbf{d}_h+\ldots+\mathbf{U}_{hp}^*\mathbf{d}_h+\mathbf{T}_{h1}^*\mathbf{c}_h+$$
$$\mathbf{T}_{h2}^*\mathbf{c}_h+\ldots+\mathbf{T}_{hp}^*\mathbf{c}_h. \qquad (20)$$

For all responses $(h=1,2,\ldots,r)$, we obtain the form of the regression curve $\boldsymbol{g}$ as follows:

$$\boldsymbol{g}=\begin{pmatrix}\mathbf{U}_{11}^*\mathbf{d}_1+\mathbf{U}_{12}^*\mathbf{d}_1+\ldots+\mathbf{U}_{1p}^*\mathbf{d}_1+\mathbf{T}_{11}^*\mathbf{c}_1+\mathbf{T}_{12}^*\mathbf{c}_1+\ldots+\mathbf{T}_{1p}^*\mathbf{c}_1\\ \mathbf{U}_{21}^*\mathbf{d}_2+\mathbf{U}_{22}^*\mathbf{d}_2+\ldots+\mathbf{U}_{2p}^*\mathbf{d}_2+\mathbf{T}_{21}^*\mathbf{c}_2+\mathbf{T}_{22}^*\mathbf{c}_2+\ldots+\mathbf{T}_{2p}^*\mathbf{c}_2\\ \vdots\\ \mathbf{U}_{r1}^*\mathbf{d}_r+\mathbf{U}_{r2}^*\mathbf{d}_r+\ldots+\mathbf{U}_{rp}^*\mathbf{d}_r+\mathbf{T}_{r1}^*\mathbf{c}_r+\mathbf{T}_{r2}^*\mathbf{c}_r+\ldots+\mathbf{T}_{rp}^*\mathbf{c}_r\end{pmatrix}$$

$$\boldsymbol{g}=\begin{pmatrix}\mathbf{U}_1 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{U}_2 & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{U}_r\end{pmatrix}\begin{pmatrix}\mathbf{d}_1\\ \mathbf{d}_2\\ \vdots\\ \mathbf{d}_r\end{pmatrix}+$$
$$\begin{pmatrix}\mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{T}_2 & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_r\end{pmatrix}\begin{pmatrix}\mathbf{c}_1\\ \mathbf{c}_2\\ \vdots\\ \mathbf{c}_r\end{pmatrix}. \qquad (21)$$

In matrix notation, (21) can be expressed as

$$\boldsymbol{g}=\mathbf{U}\mathbf{d}+\mathbf{T}\mathbf{c}\,,\ \blacksquare \qquad (22)$$

where $\mathbf{U}_h=\mathbf{U}_{h1}^*+\mathbf{U}_{h2}^*+\ldots+\mathbf{U}_{hp}^*$ and $\mathbf{T}_h=\mathbf{T}_{h1}^*+\mathbf{T}_{h2}^*+\ldots+\mathbf{T}_{hp}^*$, the components of matrix $\mathbf{U}_{hj}^*$ can be described as $\langle \eta_{hji},\theta_{hjs}\rangle=\dfrac{x_{ji}^{s-1}}{(s-1)!},\ i=1,2,\ldots,n;$ $h=1,2,\ldots,r;\ \ j=1,2,\ldots,p;\ \ s=1,2,\ldots,m$, whereas the components of the matrix $\mathbf{T}_{hj}^*$ can be described as

$$\langle \eta_{hji},\psi_{hji}\rangle=\langle \psi_{hji},\psi_{hji^*}\rangle=\int_a^b\frac{(x_{ji}-u)_+^{m-1}(x_{ji^*}-u)_+^{m-1}}{((m-1)!)^2}\,du,$$

$i=1,2,...,n;\; i^{*}=1,2,...,n;\;\; j=1,2,...,p;\;\; h=1,2,...,r.$

Lemma 2: Suppose the given function $\sum_{k=1}^{q} f_{hk}(t_{ki})$ is expressed in vector notation $\boldsymbol{f}$. The function $\boldsymbol{f}$ is approached by the Kernel estimator, then the form function of $\boldsymbol{f}$ can be expressed as

$$\boldsymbol{f}=\boldsymbol{\tau}_0(t_0), \tag{23}$$

where

$$\boldsymbol{\tau}_0(t_0)=\left[\left(\sum_{k=1}^{q}\boldsymbol{\tau}_{1k,0}(t_{k0})\right)^T \quad \left(\sum_{k=1}^{q}\boldsymbol{\tau}_{2k,0}(t_{k0})\right)^T \quad \cdots \quad \left(\sum_{k=1}^{q}\boldsymbol{\tau}_{rk,0}(t_{k0})\right)^T\right]^T.$$

Proof: The function $f_{hk}(t_{ki})$ is part of the function $\sum_{k=1}^{q} f_{hk}(t_{ki})$ for $h=1,2,...,r.$ This function can be approached by the Taylor series with t_k around t_{k0} as follows [13], [22]:

$$f_{hk}(t_k)\approx f_{hk}(t_{k0})+\frac{(t_k-t_{k0})^1 f_{hk}^1(t_{k0})}{1!}+\frac{(t_k-t_{k0})^2 f_{hk}^2(t_{k0})}{2!}+...+\frac{(t_k-t_{k0})^{m_h} f_{hk}^{m_h}(t_{k0})}{m_h!}. \tag{24}$$

If $\tau_{hk,v}(t_{k0})=\frac{f_{hk}^{v}(t_{k0})}{v!}$, $v=0,1,...,m_h$, then (24) can be stated as follows:

$$f_{hk}(t_k)\approx \tau_{hk,0}(t_{k0})+(t_k-t_{k0})^1\tau_{hk,1}(t_{k0})+(t_k-t_{k0})^2\tau_{hk,2}(t_{k0})+...+(t_k-t_{k0})^{m_h}\tau_{hk,m_h}(t_{k0}), \tag{25}$$

with $t_k\in(t_{k0}-\alpha_h, t_{k0}+\alpha_h)$, where t_{k0} is the value of predictor variable for prediction and α_h is bandwidth parameter for each response.

The Kernel estimator is obtained when polynomial order $m_h=0$; then, (25) with $m_h=0$ can be stated as

$$f_{hk}(t_k)\approx \tau_{hk,0}(t_{k0}). \tag{26}$$

By involving all observations, the function form for each response can be stated as

$$\begin{aligned}\boldsymbol{f}_{hk}(\mathbf{t}_k)&=\left(f_{hk}\left(t_{k(1)}\right)\quad f_{hk}\left(t_{k(2)}\right)\quad\cdots\quad f_{hk}\left(t_{k(n)}\right)\right)^T\\&=\left(\tau_{hk,0}\left(t_{k0(1)}\right)\quad \tau_{hk,0}\left(t_{k0(2)}\right)\quad\cdots\quad \tau_{hk,0}\left(t_{k0(n)}\right)\right)^T\\&=\boldsymbol{\tau}_{hk,0}\left(t_{k0}\right).\end{aligned}$$

Therefore, by involving all responses, we can obtain

$$\boldsymbol{f}_k(\mathbf{t}_k)=\boldsymbol{\tau}_{k,0}(t_{k0}), \tag{27}$$

where

$$\boldsymbol{f}_k(\mathbf{t}_k)=\left(\left(\boldsymbol{f}_{1k}(\mathbf{t}_k)\right)^T\quad\left(\boldsymbol{f}_{2k}(\mathbf{t}_k)\right)^T\quad\cdots\quad\left(\boldsymbol{f}_{rk}(\mathbf{t}_k)\right)^T\right)^T,$$

$$\boldsymbol{\tau}_{k,0}(t_{k0})=\left(\boldsymbol{\tau}_{1k,0}(t_{k0})\quad\boldsymbol{\tau}_{2k,0}(t_{k0})\quad\cdots\quad\boldsymbol{\tau}_{rk,0}(t_{k0})\right)^T.$$

Therefore, the function form of the regression function $\sum_{k=1}^{q} f_{hk}(t_{ki})$ can be stated in vector notation as

$$\boldsymbol{f}=\sum_{k=1}^{q}\boldsymbol{\tau}_{k,0}(t_{k0})=\boldsymbol{\tau}_0(t_0),\ \blacksquare \tag{28}$$

where

$$\boldsymbol{\tau}_0(t_0)=\left[\left(\sum_{k=1}^{q}\boldsymbol{\tau}_{1k,0}(t_{k0})\right)^T \quad \left(\sum_{k=1}^{q}\boldsymbol{\tau}_{2k,0}(t_{k0})\right)^T \quad \cdots \quad \left(\sum_{k=1}^{q}\boldsymbol{\tau}_{rk,0}(t_{k0})\right)^T\right]^T.$$

Theorem 1: The model of multi-response nonparametric regression is given in (7), and each component of the regression curve is additive. The form of the function $\boldsymbol{g}$ is presented in Lemma 1, and the form of the function $\boldsymbol{f}$ is presented in Lemma 2. Using the Penalized Weighted Least Square estimation method in the first stage of estimation and Weighted Least Square estimation method in the second stage of estimation, the mixture of Spline Smoothing and Kernel estimator in multi-response nonparametric regression is obtained as follows:

$$\hat{\boldsymbol{\omega}}=\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})\,\boldsymbol{y}, \tag{29}$$

where

$$\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})=\left(\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})+\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right),$$

$$\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})=\mathbf{U}\left(\mathbf{R}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\boldsymbol{\Sigma}^{-1}+\mathbf{T}\mathbf{R}^{-1}\boldsymbol{\Sigma}^{-1}\left(\mathbf{I}-\mathbf{U}\left(\left(\mathbf{R}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\boldsymbol{\Sigma}^{-1}\right)\right),$$

$$\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})=\left(\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})^T\left(\mathbf{D}\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})-2\mathbf{D}\right)+\mathbf{D}\right)^{-1}\left(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})^T\right)\mathbf{D}\left(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right).$$

Proof: To obtain the mixture of Spline Smoothing and Kernel estimator in multi-response nonparametric regression, a two-stage estimation is needed. The first stage of estimation is carried out to estimate the Spline Smoothing estimator component $(\hat{\boldsymbol{g}})$ using the PLWS method. By substituting the function form of $\boldsymbol{g}$ in (22) into (12), the PWLS formula can be stated as follows:

$$\underset{g_h\in W_2^m(a_h,b_h)}{Min}\left\{N^{-1}\left(\boldsymbol{h}-\mathbf{U}\mathbf{d}-\mathbf{T}\mathbf{c}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{h}-\mathbf{U}\mathbf{d}-\mathbf{T}\mathbf{c}\right)+\sum_{h=1}^{r}\lambda_h\int_{a_h}^{b_h}\left[g_h^{(m)}(x)\right]^2dx\right\},0<\lambda_h<\infty. \tag{30}$$

The penalty component $\sum_{h=1}^{r}\lambda_h\int_{a_h}^{b_h}\left[g_h^{(m)}(x)\right]^2dx$ in (30) needs to be elaborated. This penalty component can be obtained through the decomposition as follows:

$$\begin{aligned}\int_{a_h}^{b_h}\left[g_h^{(m)}(x)\right]^2dx&=\left\|P_1g_h\right\|^2\\&=\left\langle P_1g_h,P_1g_h\right\rangle\\&=\left\langle P_1\left(\boldsymbol{\theta}_h^T\mathbf{d}_h+\boldsymbol{\psi}_h^T\mathbf{c}_h\right),P_1\left(\boldsymbol{\theta}_h^T\mathbf{d}_h+\boldsymbol{\psi}_h^T\mathbf{c}_h\right)\right\rangle\\&=\left\langle\boldsymbol{\psi}_h^T\mathbf{c}_h,\boldsymbol{\psi}_h^T\mathbf{c}_h\right\rangle\\&=\mathbf{c}_h^T\left(\boldsymbol{\psi}_h\boldsymbol{\psi}_h^T\right)\mathbf{c}_h\\&=\mathbf{c}_h^T\mathbf{T}_h\mathbf{c}_h.\end{aligned} \tag{31}$$

P_1 is the orthogonal projection of g_h to W_1 in W space. By substituting (31) into the penalty component, we obtain

$$\sum_{h=1}^{r}\lambda_h\int_{a_h}^{b_h}\left[g_h^{(m)}(x)\right]^2dx=\sum_{h=1}^{r}\lambda_h\mathbf{c}_h^T\mathbf{T}_h\mathbf{c}_h$$

$$=\mathbf{c}_1^T\lambda_1\mathbf{T}_1\mathbf{c}_1+\mathbf{c}_2^T\lambda_2\mathbf{T}_2\mathbf{c}_2+\ldots+\mathbf{c}_r^T\lambda_r\mathbf{T}_r\mathbf{c}_r$$

$$=\begin{pmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_r\end{pmatrix}\begin{pmatrix}\lambda_1\mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \lambda_2\mathbf{I}_n & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \lambda_r\mathbf{I}_n\end{pmatrix}$$

$$\begin{pmatrix}\mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{T}_2 & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_r\end{pmatrix}\begin{pmatrix}\mathbf{c}_1\\ \mathbf{c}_2\\ \vdots\\ \mathbf{c}_r\end{pmatrix}$$

$$=\mathbf{c}^T\mathbf{\Lambda}\mathbf{T}\mathbf{c}.\ [6] \tag{32}$$

By substituting (32) into (30), the PWLS optimization formula can be written in matrix notation as follows:

$$\underset{g_h\in W_2^m(a_h,b_h)}{Min}\left\{N^{-1}\left(\boldsymbol{h}-\mathbf{U}\mathbf{d}-\mathbf{T}\mathbf{c}\right)^T\mathbf{\Sigma}^{-1}\left(\boldsymbol{h}-\mathbf{U}\mathbf{d}-\mathbf{T}\mathbf{c}\right)+\mathbf{c}^T\mathbf{\Lambda}\mathbf{T}\mathbf{c}\right\}=$$

$$\underset{g_h\in W_2^m(a_h,b_h)}{Min}\left\{Q\left(\mathbf{c},\mathbf{d}\right)\right\}. \tag{33}$$

Using partial derivative $Q(\mathbf{c},\mathbf{d})$ by $\mathbf{c}$ and $\mathbf{d}$, the solution for the PWLS optimization is obtained. The partial derivative of $Q(\mathbf{c},\mathbf{d})$ by $\mathbf{c}$ is as follows:

$$\frac{\partial Q(\mathbf{c},\mathbf{d})}{\partial\mathbf{c}}=-2\mathbf{T}^T\mathbf{\Sigma}^{-1}\boldsymbol{h}+2\mathbf{T}^T\mathbf{\Sigma}^{-1}\mathbf{U}\mathbf{d}+$$

$$2\mathbf{T}^T\mathbf{\Sigma}^{-1}\mathbf{T}\hat{\mathbf{c}}+2N\mathbf{\Lambda}\mathbf{T}\hat{\mathbf{c}}=\mathbf{0}. \tag{34}$$

Suppose $\mathbf{R}=\mathbf{\Sigma}^{-1}\mathbf{T}+N\mathbf{\Lambda}$; we obtain

$$\hat{\mathbf{c}}=\mathbf{R}^{-1}\left(\mathbf{\Sigma}^{-1}\boldsymbol{h}-\mathbf{\Sigma}^{-1}\mathbf{U}\mathbf{d}\right). \tag{35}$$

The partial derivative of $Q(\mathbf{c},\mathbf{d})$ by $\mathbf{d}$ is as follows:

$$\frac{\partial Q(\mathbf{c},\mathbf{d})}{\partial\mathbf{d}}=-2\mathbf{U}^T\mathbf{\Sigma}^{-1}\boldsymbol{h}+2\mathbf{U}^T\mathbf{\Sigma}^{-1}\mathbf{U}\hat{\mathbf{d}}+2\mathbf{U}^T\mathbf{\Sigma}^{-1}\mathbf{T}\hat{\mathbf{c}}=\mathbf{0}. \tag{36}$$

By substituting (35) into (36), we obtain

$$-2\mathbf{U}^T\mathbf{\Sigma}^{-1}\boldsymbol{h}+2\mathbf{U}^T\mathbf{\Sigma}^{-1}\mathbf{U}\hat{\mathbf{d}}+$$

$$2\mathbf{U}^T\mathbf{\Sigma}^{-1}\mathbf{T}\mathbf{R}^{-1}\left(\mathbf{\Sigma}^{-1}\boldsymbol{h}-\mathbf{\Sigma}^{-1}\mathbf{U}\hat{\mathbf{d}}\right)=\mathbf{0}. \tag{37}$$

Recalling that $\mathbf{R}=\mathbf{\Sigma}^{-1}\mathbf{T}+N\mathbf{\Lambda}$, we obtain $\mathbf{T}=\mathbf{\Sigma}(\mathbf{R}-N\mathbf{\Lambda})$. Next, we can modify the following part of (37):

$$\mathbf{\Sigma}^{-1}\mathbf{T}\mathbf{R}^{-1}=\mathbf{\Sigma}^{-1}\left(\mathbf{\Sigma}(\mathbf{R}-N\mathbf{\Lambda})\right)\mathbf{R}^{-1},$$

$$=\mathrm{I}-N\mathbf{\Lambda}\mathbf{R}^{-1}. \tag{38}$$

By substituting (38) into (37) and solving it, we obtain

$$\hat{\mathbf{d}}=\left(\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\boldsymbol{h}. \tag{39}$$

Then, (39) is substituted into (35), and we get

$$\hat{\mathbf{c}}=\mathbf{R}^{-1}\left(\mathbf{\Sigma}^{-1}\boldsymbol{h}-\mathbf{\Sigma}^{-1}\mathbf{U}\left(\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\boldsymbol{h}\right), \tag{40}$$

$$\hat{\mathbf{c}}=\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\left(\mathbf{I}-\mathbf{U}\left(\left(\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\right)\right)\boldsymbol{h}. \tag{41}$$

The results of $\hat{\mathbf{d}}$ and $\hat{\mathbf{c}}$ are substituted into the form of function $\boldsymbol{g}$ in (22). The Spline Smoothing estimator component in the multi-response nonparametric regression model can be obtained as follows:

$$\hat{\boldsymbol{g}}=\mathbf{U}\hat{\mathbf{d}}+\mathbf{T}\hat{\mathbf{c}}$$

$$=\Big[\mathbf{U}\left(\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}+\mathbf{T}\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}$$

$$\left(\mathbf{I}-\mathbf{U}\left(\left(\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}\right)^{-1}\mathbf{R}^{-1}\mathbf{\Sigma}^{-1}\right)\right)\Big]\boldsymbol{h}$$

$$=\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{h}. \tag{42}$$

Recalling that $\boldsymbol{h}=\boldsymbol{y}-\boldsymbol{f}$, the result of the estimation of Spline Smoothing estimator component from the first stage estimation can be stated as follows:

$$\hat{\boldsymbol{g}}=\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})(\boldsymbol{y}-\boldsymbol{f})=\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{f}. \tag{43}$$

Therefore, we obtain the estimation results of the modified model in (11) as follows:

$$\hat{\boldsymbol{h}}=\hat{\boldsymbol{g}},$$

$$\hat{\boldsymbol{h}}=\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{f}. \tag{44}$$

However, this result still contains function $\boldsymbol{f}$ that needs to be estimated in the second stage of estimation.

After obtaining the estimated Spline Smoothing component, we carry out the second stage of estimation. In the second stage of estimation, the function $\boldsymbol{f}$ as the Kernel estimator component is estimated using the WLS method. By substituting the results of the first stage of estimation (44) and the result of Lemma 2 (28) into the WLS optimization formula (13), we can obtain

$$\underset{\alpha_h}{Min}\left\{\left(\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}+\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{\tau}_0(t_0)-\boldsymbol{\tau}_0(t_0)\right)^T\right.$$

$$\left.\mathbf{\Sigma}^{-1}\mathbf{K}_{\boldsymbol{\alpha}}(t_0)\left(\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}+\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{\tau}_0(t_0)-\boldsymbol{\tau}_0(t_0)\right)\right\}, \tag{45}$$

where $\mathbf{K}_{\boldsymbol{\alpha}}(t_0)$ is the weighted matrix for the Kernel estimator with the following structure:

$$\mathbf{K}_{\boldsymbol{\alpha}}(t_{k0})=diag\left(\sum_{k=1}^{q}\mathbf{K}_{k\alpha_1}(t_{ki}-t_{k0}),\right.$$

$$\left.\sum_{k=1}^{q}\mathbf{K}_{k\alpha_2}(t_{ki}-t_{k0}),\ldots,\sum_{k=1}^{q}\mathbf{K}_{k\alpha_r}(t_{ki}-t_{k0})\right), \tag{46}$$

where $\mathbf{K}_{k\alpha_h}(t_{ki}-t_{k0})=\begin{pmatrix}K_{k\alpha_h}(t_{k1}-t_{k0}) & K_{k\alpha_h}(t_{k2}-t_{k0}) & \cdots\end{pmatrix}$ $\left.K_{k\alpha_h}(t_{kn}-t_{k0})\right)^T$; $h=1,2,\ldots,r$; $k=1,2,\ldots,q$ and $K_{k\alpha_h}(.)$ is the Kernel function.

Supposing $\mathbf{D}=\mathbf{\Sigma}^{-1}\mathbf{K}_{\boldsymbol{\alpha}}(t_0)$, (45) can be written as

$$\underset{\alpha_h}{Min}\left\{\left(\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}+\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{\tau}_0(t_0)-\boldsymbol{\tau}_0(t_0)\right)^T\mathbf{D}\right.$$

$$\left.\left(\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}+\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{\tau}_0(t_0)-\boldsymbol{\tau}_0(t_0)\right)\right\}$$

$$=\underset{\alpha_h}{Min}\left\{Q(t_0)\right\}. \tag{47}$$

The optimization solution above is obtained by performing the partial derivative $Q(t_0)$ by $\boldsymbol{\tau}_0(t_0)$, and the result is equal to zero. The result is obtained as follows:

$$\hat{\boldsymbol{\tau}}_0(t_0)=\left(\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})^T\left(\mathbf{D}\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})-2\mathbf{D}\right)+\mathbf{D}\right)^{-1}$$

$$\left(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})^T\right)\mathbf{D}\left(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)\boldsymbol{y}. \tag{48}$$

Therefore, the estimation result for the Kernel estimator component $\hat{\boldsymbol{f}}$ can be written as follows:

$$\hat{\boldsymbol{f}}=\hat{\boldsymbol{\tau}}_0(t_0)$$

$$=\left(\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})^T\left(\mathbf{D}\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})-2\mathbf{D}\right)+\mathbf{D}\right)^{-1}\left(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})^T\right)$$
$$\mathbf{D}\left(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)\boldsymbol{y}$$
$$=\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}. \quad (49)$$

By substituting the first-stage estimation results in (41) and the second-stage estimation results in (49) into (9), we can obtain the following mixture Spline Smoothing and Kernel estimator in multi-response nonparametric regression:

$$\hat{\boldsymbol{\omega}}=\hat{\boldsymbol{g}}+\hat{\boldsymbol{f}}$$
$$=\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\hat{\boldsymbol{f}}+\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}$$
$$=\left(\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})-\mathbf{A}(\boldsymbol{\lambda},\boldsymbol{\alpha})\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})+\mathbf{B}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)\boldsymbol{y}$$
$$=\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})\boldsymbol{y}. \blacksquare \quad (50)$$

B. Selection of Smoothing and Bandwidth Parameters

In nonparametric regression with the Spline Smoothing estimator and the Kernel estimator, there are smoothing parameters that play a role in regulating the smoothness of the estimated regression function. In the mixture of Spline Smoothing and Kernel estimator in multi-response nonparametric regression, there are two types of smoothing parameters: the smoothing parameters $\boldsymbol{\lambda}=(\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_r)^T$ as smoothing parameters for the Spline Smoothing component and bandwidth parameters $\boldsymbol{\alpha}=(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_r)^T$ as a smoothing parameter for the estimator Kernel component. The Smoothing parameters $(\boldsymbol{\lambda})$ and bandwidth $(\boldsymbol{\alpha})$ are tuning parameters.

One of the methods to determine the optimal smoothing and bandwidth parameters in nonparametric regression is Generalized Cross Validation (GCV) [25]. The advantage of the GCV method compared to other methods is that the GCV method has optimal asymptotic properties [24]. In the multi-response mixture Spline Smoothing and Kernel estimator, the smallest GCV value is used to determine the best model with the optimal smoothing parameters $(\boldsymbol{\lambda}_{opt})$ and bandwidth parameters $(\boldsymbol{\alpha}_{opt})$. Based on the mixture Spline Smoothing and Kernel estimator in multi-response nonparametric regression in (29), we can obtain the following GCV formula:

$$GCV(\boldsymbol{\lambda},\boldsymbol{\alpha})=\frac{MSE\left(\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)}{\left(N^{-1}trace\left(\mathbf{I}-\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)\right)^2}, \quad (51)$$

where

$$MSE\left(\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)=N^{-1}(\boldsymbol{y}-\hat{\boldsymbol{\omega}})^T(\boldsymbol{y}-\hat{\boldsymbol{\omega}})$$
$$=N^{-1}\left\|\left(\mathbf{I}-\mathbf{C}(\boldsymbol{\lambda},\boldsymbol{\alpha})\right)\boldsymbol{y}\right\|^2.$$

C. Simulation Study

In this study, a simulation study was conducted to show the performance of the proposed mixture of Spline Smoothing and Kernel estimator in multi-response nonparametric regression (29). The simulation data were generated from a formula that contains two different nonlinear functions. One function is a polynomial function $(g_{hi}(x_i))$ and the other is an exponential function $(f_{hi}(t_i))$. The polynomial function is used to generate data that match the Smoothing Spline pattern, and the exponential function is used to generate data that were suitable for the Kernel estimator. The formula for generating simulation data with three response variables and two predictor variables is as follows:

$$y_{hi}=g_{hi}(x_i)+f_{hi}(t_i)+\varepsilon_{hi},\ h=1,2,3;\ i=1,2,...,n, \quad (52)$$

where

$$g_{1i}(x_i)=4.73(x_i-1)(1-x_i)^2,$$
$$f_{1i}(t_i)=2.34\left(e^{-4.03t_i}-4e^{-7.13t_i}+3e^{-9.76t_i}\right),$$
$$g_{2i}(x_i)=6.89(x_i-1)(1-x_i)^2,$$
$$f_{2i}(t_i)=-0.23\left(e^{-2.36t_i}-4e^{-5.74t_i}+3e^{-4.23t_i}\right),$$
$$g_{3i}(x_i)=-7.53(x_i-1)(1-x_i)^2,$$
$$f_{3i}(t_i)=0.79\left(e^{-2.94t_i}-4e^{-4.82t_i}+3e^{-8.34t_i}\right),$$

the predictors are generated from $x_i \sim U(0,2)$ and $t_i \sim U(0,2)$ with a sample size $n=100$, and the random errors ε_{hi} are generated from bivariate normal distributions with $\mu_1=0,\ \mu_2=0, \mu_3=0,\ \sigma_1^2=0.2,\ \sigma_2^2=0.25, \sigma_3^2=0.3$, and $\rho=0.5$. Gaussian Kernel is employed in this simulation. Fig. 1 shows the scatter plots from the simulation data. The plots between y_1, y_2, y_3 and x tend to change at certain subintervals like the Spline Smoothing pattern, whereas the plots between y_1, y_2, y_3 and t are like random patterns that are commonly modeled with Kernel regression. In this simulation, we use the Gaussian Kernel.

The selection of the best model for the proposed estimator was carried out using minimum GCV criteria. The higher smoothing parameters and bandwidth parameters do not necessarily result in high GCV and vice versa. Therefore, we need to try several combinations of smoothing parameters and bandwidth parameter values to choose the best model. Using two-stage estimation and tuning parameters, we can obtain a set of optimal smoothing and bandwidth parameters for the simulation data. Table 1 shows some combinations of smoothing and bandwidth parameters around their optimal values. Due to limited space, we only show a few of all parameter combinations. From the combinations of smoothing and bandwidth parameters in Table 1 and the plot of GCV value in Fig. 2, the smallest GCV value is 0.2395237. The model with the smallest GCV value is chosen as the best model for these simulation data. This model is obtained from optimal smoothing parameters $\lambda_{1(opt)}=0.000725,\ \lambda_{2(opt)}=0.000395,$ $\lambda_{3(opt)}=0.000751$ and optimal bandwidth parameters $\alpha_{1(opt)}=0.0001937, \alpha_{2(opt)}=0.0003563, \alpha_{2(opt)}=0.0001875$. This model produces $R^2=95.83\%$, and $MSE=0.2066491$.

Fig. 3 presents the 3D scatterplots between the estimation results and the original simulation data for each response, where the estimated values (gray dots) are very close to the original simulation data (black dots). These show that the proposed model and estimation procedure can be used to make a prediction correctly.

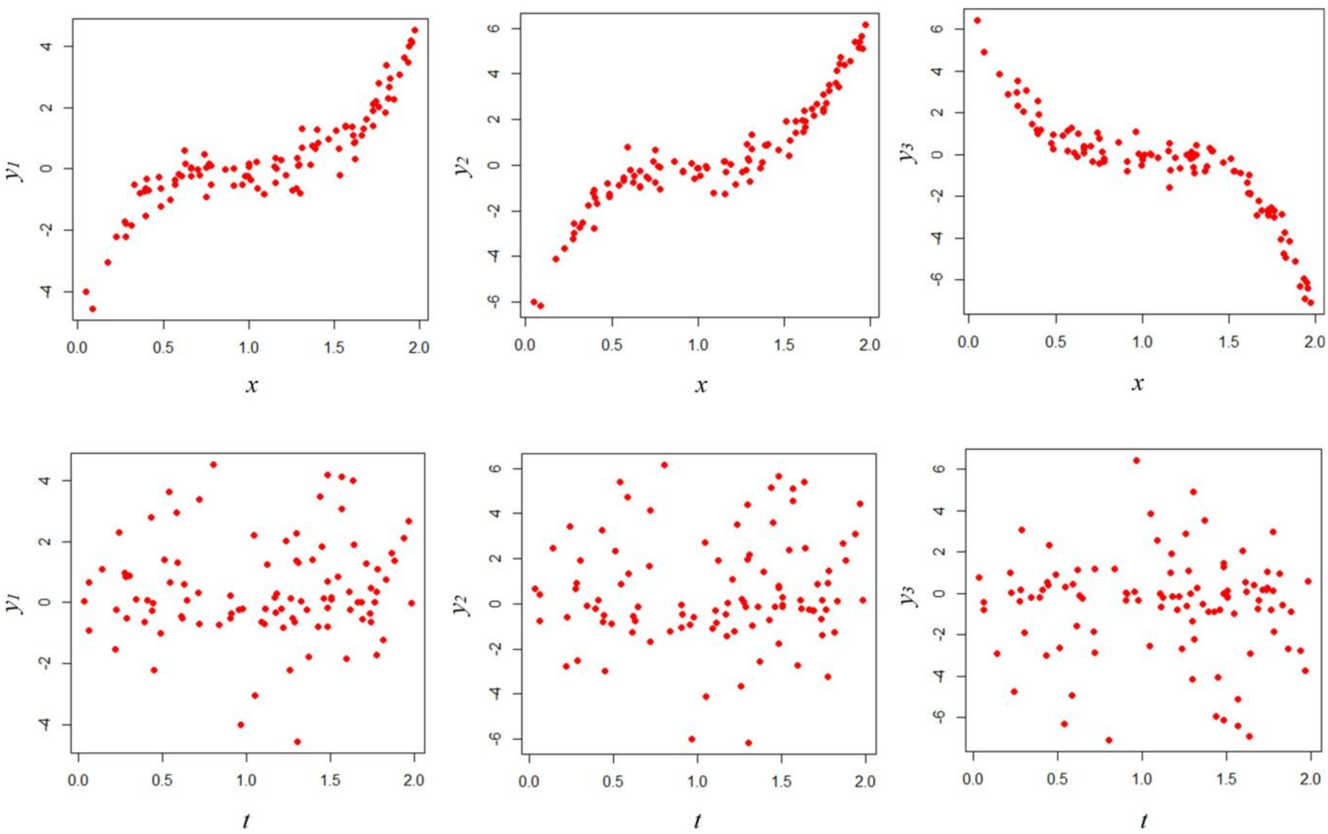

Fig. 1. Scatterplot of the simulation data.

TABLE I
COMBINATIONS OF SMOOTHING AND BANDWIDTH PARAMETERS AROUND THEIR OPTIMAL VALUES

λ_1	λ_2	λ_3	α_1	α_2	α_3	GCV
0.0000193	0.0000356	0.0000187	0.0000725	0.0000395	0.0000751	0.24747
0.0000193	0.0000356	0.0000187	0.000725	0.000395	0.000751	0.24513
0.000193	0.000356	0.000187	0.0000725	0.0000395	0.0000751	0.24103
0.000193	**0.000356**	**0.000187**	**0.000725**	**0.000395**	**0.000751**	**0.23952**
0.000193	0.000356	0.000187	0.00725	0.00395	0.00751	0.2498
0.00193	0.00356	0.00187	0.000725	0.000395	0.000751	0.32076
0.00193	0.00356	0.00187	0.00725	0.00395	0.00751	0.35638

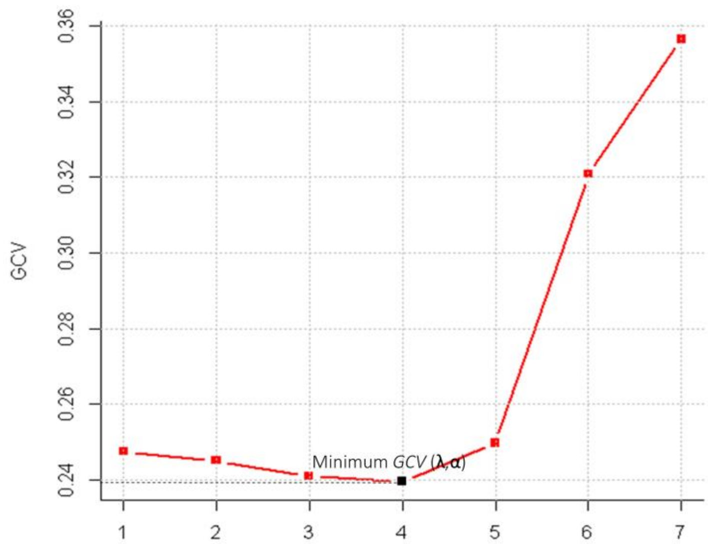

Fig. 2. Plot of Generalized Cross Validation (GCV) value.

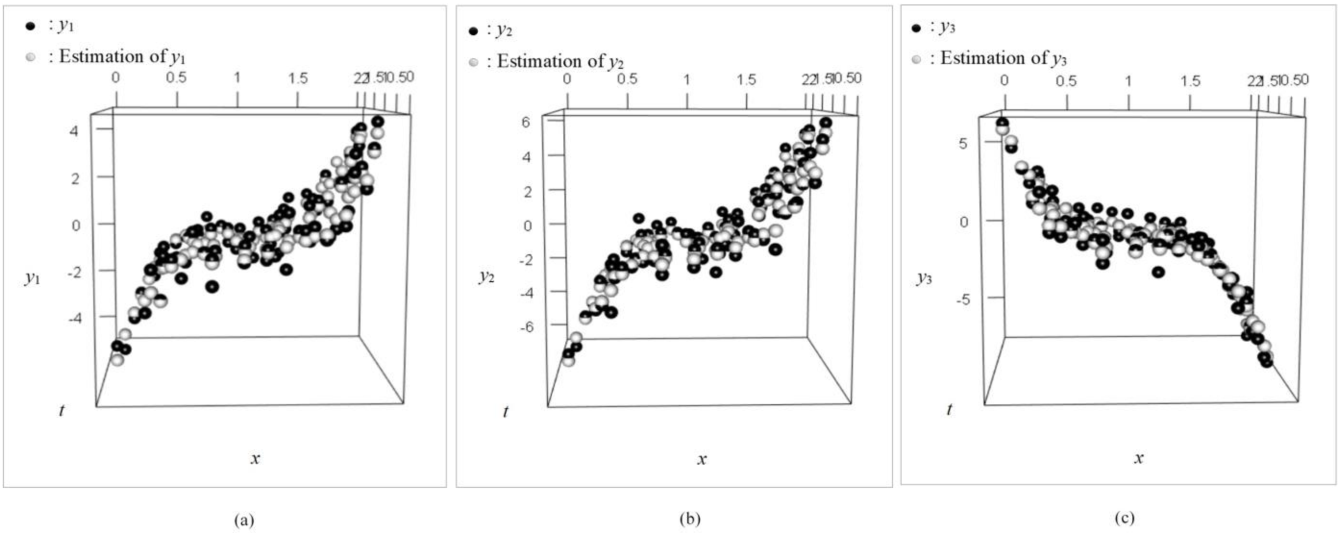

Fig. 3. 3D scatterplots between the estimation results and the actual simulation data.

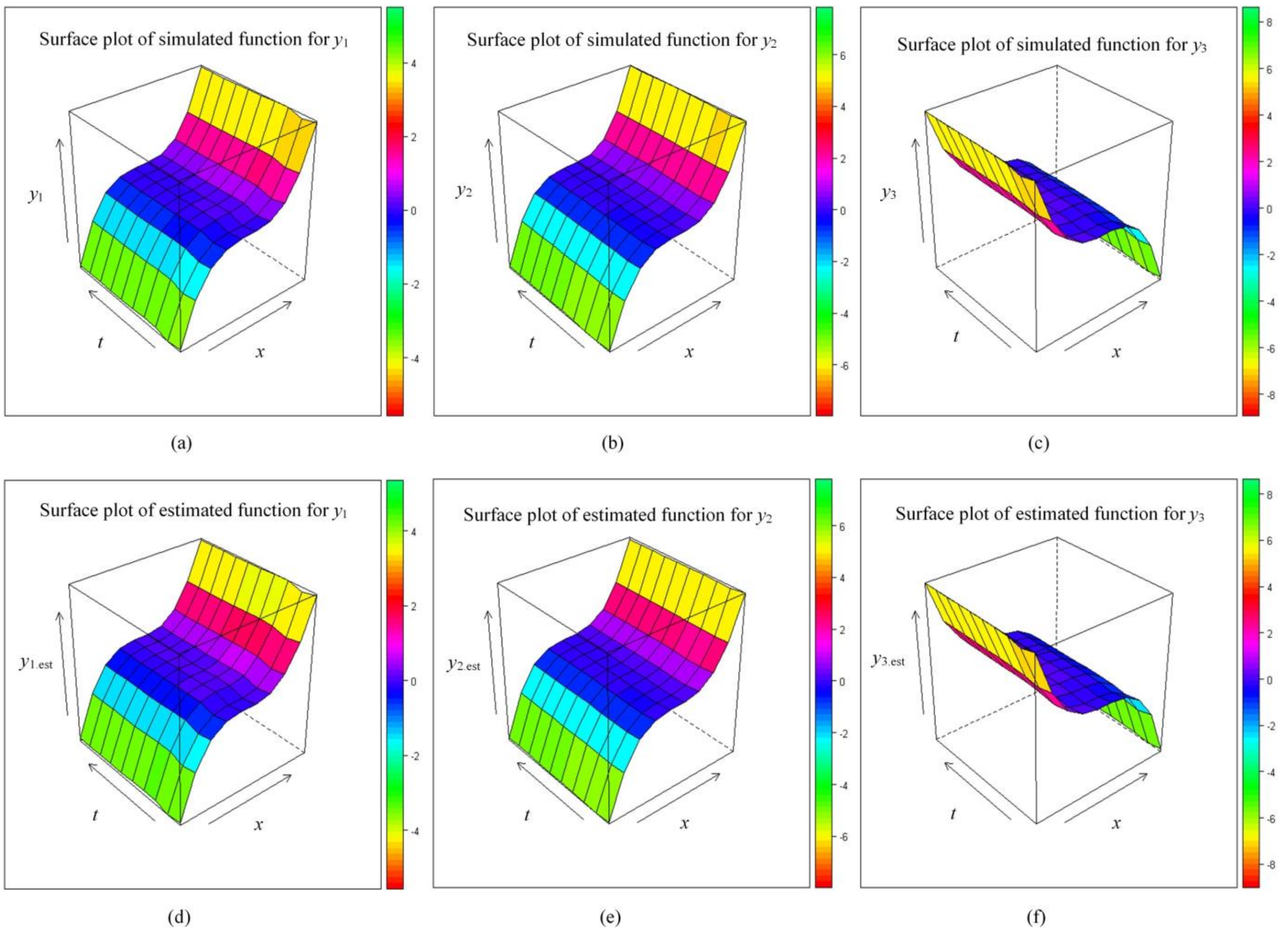

Fig. 4. 3D surface plots of the function for simulation (a,b,c) and its estimator (d,e,f).

TABLE II
COMPARISON OF THREE KINDS OF ESTIMATORS

Estimator	Minimum GCV
Mixture Spline Smoothing and Kernel	**0.2395**
Spline Smoothing	0.3237
Kernel	1.1493

Furthermore, Fig. 4 (a, b, c) shows the surface plots from the functions for generating simulation data (52). In comparison, Fig. 4 (d, e, f) is formed from (29), which is obtained from two-stage estimation, i.e., PWLS and WLS. The two sides of Fig. 4 display plots that appear to have similar surface shape. This evidence shows that the estimation procedure proposed in (29) can be used appropriately to estimate the function generated from the simulation.

We also compare the modeling on simulation data with our proposed multi-response mixture estimator against the multi-response Spline Smoothing estimator only and the multi-response Kernel estimator only. The minimum GCV values of these models are presented in Table 2. The model using our proposed estimator produces the smallest minimum GCV values. This empirical finding suggests that modeling for data with different patterns gives better results when using a mixture estimator.

D. Application Study

In this section, we discuss the results of applying the proposed multi-response mixture estimator in real cases. The data used are secondary data obtained from the Statistics Indonesia of Papua Province. They consist of Y_1: Life Expectancy Index (LEI), Y_2: Education Index (EI), and Y_3: Gross National Income Index (GNII) as three response variables that represent the three dimensions of the Human Development Index (HDI) with four predictor variables that influence it, namely X_1: the population growth rate, X_2: dependency ratio, X_3: percentage of population working in agriculture, and X_4: economic growth rate in the Papua Province. The observation units used were 29 districts/cities in the Papua Province in 2018.

In the case of multi-response, there is an assumption that there must be a correlation between the response variables. According to Rahayu et al. [26], the three response variables that represent the three dimensions of the Human Development Index are correlated so that the data are feasible to be modeled with multi-response regression. To ensure that the three response variables satisfy the correlation assumptions for multi-response modeling, we perform a correlation test with the Bartlett sphericity test. The test gives the chi-square value $X^2 = 41.7576$ with $p-value = 5\times10^{-9}$. The $p-value$ of the Bartlett Sphericity test is smaller than $\alpha = 5\%$, so it can be concluded that the correlation matrix between response variables is different from the identity matrix. Because the correlation matrix is not an identity matrix, multi-response regression modeling is feasible.

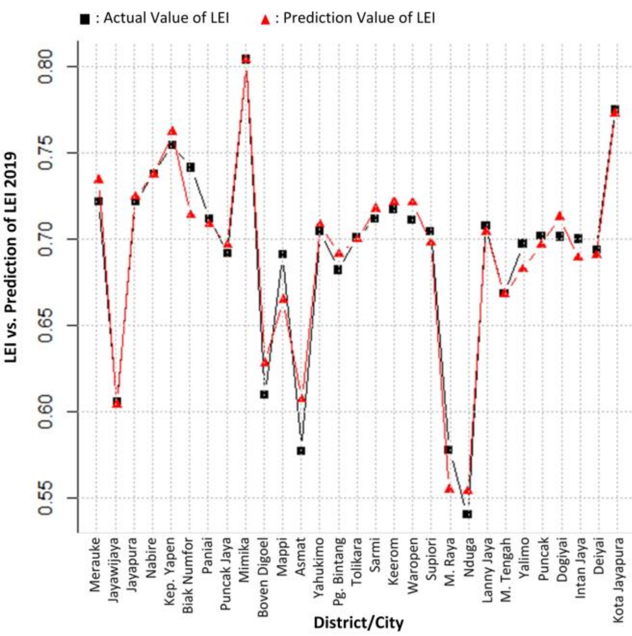

Fig. 5. Plot of Life Expectancy Index (LEI) vs. Prediction of LEI in 2019.

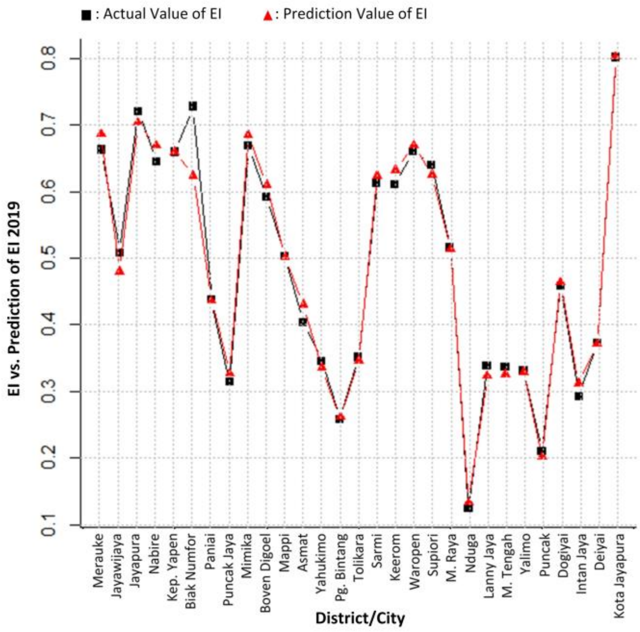

Fig. 6. Plot of Education Index (EI) vs. Prediction of EI in 2019.

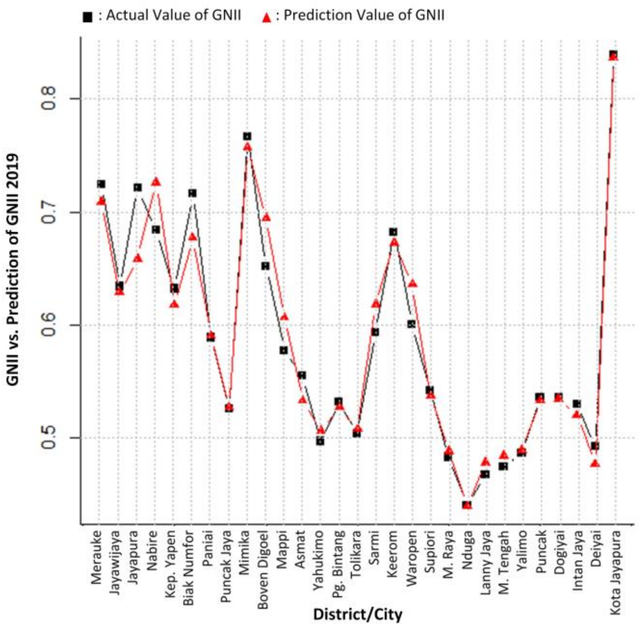

Fig. 7. Plot of Gross National Income Index (GNII) vs. Prediction of GNII in 2019.

Meanwhile, the application of our proposed mixture estimator model is appropriate for the data of three indexes from the three HDI dimensions in Papua province because there are different patterns between the predictor variables and the three response variables. These relationship patterns are partly under the characteristics of the Spline pattern, which changes at certain sub-intervals and other parts following the Kernel pattern, which does not have a specific pattern or seems random. In this case, the variable dependency ratio and percentage of the population working in agriculture are used as predictor variables for the Spline Smoothing component because the pattern tends to change at certain subintervals, whereas the variable population growth rate and economic growth rate are used as Kernel component variables because the pattern looks random.

By applying the multi-response mixture Spline

Smoothing and Kernel estimator (29) on the data of three indexes from the three HDI dimensions in Papua province in 2018 and trying several combinations of smoothing and bandwidth parameters, we obtained a minimum GCV value of 0.00066 with optimal smoothing parameters $\lambda_{1(opt)} = 1.64$, $\lambda_{2(opt)} = 7.319$, $\lambda_{3(opt)} = 1.1$ and optimal bandwidth parameters $\alpha_{1(opt)}$=4.059, $\alpha_{2(opt)}$=3.001, $\alpha_{3(opt)}$=1.001. This model produces $R^2 = 99.09\%$ and $MSE = 0.000201$. Based on the R^2 value, this model can describe the relationship between the predictor variables and the response variables up to 99.09%. These findings indicate that our proposed multi-response mixture estimator is suitable for modeling the data of three indexes from the three HDI dimensions in Papua Province.

Once the modeling process is complete, the next step is to validate the model to see its prediction ability. In this validation stage, the prediction model that has been obtained from the data of three indexes from the three HDI dimensions in 2018 is used to predict the three indexes from the three HDI dimensions in 2019. By applying the values of the predictor variables (population growth rate, the dependency ratio, the percentage of the population working in agriculture, and economic growth rate of the Papua Province) in 2019 to the obtained model, we can obtain predicted values of the three indexes from the three HDI dimensions in 2019. Once the prediction is made, we calculate the Mean Absolute Percentage Error (MAPE) value, which can be used as the model's eligibility criteria. The MAPE value for the ability of the model to predict the three indexes from the three HDI dimensions in 2019 is 2.5798% or the level of accuracy is 97.4202%. This MAPE value is less than 5%, which means that the multi-response mixture Spline Smoothing and Kernel model has an excellent ability to predict the three indexes from the three HDI dimensions in Papua Province. This good predictive ability is also shown in Fig. 5 until Fig. 7, where the predicted values of the three indexes from the three HDI dimensions in 2019 are very close to the actual values.

IV. CONCLUSION

This paper proposes the multi-response nonparametric regression model with a mixture of Spline Smoothing and Kernel estimator (29). This proposed multi-response mixture estimator is developed to handle the issue of how to model data with different patterns between each predictor in the multi-response case. Our proposed estimator is obtained through the two-stage estimation i.e., the first stage using PWLS to obtain the Spline Smoothing component, followed by the second stage using WLS to obtain the Kernel component. The selection of the best model for the proposed estimator was carried out using minimum GCV criteria. Our simulation study shows that the data with different patterns between each predictor are better modeled using our proposed mixture estimator rather than using only one kind of estimator. The proposed estimator can also be well implemented to model the three indexes from the three HDI dimensions of Papua Province in 2018 and provides satisfactory results. The predictive ability of the proposed model to predict the three indexes from the three HDI dimensions of Papua province in 2019 is very good, with the MAPE value less than 5%.

A limitation of this study is that we only use smoothing and bandwidth parameters according to the response variables. This proposed estimator can be developed for future work by considering smoothing and bandwidth parameters according to both response and predictor variables. Despite this limitation, this research certainly adds to our understanding of the mixture Spline Smoothing and Kernel estimator in multi-response nonparametric regression.

REFERENCES

[1] W. Härdle, *Applied nonparametric regression*, no. 19. Cambridge U.K: Cambridge university press, 1990.

[2] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, Second Edi. New York: Marcel Dekker, Inc, 1999.

[3] Y. Wang, W. Guo, and M. B. Brown, “Spline smoothing for bivariate data with applications to association between hormones,” *Statistica Sinica*, vol. 10, no. 2, pp. 377–397, 2000.

[4] B. Lestari, I. N. Budiantara, S. Sunaryo, and M. Mashuri, “Spline estimator in multi-response nonparametric regression model with unequal correlation of errors,” *Journal of Mathematics and Statistics*, vol. 6, no. 3, pp. 327–332, 2010, doi: 10.3844/jmssp.2010.327.332.

[5] D. Aydin, M. Memmedli, and R. E. Omay, “Smoothing Parameter Selection for Nonparametric Regression Using Smoothing Spline,” *European Journal of Pure and Applied Mathematics*, vol. 6, no. 2, pp. 222–238, 2013.

[6] A. A. R. Fernandes, I. N. Budiantara, B. W. Otok, and Suhartono, “Spline estimator for bi-responses nonparametric regression model for longitudinal data,” *Applied Mathematical Sciences*, vol. 8, no. 113–116, pp. 5653–5665, 2014, doi: 10.12988/ams.2014.47566.

[7] R. Diana, “Smoothing spline in multivariable semiparametric regression through fully Bayesian approach,” *Model Assisted Statistics and Applications*, vol. 13, no. 2, pp. 123–132, 2018, doi: 10.3233/MAS-180424.

[8] T. Gasser and H. Muller, “Estimating regression functions and their derivatives by the kernel method,” *Scand. J. Stat.*, vol. 11, no. 3, pp. 171–185, 1984.

[9] P. Hall and L. S. Huang, “Nonparametric kernel regression subject to monotonicity constraints,” *Ann. Stat.*, vol. 29, no. 3, pp. 624–647, 2001, doi: 10.1214/aos/1009210683.

[10] H. Okumura and K. Naito, “Non-parametric kernel regression for multinomial data,” *J. Multivar. Anal.*, vol. 97, no. 9, pp. 2009–2022, 2006, doi: 10.1016/j.jmva.2005.12.008.

[11] L. S. Dharmasena, P. Zeephongsekul, and B. M. De Silva, “Sample size determination for kernel regression estimation using sequential fixed-width confidence bands,” *IAENG International Journal of Applied Mathematics*, vol. 38, no. 3, pp. 129–135, 2008.

[12] P. Du, C. F. Parmeter, and J. S. Racine, “Nonparametric kernel regression with multiple predictors and multiple shape constraints,” *Stat. Sin.*, vol. 23, no. 3, pp. 1347–1371, 2013, doi: 10.5705/ss.2012.024.

[13] N. Chamidah and T. Saifudin, “Estimation of children growth curve based on kernel smoothing in multi-response nonparametric regression,” *Appl. Math. Sci.*, vol. 7, no. 37–40, pp. 1839–1847, 2013, doi: 10.12988/ams.2013.13168.

[14] E. Erçelik and M. Nadar, “Nonparametric density estimation based on beta prime kernel,” *Commun. Stat. - Theory Methods*, vol. 49, no. 2, pp. 325–342, 2020, doi: 10.1080/03610926.2018.1538458.

[15] I. N. Budiantara, V. Ratnasari, M. Ratna, W. Wibowo, N. Afifah, D. P. Rahmawati and M. A. D. Octavanny, “Modeling percentage of poor people in Indonesia using kernel and Fourier series mixed estimator in nonparametric regression,” *Investig. Operacional*, vol. 40, no. 4, pp. 538–550, 2019.

[16] Rismal, I. N. Budiantara, and D. D. Prastyo, “Mixture model of spline truncated and kernel in multivariable nonparametric regression,” in *AIP Conference Proceedings*, 2016, vol. 1739, no. 1, doi: 10.1063/1.4952565.

[17] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, “The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model,” *Commun. Stat. - Theory Methods*, 2020, doi: 10.1080/03610926.2019.1710201.

[18] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, “Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression,” *Journal of Mathematics*, vol. 2020, 2020, doi:

10.1155/2020/4712531.
[19] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanto, and D. P. Rahmawati, "Nonparametric Regression Model for Longitudinal Data with Mixed Truncated Spline and Fourier Series," *Abstract and Applied Analysis*, 2020, vol. 2020, doi: 10.1155/2020/4710745.
[20] M. I. E. Zulkifly, A. F. Wahab, and R. Zakaria, "B-Spline Curve Interpolation Model by using Intuitionistic Fuzzy Approach," *IAENG International Journal of Applied Mathematics*, vol. 50, no. 4, pp. 760-766, 2020.
[21] B. Rahmat, E. Joelianto, I. K. E. Purnama, and M. H. Purnomo, "An Improved Mean Shift Using Adaptive Fuzzy Gaussian Kernel for Indonesia Vehicle License Plate Tracking," *IAENG International Journal of Computer Science*, vol. 45, no. 3, pp. 458-471, 2018.
[22] M. G. Schimek, *Smoothing and regression: approaches, computation, and application*. John Wiley & Sons, 2013.
[23] T. Hastie and R. Tibshirani, "Generalized additive models," *Stat. Sci.*, vol. 1, no. 3, pp. 297–310, 1986, doi: 10.1214/ss/1177013604.
[24] G. Wahba, *Spline Models for Observational Data*. Pensylvania: SIAM, 1990.
[25] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numer. Math.*, 1978, doi: 10.1007/bf01404567.
[26] A. Rahayu, Purhadi, Sutikno, and D. D. Prastyo, "Multivariate gamma regression: Parameter estimation, hypothesis testing, and its application," *Symmetry*, vol. 12, no. 5, 2020, doi: 10.3390/SYM12050813.