Article

Reproducing Kernel Hilbert Space Approach to Multiresponse Smoothing Spline Regression Function

Budi Lestari [1,2], Nur Chamidah [2,3,*], Dursun Aydin [4] and Ersin Yilmaz [4]

[1] Department of Mathematics, Faculty of Mathematics and Natural Sciences, The University of Jember, Jember 68121, Indonesia

[2] Research Group of Statistical Modeling in Life Sciences, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[3] Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

[4] Department of Statistics, Faculty of Science, Muğla Sıtkı Koçman University, Muğla 48000, Turkey

[*] Correspondence: nur-c@fst.unair.ac.id

Abstract: In statistical analyses, especially those using a multiresponse regression model approach, a mathematical model that describes a functional relationship between more than one response variables and one or more predictor variables is often involved. The relationship between these variables is expressed by a regression function. In the multiresponse nonparametric regression (MNR) model that is part of the multiresponse regression model, estimating the regression function becomes the main problem, as there is a correlation between the responses such that it is necessary to include a symmetric weight matrix into a penalized weighted least square (PWLS) optimization during the estimation process. This is, of course, very complicated mathematically. In this study, to estimate the regression function of the MNR model, we developed a PWLS optimization method for the MNR model proposed by a previous researcher, and used a reproducing kernel Hilbert space (RKHS) approach based on a smoothing spline to obtain the solution to the developed PWLS optimization. Additionally, we determined the symmetric weight matrix and optimal smoothing parameter, and investigated the consistency of the regression function estimator. We provide an illustration of the effects of the smoothing parameters for the estimation results using simulation data. In the future, the theory generated from this study can be developed within the scope of statistical inference, especially for the purpose of testing hypotheses involving multiresponse nonparametric regression models and multiresponse semiparametric regression models, and can be used to estimate the nonparametric component of a multiresponse semiparametric regression model used to model Indonesian toddlers' standard growth charts.

Keywords: smoothing spline regression function; MNR; RKHS; consistency; standard growth charts

1. Introduction

A reproducing kernel Hilbert space (RKHS) theory was first introduced by Aronszajn in 1950 [1]. This theory was later developed by [1,2] to solve optimization problems in regression, especially nonparametric spline original regression. The RKHS approach was used by [3] for an M-type spline estimator. Next, Ref. [4] used the RKHS approach for a relaxed spline estimator.

There are many cases in our daily life that we have to analyze, especially cases involving the functional relationship between different variables. In statistics, to analyze the functional relationship between several variables, namely, the influence of the independent variable or predictor variable on the dependent variable or response variable, regression analysis is used. In regression analysis, it is necessary to build a mathematical model,

which is commonly referred to as a regression model, and this functional relationship is expressed by a regression function. In regression analysis, there are two kinds of basic regression model approaches, namely, parametric regression models and nonparametric regression models. In general, the main problem in regression analysis whether using a parametric regression model approach or a nonparametric regression model approach is the problem of estimating the regression function. In the parametric regression model, the problem of estimating the regression function is the same as the problem of estimating the parameters of the parametric regression model where this is different from the nonparametric regression model. In nonparametric regression models, estimating the regression function is equivalent to estimating an unknown smooth function contained in a Sobolev space using smoothing techniques.

There are several frequently used smoothing techniques for estimating nonparametric regression functions, for example, local linear, local polynomial, kernel, and spline. The research results of several previous researchers have shown that smoothing techniques such as local linear, local polynomial, and kernel are highly recommended for estimating nonparametric regression functions for prediction purposes. These researchers include [5,6], who used local linear for predicting hypertension risk and predicting Mycobacterium tuberculosis numbers, respectively; Ref. [7] used local linear for determining boundary correction of nonparametric regression function; Ref. [8] used local linear for determining the bias reduction of a regression function estimate; Ref. [9] used local linear to design a standard growth chart for assessing the nutritional status of toddlers; Refs. [10,11] used local polynomial for estimating regression functions in cases of errors-in-variable and correlated errors, respectively; Refs. [12,13] used local polynomial to estimate the regression function for functional data and for finite population, respectively; Ref. [14] discussed smoothing techniques using kernel; Refs. [15,16] discussed consistency kernel regression estimation and estimating regression functions for cases of correlated errors using kernel, respectively; and Refs. [17,18] discussed estimating covariance matrix and selecting bandwidth using kernel, respectively. However, local linear, local polynomial, and kernel are highly dependent on the bandwidth in the neighborhood of the target point. Thus, if we use these local linear, local polynomial, or kernel approaches to estimate a model with fluctuating data, then we require a small bandwidth, and this results in too rough an estimation of the curve. This means that these local linear, local polynomial, and kernel approaches do not consider smoothness, only the goodness of fit. Therefore, for estimation models with fluctuating data in the sub-intervals, these local linear, local polynomial, and kernel methods are not good to use, as the results of estimation result in a large value of the mean square error (MSE). This is different from spline approaches, which consider goodness of fit and smoothness factors, as discussed by [1,19], who used splines for modeling observational data and estimating nonparametric regression functions. Furthermore, for prediction and interpretation purposes, smoothing techniques such as smoothing spline and truncated spline are better and more flexible for estimating the nonparametric regression functions [20]. Due to the flexible nature of these splines, many researchers have been interested in using and developing them in several cases. For examples, M-type splines were used by [21] to analyze variance for correlated data, and by [22] for estimating both nonparametric and semiparametric regression functions; truncated splines have been discussed by [23] to estimate mean arterial pressure for prediction purpose and by [24] to estimate blood pressure for prediction and interpretation purposes. Additionally, Ref. [25] developed truncated spline for estimating a semiparametric regression model and determining the asymptotic properties of the estimator. Furthermore, Ref. [26] discussed the flexibility of B-spline and penalties in estimating regression function; Ref. [27] discussed analyzing current status data using penalized spline; Ref. [28] analyzed the association between cortisol and ACTH hormones using bivariate spline; and Ref. [29] analyzed censored data using spline regression. In addition, Ref. [30] used both kernel and spline for estimating the regression function and selecting the optimal smoothing parameter of a uniresponse nonparametric regression (UNR) model; Refs. [31,32] developed

both kernel and spline for estimating the regression function and for selecting the optimal smoothing parameter of a multiresponse nonparametric regression (MNR) model; and Ref. [33] discussed smoothing techniques, namely, kernel and spline, to estimate the coefficient of a rates model.

In regression modeling, a common problem involves more than one response variable observed at several values of predictor variables and between responses that are correlated with each other. The multiresponse nonparametric regression (MNR) model approach is appropriate for modeling the functions which represent the relationship between the response variable and predictor variable with correlated responses. In this model there is a correlation between the responses. Because of this correlation, it is necessary to construct a matrix called a weight matrix. Constructing the weight matrix is one of the things that distinguishes the MNR model approach from a classical model approach, that is, a parametric regression model or uniresponse nonparametric regression model approach. Thus, in the estimation process the regression function requires a weight matrix in the form of a symmetric matrix, especially a diagonal matrix. Furthermore, in the MNR model there are several smoothing techniques which can be used to estimate the regression function. One of these smoothing techniques is the smoothing spline approach. In recent years, studies on smoothing splines have attracted a great deal of attention and the methodology has been widely used in many areas of research, for example, for estimating regression functions of nonparametric regression models, in [34,35] used smoothing spline, mixed smoothing spline, and Fourier series; estimating regression functions were conducted by [36,37] for a semiparametric nonlinear regression model and a semiparametric regression model; and smoothing spline in an ANOVA model was discussed by [38]. Smoothing spline estimator, with its powerful and flexible properties, is one of the most popular estimators used for estimating regression function of the nonparametric regression model. Although the researchers mentioned above have previously discussed splines for estimating regression functions in many cases, none of them have used a reproducing kernel Hilbert space (RKHS) approach to estimate the regression function of the MNR model. On the other hand, even though there are studies, as mentioned above, that have used the RKHS approach for estimating regression functions, those researchers used the RKHS for estimating regression functions of single–response or uniresponse linear regression models only. This means that RKHS approaches were not used for estimating the regression function of the MNR model based on a smoothing spline estimator. In addition, although [34] used the RKHS approach to estimate the regression function of the MNR model, and also discussed it in a special case involving a simulation study; but Ref. [34] assumed that the three responses of the MNR model have the same smoothing parameter values, which in real life situation is a difficult assumption to fulfill. In addition, Ref. [34] did not discuss the consistency of the smoothing spline regression function estimator. Therefore, in this study we provide a theoretical discussion on estimating the smoothing spline regression function of the MNR model in case of unequal values of the smoothing parameters using the RKHS approach. In other words, in this study we discuss it for the more general case.

2. Materials and Methods

In this section, we briefly describe the materials and methods used according to the needs of this study, following the steps in the order in which they were carried out.

2.1. Multiresponse Nonparametric Regression Model

Suppose, given a paired observation $\{y_{ri}, t_{ri}\}$ which satisfies the following multiresponse nonparametric regression (MNR) model:

$$y_{ri} = f_r(t_{ri}) + \varepsilon_{ri}, \ i = 1,2, \dots, n_r, \ r = 1,2, \dots, p \tag{1}$$

where y_{ri} is the observation value of the response variable on the rth response and the ith observation; $f_r(\cdot)$ represents an unknown nonparametric regression function of rth

response which is assumed to be smooth in the sense that it is contained in a Sobolev space; t_{ri} is the observation value of a predictor variable on the rth response and the ith observation; and ε_{ri} represents the value of the random error on the rth response and the ith observation, which is assumed to have zero mean and variance σ_{ri}^2 (heteroscedastic). In this study, we assume that the correlation between responses is $\rho_{rs} = \begin{cases} \rho_r & for \quad r = s \\ 0 & for \quad r \neq s \end{cases}$.

In general, the main problem in MNR modelling is how we estimate the MNR model, which in this case is equivalent to the problem of estimating the regression function of the MNR model. There are many smoothing techniques that can be used to estimate the MNR model presented in (1), for example, kernel, local linear, splines, and local polynomial. One of these smoothing techniques is the spline approach, in which the smoothing spline is the most flexible estimator for estimating fluctuating data on sub-intervals. The following briefly presents the estimation method using the smoothing spline estimator. Further details related to the smoothing spline estimator can be found in [20].

2.2. Smoothing Spline Estimator

In this study, we estimated the regression function, $f_r(t_{ri})$ of the MNR model presented in (1) based on the smoothing spline estimator using the reproducing kernel Hilbert space approach, which is discussed in the following section. An estimate of the regression function of the MNR model presented in (1) can be obtained by developing the penalized weighted least squares (PWLS) optimization method proposed by [31], which is only used for the two-response nonparametric regression model with the same variance of errors, namely, the homoscedastic case. We then develop the PWLS optimization to estimate a nonparametric regression model with more than two responses, namely, the MNR model, in case of unequal variance of errors, which is called as heteroscedastic case. Hence, the estimated smoothing spline of the MNR model presented in (1) can be obtained by carrying out the following PWLS optimization:

$$\min_{f_1,\dots,f_p \in W_2^m[a_r,b_r]} \{N^{-1}(\mathbf{y_1} - \mathbf{f_1})^T \mathbf{W}_1 (\mathbf{y_1} - \mathbf{f_1}) + \cdots + (\mathbf{y_p} - \mathbf{f_p})^T \mathbf{W}_p (\mathbf{y_p} - \mathbf{f_p}) +$$

$$\lambda_1 \int_{a_1}^{b_1} \left(f_1^{(2)}(t) \right)^2 dt + \cdots + \lambda_p \int_{a_p}^{b_p} \left(f_p^{(2)}(t) \right)^2 dt \} \tag{2}$$

where $N = \sum_{r=1}^p n_r$; $\mathbf{W_1}, \dots, \mathbf{W_p}$ are symmetric weight matrices, $\lambda_1, \dots, \lambda_p$ are smoothing parameters, and $\mathbf{f_1}, \mathbf{f_2}, \dots, \mathbf{f_p}$ are unknown regression functions in a Sobolev space $W_2^m[a_r, b_r]$, where the Sobolev space $W_2^m[a_r, b_r]$ is defined as follows:

$W_2^m[a_r, b_r] = \{f | f^{(v)}, v = 0,1,2, \dots, m-1$ are absolutely continuous on $[a_r, b_r]$ and

$f^{(m)} \in L_2[a_r, b_r]$, where $L_2[a_r, b_r]$ is the collection of square integrable function on

$L_2[a_r, b_r], r = 1,2, \dots, p\}$

Furthermore, to obtain the solution to the PWLS provided in (2), we used the reproducing kernel Hilbert space (RKHS) approach. In the following section, we provide a brief review of RKHS. Further details related to RKHS can be found in [39], a paper concerning the theory of RKHS, and in [40], a textbook which discusses the use of RKHS in probability and statistics.

2.3. Reproducing Kernel Hilbert Space

The need to reproducing kernel Hilbert space (RKHS) arises in various fields, including statistics, theory of approximation, theory of machine learning, theory of group representation, and complex analysis. In statistics, the RKHS method is often used as a method for estimating a regression function based on the smoothing spline estimator for prediction purposes. In machine learning, the RKHS method is arguably the most popular approach for dealing with nonlinearity in data. Several researchers have discussed the RKHS method; for example, Refs. [41,42] discussed the use of RKHS in Support Vector Machines (SVM) and optimization problems, respectively, and Refs. [43,44] discussed the use of RKHS in asymptotic distribution for regression and machine learning.

A Hilbert space $\mathcal{H}$ is called an RKHS on a set X over field $\mathbb{F}$ if the following conditions are met [1,39]:

(i) $\mathcal{H}$ is a vector subspace of $\mathcal{F}(X,\mathbb{F})$, where $\mathcal{F}(X,\mathbb{F})$ is a vector space over $\mathbb{F}$;

(ii) $\mathcal{H}$ is endowed with an inner product $\langle\,,\rangle$, making it into a Hilbert space;

(iii) the linear evaluation functional $E_y:\mathcal{H} \to \mathbb{F}$, defined by $E_y(f) = f(y)$, is bounded, for every $y \in X$.

Furthermore, if $\mathcal{H}$ is an RKHS on X, then for every $y \in X$ there exists a unique vector $k_y \in \mathcal{H}$ such that for every $f \in \mathcal{H}$, $f(y) = \langle f, k_y \rangle$. This is because every bounded linear functional is provided by the inner product with a unique vector in $\mathcal{H}$. The function k_y is called a reproducing kernel (RK) for point y. The reproducing kernel (RK) for $\mathcal{H}$ is a two–variable function defined by $K(x,y) = k_y(x)$. Hence, we have $K(x,y) = k_y(x) = \langle k_y, k_x \rangle$ and $\|E_y\|^2 = \|k_y\|^2 = \langle k_y, k_y \rangle = K(y,y)$.

In this study, we provide a simulation study to evaluate the performance of the proposed MNR model estimation method.

2.4. Simulation

The simulation in this study consists of a simulation to determine the optimal smoothing spline based on a generalized cross-validation (GCV) criterion to obtain the best estimated MNR model and a simulation to describe the effect of the smoothing parameters on the estimation results of the regression function of the MNR model based on minimal GCV value. We generate samples sized $n = 100$ from the MNR model and provide an illustration of the effects of the smoothing parameters in order to estimate the results of the MNR model by comparing three kinds of different smoothing parameter values, namely, small, optimal, and large smoothing parameters values.

In the following section, we provide the results and discussion of this study covering estimation of the regression function of the MNR model using the RKHS approach by estimating the symmetric weight matrix and optimal smoothing parameters, a simulation study, and investigating the consistency of the smoothing spline regression function estimator.

3. Results and Discussions

The results and discussion presented in this section include estimating the regression function of the MNR model using RKHS, estimating the weight matrix, estimating the optimal smoothing parameter, investigating the consistency of the regression function estimator, a simulation study, and an application example using real data.

3.1. Estimating the Regression Function of the MNR Model Using the RKHS Approach

The MNR model presented in (1) can be expressed in matrix notation as follows:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2}, \dots, \mathbf{y_p})^{\mathrm{T}}$, $\mathbf{f} = (\mathbf{f_1}, \mathbf{f_2}, \dots, \mathbf{f_p})^{\mathrm{T}}$, $\mathbf{t} = (\mathbf{t_1}, \mathbf{t_2}, \dots, \mathbf{t_p})^{T}$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_p})^{\mathrm{T}}$, $\mathbf{y_r} = (y_{r1}, y_{r2}, \dots, y_{rn_r})^{\mathrm{T}}$, $\mathbf{f_r} = (f_r(t_{r1}), f_r(t_{r2}), \dots, f_r(t_{rn_r}))^{\mathrm{T}}$, $\mathbf{t_r} = (t_{r1}, t_{r2}, \dots, t_{rn_r})^{T}$, and $\boldsymbol{\varepsilon_r} = (\varepsilon_{r1}, \varepsilon_{r2}, \dots, \varepsilon_{rn_r})^{\mathrm{T}}$.

We assume that $\boldsymbol{\varepsilon}$ is a zero mean random error with covariance $\mathbf{W}^{-1}$. In this case, the covariance matrix $\mathbf{W}^{-1}$ is a symmetrical matrix, that is, it is a diagonal matrix which can be expressed as follows:

$$\mathbf{W}^{-1} = diag(\mathbf{W_1^{-1}}, \mathbf{W_2^{-1}}, \dots, \mathbf{W_p^{-1}}) \tag{4}$$

where $\mathbf{W_r^{-1}}$ is an rth-response covariance matrix of $\boldsymbol{\varepsilon_r}$ for $r = 1,2, \dots, p$.

To determine the regression function of the MNR model (1) using the RKHS approach, we first express the MNR model in a general smoothing spline regression model [20]. Therefore, we can express the MNR model (1) as follows:

$$y_{ri} = \mathcal{L}_{t_r} f_r + \varepsilon_{ri}; \ i = 1,2, \dots, n_r; \ r = 1,2, \dots, p \tag{5}$$

where $f_r \in \mathcal{H}_r$ is an unknown smooth function, $\mathcal{L}_{t_r} \in \mathcal{H}_r$ is a bounded linear functional, and $\mathcal{H}_r$ is a Hilbert space.

Next, the Hilbert space $\mathcal{H}_r$ is decomposed into a direct sum of the Hilbert subspace G_r and Hilbert sub space K_r, where G_r has basis $\{\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rm_r}\}$, K_r has basis $\{\beta_{r1}, \beta_{r2}, \dots, \beta_{rn_r}\}$, and $G_r \perp K_r$ is as follows:

$$\mathcal{F}_r = G_r \oplus K_r. \tag{6}$$

This implies that for $g_r \in G_r$, $z_r \in K_r$, and $r = 1,2, \dots, p$ we can express every function $f_r \in \mathcal{F}_r$ as follows:

$$f_r = g_r + z_r. \tag{7}$$

Because $\{\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rm_r}\}$ is the basis of the Hilbert subspace G_r and $\{\beta_{r1}, \beta_{r2}, \dots, \beta_{rn_r}\}$ is the basis of the Hilbert subspace K_r, the function f_r in (7) can be expressed as follows:

$$f_r = \sum_{i=1}^{m_r} b_{ri} \alpha_{ri} + \sum_{j=1}^{n_r} c_{rj} \beta_{rj} = \boldsymbol{\alpha}_r^T \mathbf{b}_r + \boldsymbol{\beta}_r^T \mathbf{c}_r \tag{8}$$

where $r = 1,2, \dots, p$; $b_{ri}, c_{rj} \in \mathbb{R}$; $\boldsymbol{\alpha}_r = (\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rm_r})^T$; $\mathbf{b}_r = (b_{r1}, b_{r2}, \dots, b_{rm_r})^T$; $\boldsymbol{\beta}_r = (\beta_{r1}, \beta_{r2}, \dots, \beta_{rn_r})^T$; and $\mathbf{c}_r = (c_{r1}, c_{r2}, \dots, c_{rn_r})^T$.

Hence, for $r = 1,2, \dots, p$ and $i = 1,2, \dots, n_r$, we have

$$\mathcal{L}_{t_{ri}} f_r = \mathcal{L}_{t_{ri}}(g_r + z_r) = \mathcal{L}_{t_{ri}}(g_r) + \mathcal{L}_{t_{ri}}(z_r)$$

$$= g_r(t_{ri}) + z_r(t_{ri})$$

$$= f_r(t_{ri}).$$

Because $\mathcal{L}_{t_r} \in \mathcal{F}_r$ is a bounded linear functional in the Hilbert space $\mathcal{F}_r$, according to [20] there exists a Riesz representer $\delta_{ri} \in \mathcal{F}_r$ such that

$$\mathcal{L}_{t_{ri}} f_r = \langle \delta_{ri}, f_r \rangle = f_r(t_{ri}) \tag{9}$$

where $f_r \in \mathcal{F}_r$ and $\langle \cdot, \cdot \rangle$ denote an inner product. Next, by considering Equations (8) and (9) and applying the properties of the inner product, the function $f_r(t_{ri})$ can be written as follows:

$$f_r(t_{ri}) = \langle \delta_{ri}, \boldsymbol{\alpha}_r^T \mathbf{b}_r + \boldsymbol{\beta}_r^T \mathbf{c}_r \rangle$$
$$= \langle \delta_{ri}, \boldsymbol{\alpha}_r^T \mathbf{b}_r \rangle + \langle \delta_{ri}, \boldsymbol{\beta}_r^T \mathbf{c}_r \rangle \tag{10}$$

Then, based on Equation (10), we can obtain the regression functions $f_r(t_{ri})$ for $r = 1,2, \dots, p$, which are the regresion functions for the first response, the second response, …, and the pth response, as follows:

For $r = 1$,   we have: $f_1(t_{1i}) = \langle \delta_{1i}, \boldsymbol{\alpha}_1^T \mathbf{b}_1 \rangle + \langle \delta_{1i}, \boldsymbol{\beta}_1^T \mathbf{c}_1 \rangle$, $i = 1,2, \dots, n_1$

For $r = 2$,   we have: $f_2(t_{2i}) = \langle \delta_{2i}, \boldsymbol{\alpha}_2^T \mathbf{b}_2 \rangle + \langle \delta_{2i}, \boldsymbol{\beta}_2^T \mathbf{c}_2 \rangle$, $i = 1,2, \dots, n_2$

$$\vdots$$

For $r = p$,   we have: $f_p(t_{pi}) = \langle \delta_{pi}, \boldsymbol{\alpha}_p^T \mathbf{b}_p \rangle + \langle \delta_{pi}, \boldsymbol{\beta}_p^T \mathbf{c}_p \rangle$, $i = 1,2, \dots, n_p$ (11)

Hence, following Equation (11), we obtain the regression function for $i = 1$ as follows:

$$\mathbf{f_1}(t_1) = \big(f_1(t_{11}), f_1(t_{12}), \dots, f_1(t_{1n_1})\big)^T = \mathbf{A_1 b_1} + \mathbf{C_1 d_1} \tag{12}$$

where $\mathbf{b_1} = (b_{11}, b_{12}, \dots, b_{1m_1})^T$; $\mathbf{d_1} = (d_{11}, d_{12}, \dots, d_{1n_1})^T$; $\mathbf{A_1} =$

$$\begin{pmatrix} \langle \delta_{11}, \alpha_{11} \rangle & \langle \delta_{11}, \alpha_{12} \rangle & \cdots & \langle \delta_{11}, \alpha_{1m_1} \rangle \\ \langle \delta_{12}, \alpha_{11} \rangle & \langle \delta_{12}, \alpha_{12} \rangle & \cdots & \langle \delta_{12}, \alpha_{1m_1} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \delta_{1n_1}, \alpha_{11} \rangle & \langle \delta_{1n_1}, \alpha_{12} \rangle & \cdots & \langle \delta_{1n_1}, \alpha_{1m_1} \rangle \end{pmatrix}$$;   and $\mathbf{C_1} =$

$$\begin{pmatrix} \langle \delta_{11}, \beta_{11} \rangle & \langle \delta_{11}, \beta_{12} \rangle & \cdots & \langle \delta_{11}, \beta_{1n_1} \rangle \\ \langle \delta_{12}, \beta_{11} \rangle & \langle \delta_{12}, \beta_{12} \rangle & \cdots & \langle \delta_{12}, \beta_{1n_1} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \delta_{1n_1}, \beta_{11} \rangle & \langle \delta_{1n_1}, \beta_{12} \rangle & \cdots & \langle \delta_{1n_1}, \beta_{1n_1} \rangle \end{pmatrix}.$$

Similarly, we obtain the regression functions for $i = 2, \dots, n_1$, which are $\mathbf{f_2}(t_2)$, $\mathbf{f_3}(t_3)$, …, $\mathbf{f_p}(t_p)$, as follows:

$$\mathbf{f_2}(t_2) = \big(f_2(t_{11}), f_2(t_{12}), \dots, f_2(t_{1n_2})\big)^T = \mathbf{A_2 b_2} + \mathbf{C_2 d_2}$$

$$\mathbf{f_3}(t_3) = \big(f_3(t_{11}), f_3(t_{12}), \dots, f_3(t_{1n_3})\big)^T = \mathbf{A_3 b_3} + \mathbf{C_3 d_3} \tag{13}$$

$$\vdots$$

$$\mathbf{f_p}(t_p) = \big(f_p(t_{11}), f_p(t_{12}), \dots, f_p(t_{1n_p})\big)^T = \mathbf{A_p b_p} + \mathbf{C_p d_p}$$

Hence, based on Equations (12) and (13), we obtain the regression function $\mathbf{f}(t)$ of the MNR model as follows:

$$\mathbf{f}(t) = \big(\mathbf{f_1}(t_1), \mathbf{f_2}(t_2), \dots, \mathbf{f_p}(t_p)\big)^T = \big(\mathbf{A_1 b_1}, \mathbf{A_2 b_2}, \dots, \mathbf{A_p b_p}\big)^T + \big(\mathbf{C_1 d_1}, \mathbf{C_2 d_2}, \dots, \mathbf{C_p d_p}\big)^T$$

$$= diag\Big(\big(\mathbf{A_1}, \mathbf{A_2}, \dots, \mathbf{A_p}\big)\big(\mathbf{b_1}, \mathbf{b_2}, \dots, \mathbf{b_p}\big)^T\Big) + \tag{14}$$

$$diag\Big(\big(\mathbf{C_1}, \mathbf{C_2}, \dots, \mathbf{C_p}\big)\big(\mathbf{d_1}, \mathbf{d_2}, \dots, \mathbf{d_p}\big)^T\Big) = \mathbf{Ab} + \mathbf{Cd}$$

Thus, we can express the MNR model presented in (1) in matrix notation as follows:

$$\mathbf{y} = \mathbf{Ab} + \mathbf{Cd} + \boldsymbol{\varepsilon} \tag{15}$$

where $\mathbf{A} = diag(\mathbf{A_1}, \mathbf{A_2}, \dots, \mathbf{A_p})$ is an $(N \times M)$ diagonal matrix with $N = \sum_{r=1}^p n_r$, $M = \sum_{r=1}^p m_r$; $\mathbf{b} = \big(\mathbf{b_1^T}, \mathbf{b_2^T}, \dots, \mathbf{b_p^T}\big)^T$ is an $(M \times 1)$ vector of parameters; $\mathbf{C} = diag(\mathbf{C_1}, \mathbf{C_2}, \dots, \mathbf{C_p})$ is an $(N \times N)$ diagonal matrix; and $\mathbf{d} = \big(\mathbf{d_1^T}, \mathbf{d_2^T}, \dots, \mathbf{d_p^T}\big)^T$ is an $(N \times 1)$ vector of parameters.

Now, we can determine an estimated smoothing spline regression function of the MNR model presented in (1) using the RKHS approach by taking the solution of the following optimization:

$$\underset{\substack{f_r \in \mathcal{F}_r \\ r=1,2,\dots,p}}{\text{Min}} \left\{ \left\| \mathbf{W}^{\frac{1}{2}} \boldsymbol{\varepsilon} \right\|^2 \right\} = \underset{\substack{f_r \in \mathcal{F}_r \\ r=1,2,\dots,p}}{\text{Min}} \left\{ \left\| \mathbf{W}^{\frac{1}{2}} (\mathbf{y} - \mathbf{f}) \right\|^2 \right\} \tag{16}$$

with a constraint $\int_{a_r}^{b_r} \big(f_r^{(m)}(t_r)\big)^2 dt_r < \gamma_r$, where $\gamma_r \geq 0$ and $r = 1,2, \dots, p$.

Note that determining the solution to Equation (16) is equivalent to determining the solution to the following PWLS optimization:

$$\underset{\substack{f_r \in W_2^m[a_r, b_r] \\ r=1,2,\dots,p}}{\text{Min}} \left\{ N^{-1}(\mathbf{y} - \mathbf{f})^{\mathbf{T}}\mathbf{W}(\mathbf{y} - \mathbf{f}) + \sum_{r=1}^{p} \lambda_r \int_{a_r}^{b_r} \left(f_r^{(m)}(t_r)\right)^2 dt_r \right\} \tag{17}$$

where $N = \sum_{r=1}^{p} n_r$; $N^{-1}(\mathbf{y} - \mathbf{f})^{\mathbf{T}}\mathbf{W}(\mathbf{y} - \mathbf{f})$ is a weighted least square that represents the goodness of fit, $\sum_{r=1}^{p} \lambda_r \int_{a_r}^{b_r} \left(f_r^{(m)}(t_r)\right)^2 dt_r$ represents a penalty that measures smoothness, and $\lambda_r, r = 1,2,\dots,p$ represents a smoothing parameter which controls the trade-off between the goodness of fit and the penalty.

Next, we decompose the penalty presented in (17) as follows:

$$\sum_{r=1}^{p} \lambda_r \int_{a_r}^{b_r} \left(f_r^{(m)}(t_r)\right)^2 dt_r = \lambda_1 \int_{a_1}^{b_1} \left(f_1^{(m)}(t_1)\right)^2 dt_1 + \dots + \lambda_p \int_{a_p}^{b_p} \left(f_p^{(m)}(t_p)\right)^2 dt_p \tag{18}$$

Because we have
$$\begin{cases} \int_{a_1}^{b_1} \left(f_1^{(m)}(t_1)\right)^2 dt_1 = \langle \boldsymbol{\beta}_1^{\mathbf{T}}\mathbf{d}_1, \boldsymbol{\beta}_1^{\mathbf{T}}\mathbf{d}_1 \rangle = \mathbf{d}_1^{\mathbf{T}}\langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_1^{\mathbf{T}}\rangle\mathbf{d}_1 = \mathbf{d}_1^{\mathbf{T}}\mathbf{C}_1\mathbf{d}_1 \\ \int_{a_2}^{b_2} \left(f_2^{(m)}(t_2)\right)^2 dt_2 = \langle \boldsymbol{\beta}_2^{\mathbf{T}}\mathbf{d}_2, \boldsymbol{\beta}_2^{\mathbf{T}}\mathbf{d}_2 \rangle = \mathbf{d}_2^{\mathbf{T}}\langle \boldsymbol{\beta}_2, \boldsymbol{\beta}_2^{\mathbf{T}}\rangle\mathbf{d}_2 = \mathbf{d}_2^{\mathbf{T}}\mathbf{C}_2\mathbf{d}_2 \\ \qquad\qquad\qquad\qquad\vdots \\ \int_{a_p}^{b_p} \left(f_p^{(m)}(t_p)\right)^2 dt_p = \langle \boldsymbol{\beta}_p^{\mathbf{T}}\mathbf{d}_p, \boldsymbol{\beta}_p^{\mathbf{T}}\mathbf{d}_p \rangle = \mathbf{d}_p^{\mathbf{T}}\langle \boldsymbol{\beta}_p, \boldsymbol{\beta}_p^{\mathbf{T}}\rangle\mathbf{d}_p = \mathbf{d}_p^{\mathbf{T}}\mathbf{C}_p\mathbf{d_p} \end{cases}$$
we are able to obtain the penalty presented in (17) or (18) as follows:

$$\sum_{r=1}^{p} \lambda_r \int_{a_r}^{b_r} \left(f_r^{(m)}(t_r)\right)^2 dt_r = \mathbf{d}^{\mathbf{T}}\boldsymbol{\Lambda}\mathbf{C}\mathbf{d} \tag{19}$$

where $\boldsymbol{\Lambda} = diag\left(\lambda_1\mathbf{I}_{n_1}, \lambda_2\mathbf{I}_{n_2}, \dots, \lambda_p\mathbf{I}_{n_p}\right)$. Furthermore, we can write the goodness of fit component in (17) as follows:

$$N^{-1}(\mathbf{y} - \mathbf{f})^{\mathbf{T}}\mathbf{W}(\mathbf{y} - \mathbf{f}) = N^{-1}(\mathbf{y} - \mathbf{Ab} - \mathbf{Cd})^{\mathbf{T}}\mathbf{W}(\mathbf{y} - \mathbf{Ab} - \mathbf{Cd}) \tag{20}$$

Based on Equations (19) and (20), we can express the PWLS optimization presented in (17) as follows:

$$\underset{\mathbf{b} \in \mathbb{R}^{pn}, \mathbf{d} \in \mathbb{R}^{pm}}{\text{Min}}\{\mathbf{Q}(\mathbf{b}, \mathbf{d})\} = \underset{\mathbf{b} \in \mathbb{R}^{pn}, \mathbf{d} \in \mathbb{R}^{pm}}{\text{Min}}\{ N^{-1}(\mathbf{y} - \mathbf{Ab} - \mathbf{Cd})^{\mathbf{T}} \times$$
$$\mathbf{W}(\mathbf{y} - \mathbf{Ab} - \mathbf{Cd}) + \mathbf{d}^{\mathbf{T}}\boldsymbol{\Lambda}\mathbf{Cd}\} \tag{21}$$

The solution to (21) can be obtained by taking the partial diferentiation $\mathbf{Q}(\mathbf{b}, \mathbf{d})$ with respecct to $\mathbf{b}$ and $\mathbf{d}$. In this step, we obtain the estimations of $\mathbf{b}$ and $\mathbf{d}$ as follows:

$$\hat{\mathbf{b}} = (\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{Wy}, \text{ and } \hat{\mathbf{d}} = \mathbf{D}^{-1}\mathbf{W}[\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W}]\mathbf{y}$$

where $\mathbf{D} = \mathbf{WC} + N\boldsymbol{\Lambda}\mathbf{I}$.

From this step, we obtain the estimation of the smoothing spline regression function of the MNR model presented in (1) or (15) as follows:

$$\hat{\mathbf{f}} = \mathbf{A}\hat{\mathbf{b}} + \mathbf{C}\hat{\mathbf{d}} = \mathbf{Hy} \tag{22}$$

where $\mathbf{H} = \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W} + \mathbf{CD}^{-1}\mathbf{W}(\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W})$, $\mathbf{D} = \mathbf{WC} + N\boldsymbol{\Lambda}\mathbf{I}$, $\boldsymbol{\Lambda} = diag\left(\lambda_1\mathbf{I}_{n_1}, \lambda_2\mathbf{I}_{n_2}, \dots, \lambda_p\mathbf{I}_{n_p}\right)$, $\mathbf{I}$ is an identity matrix with dimension N, and $N = \sum_{r=1}^{p} n_r$.

3.2. Estimating the Symmetric Weight Matrix

Based on MNR model presented in (3), the $\mathbf{W^{-1}}$ from Equation (4) is a covariance matrix of the random error $\boldsymbol{\varepsilon}$. To obtain the estimated weight matrix$\hat{\mathbf{W}}$, where the weight matrix $\mathbf{W}$ is the inverse of the covariance matrix, we first we consider a paired observation $\{y_{ri}, t_{ri}\}$, $r = 1,2,\dots,p$; $i = 1,2,\dots,n_r$ which follows the MNR model presented in (3). Second, supposing that $\mathbf{y} = \left(\mathbf{y_1}, \mathbf{y_2}, \dots, \mathbf{y_p}\right)^{\mathbf{T}}$ is a multivariate (i.e., N-variates where $N = \sum_{r=1}^{p} n_r$) normally distributed random sample with mean $\mathbf{f}$ and covariance $\mathbf{W^{-1}}$, we have the following likelihood function:

$$L(\mathbf{f}, \mathbf{W}|\mathbf{y}) = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{\frac{N}{2}}(\mathbf{W}^{-1})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}_j - \mathbf{f}_j)^{\mathbf{T}}\mathbf{W}(\mathbf{y}_j - \mathbf{f}_j)\right)\right\} \qquad (23)$$

Because $N = \sum_{r=1}^{p} n_r$ and $\mathbf{W} = diag(\mathbf{W_1}, \mathbf{W_2}, \dots, \mathbf{W_p})$, the likelihood function presented in (23) can be written as follows:

$$L(\mathbf{f}, \mathbf{W}|\mathbf{y}) = \frac{1}{(2\pi)^{\frac{nn_1}{2}}(\mathbf{W}_1^{-1})^{\frac{n}{2}}} exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y_{1j}} - \mathbf{f_{1j}})^{\mathbf{T}}\mathbf{W_1}\,(\mathbf{y_{1j}} - \mathbf{f_{1j}})\right) \times$$

$$\frac{1}{(2\pi)^{\frac{nn_2}{2}}(\mathbf{W}_2^{-1})^{\frac{n}{2}}} exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y_{2j}} - \mathbf{f_{2j}})^{\mathbf{T}}\mathbf{W_2}\,(\mathbf{y_{2j}} - \mathbf{f_{2j}})\right) \times \dots \times \quad (24)$$

$$\frac{1}{(2\pi)^{\frac{nn_p}{2}}(\mathbf{W}_p^{-1})^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y_{pj}} - \mathbf{f_{pj}})^{\mathbf{T}}\mathbf{W_p}\,(\mathbf{y_{pj}} - \mathbf{f_{pj}})\right)$$

Next, based on (24), the estimated weight matrix can be obtained by carrying out the following optimization:

$$L(\mathbf{f}, \mathbf{W}|\mathbf{y}) = \underset{\mathbf{W_1}}{\text{Max}}\left\{\frac{1}{(2\pi)^{\frac{nn_1}{2}}(\mathbf{W}_1^{-1})^{\frac{n}{2}}} exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y_{1j}} - \mathbf{f_{1j}})^{\mathbf{T}}\mathbf{W_1}, \dots (\mathbf{y_{1j}} - \mathbf{f_{1j}})\right)\right\} \times$$

$$\underset{\mathbf{W_2}}{\text{Max}}\left\{\frac{1}{(2\pi)^{\frac{nn_2}{2}}(\mathbf{W}_2^{-1})^{\frac{n}{2}}} exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y_{2j}} - \mathbf{f_{2j}})^{\mathbf{T}}\mathbf{W_2}, \dots (\mathbf{y_{2j}} - \mathbf{f_{2j}})\right)\right\} \times \cdots \times \quad (25)$$

$$\underset{\mathbf{W_p}}{\text{Max}}\left\{\frac{1}{(2\pi)^{\frac{nn_p}{2}}(\mathbf{W}_p^{-1})^{\frac{n}{2}}} exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{y_{pj}} - \mathbf{f_{pj}})^{\mathbf{T}}\mathbf{W_p}\,(\mathbf{y_{pj}} - \mathbf{f_{pj}})\right)\right\}$$

According to [45], the maximum value of each component of the likelihood function in Equation (25) can be determined using the following equations:

$$\widehat{\mathbf{W}}_1 = \frac{\hat{\varepsilon}_1\hat{\varepsilon}_1^{\mathbf{T}}}{N} = \frac{(y_1 - \hat{f}_1)(y_1 - \hat{f}_1)^{\mathbf{T}}}{N}, \ \ \widehat{\mathbf{W}}_2 = \frac{\hat{\varepsilon}_2\hat{\varepsilon}_2^{\mathbf{T}}}{N} = \frac{(y_2 - \hat{f}_2)(y_2 - \hat{f}_2)^{\mathbf{T}}}{N}, \ \dots, \ \widehat{\mathbf{W}}_p = \frac{\hat{\varepsilon}_p\hat{\varepsilon}_p^{\mathbf{T}}}{N} = \frac{(y_p - \hat{f}_p)(y_p - \hat{f}_p)^{\mathbf{T}}}{N}.$$

We may express the estimated smoothing spline regression function presented in (22) as follows:

$$\hat{\mathbf{f}}(\lambda, \sigma^2) = \mathbf{H}(\lambda, \sigma^2)\mathbf{y} \qquad (26)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)^T$, and $\boldsymbol{\sigma^2} = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)^T$.

Hence, the maximum likelihood estimator for the weight matrix $\mathbf{W}$ is provided by:

$$\widehat{\mathbf{W}} = diag(\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2, \dots, \widehat{\mathbf{W}}_p) = diag\left(\frac{(y_1 - \hat{f}_1)(y_1 - \hat{f}_1)^{\mathbf{T}}}{N}, \frac{(y_2 - \hat{f}_2)(y_2 - \hat{f}_2)^{\mathbf{T}}}{N}, \dots, \frac{(y_p - \hat{f}_p)(y_p - \hat{f}_p)^{\mathbf{T}}}{N}\right)$$

$$= diag\left(\frac{(\mathbf{I}_{n_1} - \mathbf{H}(\lambda_1, \sigma_1^2))y_1 y_1^{\mathbf{T}}(\mathbf{I}_{n_1} - \mathbf{H}(\lambda_1, \sigma_1^2))^{\mathbf{T}}}{N}, \frac{(\mathbf{I}_{n_1} - \mathbf{H}(\lambda_1, \sigma_1^2))y_1 y_1^{\mathbf{T}}(\mathbf{I}_{n_1} - \mathbf{H}(\lambda_1, \sigma_1^2))^{\mathbf{T}}}{N}, \dots, \frac{(\mathbf{I}_{n_p} - \mathbf{H}(\lambda_p, \sigma_p^2))y_p y_p^{\mathbf{T}}(\mathbf{I}_{n_p} - \mathbf{H}(\lambda_p, \sigma_p^2))^{\mathbf{T}}}{N}\right).$$

This shows that the estimated weight matrix obtained above is a symmetric matrix, specifically, a diagonal matrix the main diagonal components of which are the estimated weight matrices of the first response, second response, etc., up to the *p*-th response.

3.3. Estimating Optimal Smoothing Parameters

In MNR modeling, selection of the optimal smoothing parameter value λ cannot be omitted, and is crucial to obtaining a good regression function fit of the MNR model based on the smoothing spline estimator. According to [46], there are several criteria that can be used to select λ, including minimizing cross-validation (CV), generalized cross-validation (GCV), Mallows' C_p, and Akaike's information criterion (AIC). However, according to [47], for good regression function fitting based on the spline estimator Mallows' C_p and GCV are the most satisfactory.

In this section, we determine the optimal smoothing parameter value for good regression function fitting of the MNR model (1). Taking into account Equation (26), we may express the estimated smoothing spline regression function presented in (22) as follows:

$$\hat{\mathbf{f}}(\lambda, \sigma^2) = \mathbf{H}(\lambda, \sigma^2)\mathbf{y}$$

where $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)^T$, $\sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)^T$. The mean squared error (MSE) of the estimated smoothing spline regression function presented in (26) is provided by

$$MSE = \frac{(\mathbf{y}-\hat{\mathbf{f}}(\lambda,\sigma^2))^{\mathbf{T}}\mathbf{W}(\mathbf{y}-\hat{\mathbf{f}}(\lambda,\sigma^2))}{\sum_{r=1}^{p} n_r} = \frac{(\mathbf{y}-\hat{\mathbf{f}}(\lambda,\sigma^2))^{\mathbf{T}}\mathbf{W}(\mathbf{y}-\hat{\mathbf{f}}(\lambda,\sigma^2))}{N}$$

$$= \frac{\left((\mathbf{I}-\mathbf{H}(\lambda,\sigma^2))\mathbf{y}\right)^{\mathbf{T}}\mathbf{W}(\mathbf{I}-\mathbf{H}(\lambda,\sigma^2))\mathbf{y}}{N}$$

$$= \frac{\left\|\mathbf{W}^{\frac{1}{2}}(\mathbf{I}-\mathbf{H}(\lambda,\sigma^2))\mathbf{y}\right\|^2}{N}$$

Hereinafter, we define this function as follows:

$$G(\lambda) = \frac{N^{-1}\left\|\mathbf{W}^{\frac{1}{2}}(\mathbf{I}-\mathbf{H}(\lambda,\sigma^2))\mathbf{y}\right\|^2}{\left(\frac{1}{N} trace(\mathbf{I_N}-\mathbf{H}(\lambda,\sigma^2))\right)^2} \tag{27}$$

Therefore, based on (27), we can obtain the optimal smoothing parameter value, $\lambda_{\mathbf{opt}} = (\lambda_{1;opt}, \lambda_{2;opt}, \dots, \lambda_{p;opt})^T$, by taking the solution of the following optimization:

$$G_{opt}(\lambda_{\mathbf{opt}}) = \underset{\lambda_1,\dots,\lambda_p \in \mathbb{R}^+}{\text{Min}} \{G(\lambda)\}$$

$$= \underset{\lambda_1,\dots,\lambda_p \in \mathbb{R}^+}{\text{Min}} \left\{ \frac{N^{-1}\left\|\mathbf{W}^{\frac{1}{2}}(\mathbf{I}-\mathbf{H}(\lambda,\sigma^2))\mathbf{y}\right\|^2}{\left(\frac{1}{N} trace(\mathbf{I_N}-\mathbf{H}(\lambda,\sigma^2))\right)^2} \right\} \tag{28}$$

where $\mathbb{R}^+$ represents a positive real number set and $N = \sum_{r=1}^{p} n_r$.

Thus, the optimal smoothing parameter value $\lambda_{\mathbf{opt}} = (\lambda_{1;opt}, \lambda_{2;opt}, \dots, \lambda_{p;opt})^T$ is obtained from the minimizing process of the function $G(\lambda)$ in (27). The function $G(\lambda)$ in (27) is called the generalized cross-validation function [1].

3.4. Simulation Study

In this section, we provide a simulation study for estimating the smoothing spline regression function of the MNR model, where the performance of the proposed MNR model estimation method depends on the selection of an optimal smoothing parameter value. For example, we generate samples with size $n = 100$ from an MNR model, namely, a three-response nonparametric regression model, as follows:

$$\left. \begin{array}{ll} y_{1i} = 5 + 3\sin(2\pi t_{1i}^2) + \varepsilon_{1i}, & \text{for } i = 1,2,\dots,n \\ y_{2i} = 3 + 3\sin(2\pi t_{1i}^2) + \varepsilon_{2i}, & \text{for } i = 1,2,\dots,n \\ y_{3i} = 1 + 3\sin(2\pi t_{1i}^2) + \varepsilon_{2i}, & \text{for } i = 1,2,\dots,n \end{array} \right\} \tag{29}$$

where $n = 100$ and with correlations $\rho_{12} = 0.6$, $\rho_{13} = 0.7$, $\rho_{23} = 0.8$ and variances $\sigma_1^2 = 2$, $\sigma_2^2 = 3$, $\sigma_3^2 = 4$. Based on the results of this simulation, we obtain a minimum generalized cross-validation (GCV) value of 2.526286 and three optimal smoothing parameter values, which are $\lambda_{1(opt)} = 2.146156 \times 10^{-7}$ (for the first response), $\lambda_{2(opt)} = 1.084013 \times 10^{-7}$ (for the second response), and $\lambda_{3(opt)} = 5.930101 \times 10^{-8}$ (for the third response).

Next, we present an illustration of the effects of the smoothing parameters on the estimation results of the MNR Model by comparing three kinds of different smoothing parameter values, namely, $\lambda_{1(small)} = 10^{-10}$, $\lambda_{2(small)} = 2 \times 10^{-10}$, and $\lambda_{3(small)} = 3 \times 10^{-10}$, which represent small smoothing parameter values; $\lambda_{1(opt)} = 2.146156 \times 10^{-7}$, $\lambda_{2(opt)} = 1.084013 \times 10^{-7}$, and $\lambda_{3(opt)} = 5.930101 \times 10^{-8}$, which represent optimal

smoothing parameter values; and $\lambda_{1(large)} = 10^{-5}$, $\lambda_{2(large)} = 2 \times 10^{-5}$, and $\lambda_{3(large)} = 3 \times 10^{-5}$, which represent large smoothing parameter values. In the following table and figures, we provide the results of this simulation study.

Table 1 shows that the smoothing parameter values of 2.146156×10^{-7}, 1.084013×10^{-7}, and 5.930101×10^{-8} are the optimal smoothing parameter values, as these smoothing parameters have the lowest GCV value (2.526286) among all the others. Thus, according to (28), these smoothing parameter values are the optimal smoothing parameters. We can write them as $\lambda_{1(opt)} = 2.146156 \times 10^{-7}$, $\lambda_{2(opt)} = 1.084013 \times 10^{-7}$, and $\lambda_{3(opt)} = 5.930101 \times 10^{-8}$. The optimal smoothing parameters provide the best estimation results for the MNR model presented in (29).

Table 1. Comparison estimation results of MNR Model in (29) for three kinds of smoothing parameter values.

$n = 100$; $\rho_{12} = 0.6$; $\rho_{13} = 0.7$; $\rho_{23} = 0.8$; $\sigma_1^2 = 2$; $\sigma_2^2 = 3$; $\sigma_3^2 = 4$		
Smoothing Parameters	**Minimum Values of GCV**	**Results of Estimation**
$\left.\begin{array}{l}\lambda_{1(small)} = 10^{-10} \\ \lambda_{2(small)} = 2 \times 10^{-10} \\ \lambda_{3(small)} = 3 \times 10^{-10}\end{array}\right\}$ (*Small Values*)	4.763234	The estimation results are too rough.
$\left.\begin{array}{l}\lambda_{1(opt)} = 2.146156 \times 10^{-7} \\ \lambda_{2(opt)} = 1.084013 \times 10^{-7} \\ \lambda_{3(opt)} = 5.930101 \times 10^{-8}\end{array}\right\}$ (*Optimal Values*)	2.526286	The estimation results are the best.
$\left.\begin{array}{l}\lambda_{1(large)} = 10^{-5} \\ \lambda_{2(large)} = 2 \times 10^{-5} \\ \lambda_{3(large)} = 3 \times 10^{-5}\end{array}\right\}$ (*Large Values*)	4.617504	The estimation results are too smooth.

The plots of the estimated regression function of the MNR model presented in (29) for the three different smoothing parameters are shown in Figures 1–3.
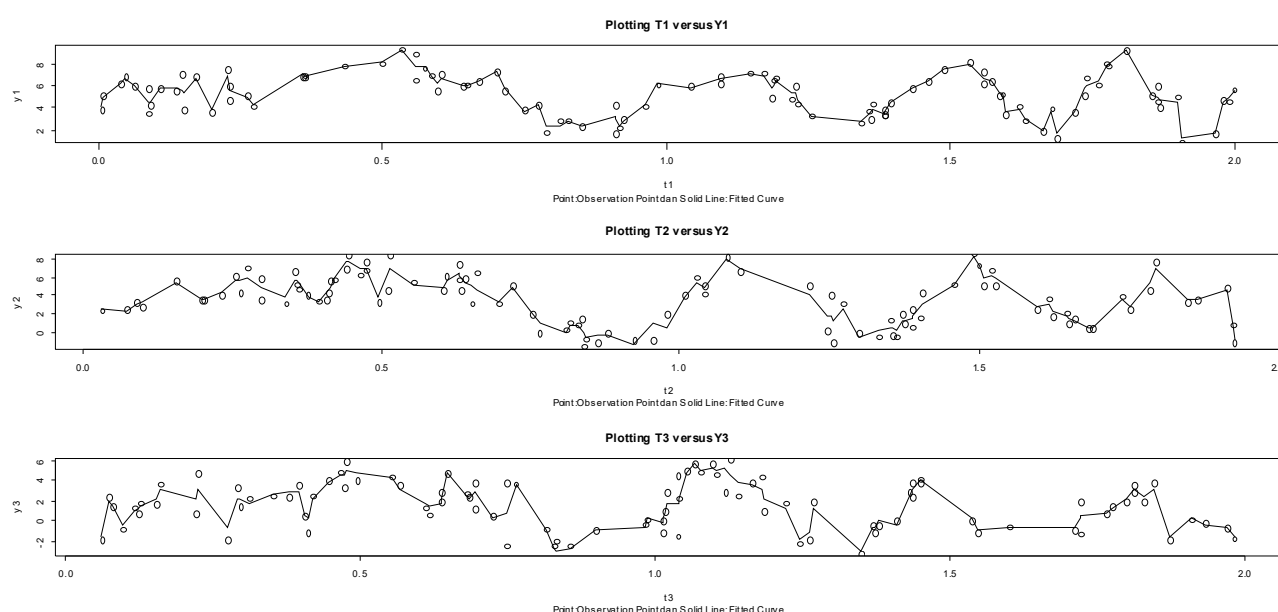

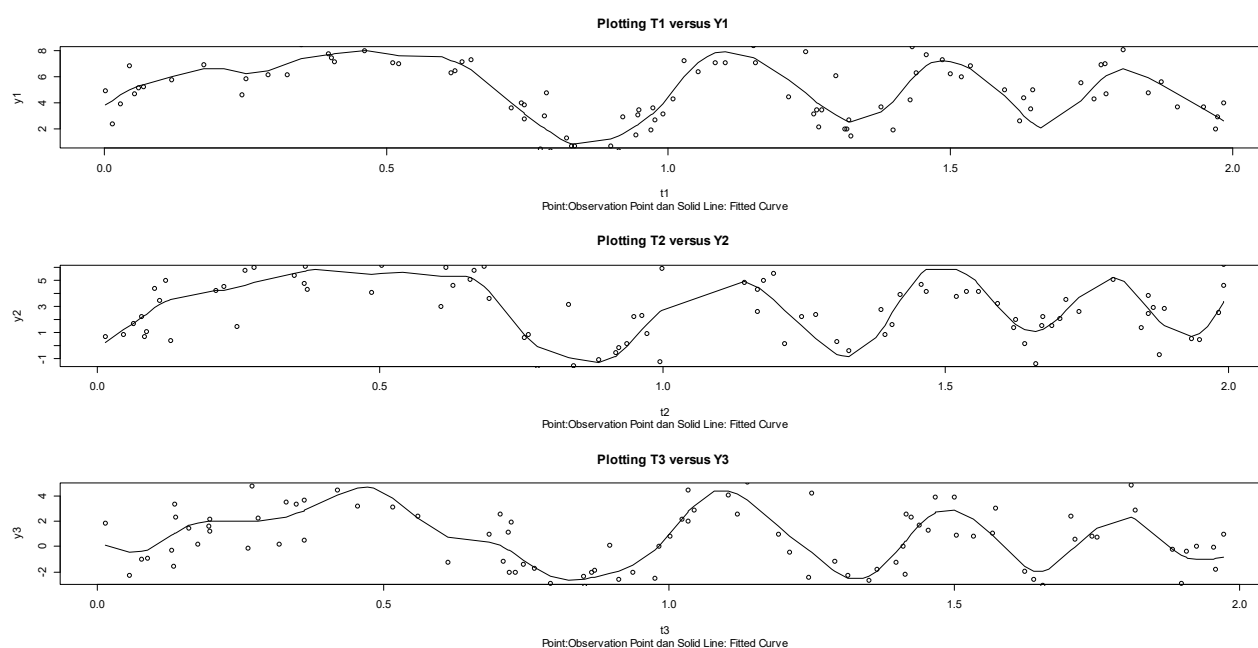**Figure 1.** Plots of estimated MNR Models in (29) for small smoothing parameters.

Figure 2. Plots of estimated MNR Model in (29) for optimal smoothing parameters.
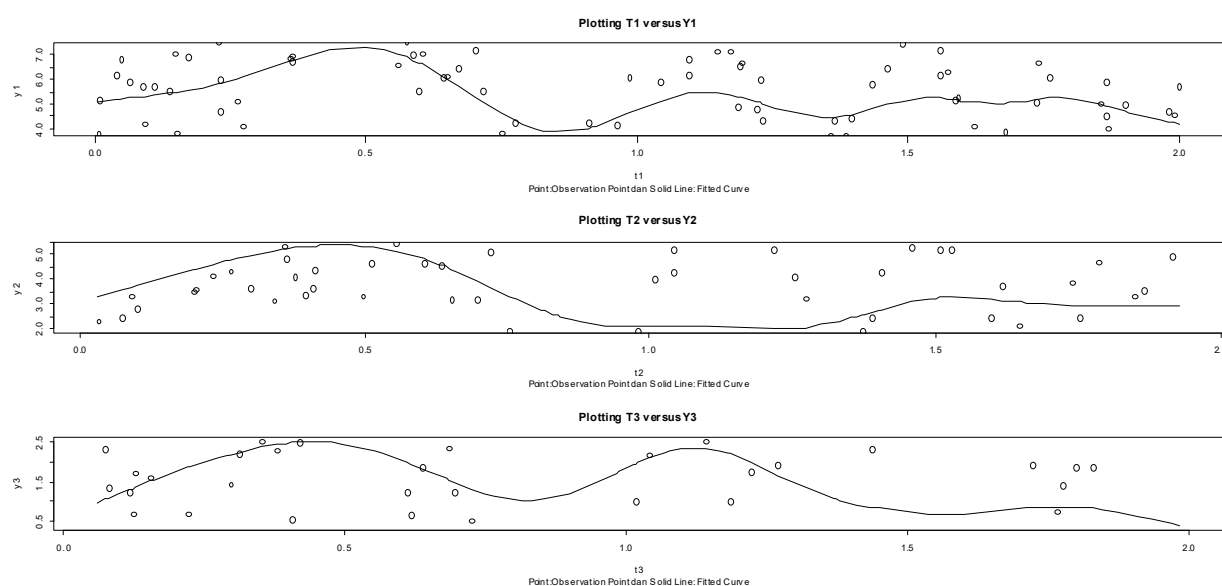

Figure 3. Plots of estimated MNR Models in (29) for large smoothing parameters.

Figure 1 shows that for all responses the small smoothing parameter values provide estimates of the regression functions of the MNR model presented in (29) that are too rough, namely, (y_1) for the first response, (y_2) for the second response, and (y_3) for the third response.

Figure 2 shows that the optimal smoothing parameter values provide the best estimates of the regression functions of the MNR model presented in (29) for all responses, namely, (y_1) for the first response, (y_2) for the second response, and (y_3) for the third response.

Figure 3 shows that for all responses the large smoothing parameter values provide estimates of the regression functions of the MNR model presented in (29) that are too smooth, namely, (y_1) for the first response, (y_2) for the second response, and (y_3) for the third response.

3.5. Investigating the Consistency of the Smoothing Spline Regression Function Estimator

We first investigate the asymptotic properties of the smoothing spline regression function estimator $\hat{\mathbf{f}}$ based on the integrated mean square error (IMSE) criterion. We develop the IMSE proposed by [14] from the uniresponse case to the multiresponse case. Suppose that we decompose the IMSE into two components, $bias^2(\boldsymbol{\lambda})$ and $Var(\boldsymbol{\lambda})$, as follows:

$$IMSE(\boldsymbol{\lambda}) = E \int_a^b [(\mathbf{f(t)} - \hat{\mathbf{f}}(\mathbf{t}))^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \hat{\mathbf{f}}(\mathbf{t}))]\, d\mathbf{t} = bias^2(\boldsymbol{\lambda}) + Var(\boldsymbol{\lambda}) \tag{30}$$

where $bias^2(\boldsymbol{\lambda}) = \int_a^b E[(\mathbf{f(t)} - E\hat{\mathbf{f}}(\mathbf{t}))^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - E\hat{\mathbf{f}}(\mathbf{t}))]d\mathbf{t}$, and $Var(\boldsymbol{\lambda}) = \int_a^b E[(E\hat{\mathbf{f}}(\mathbf{t}) - \hat{\mathbf{f}}(\mathbf{t}))^{\mathbf{T}} \mathbf{W}(E\hat{\mathbf{f}}(\mathbf{t}) - \hat{\mathbf{f}}(\mathbf{t}))]d\mathbf{t}$. Furthermore, in order to investigate the asymptotic property of $bias^2(\boldsymbol{\lambda})$, we assign the solution to PWLS optimization in the following theorem.

Theorem 1. *If* $\hat{\mathbf{f}}(\mathbf{t})$ *is the solution that minimizes the following penalized weighted least square (PWLS):*

$$N^{-1}(\mathbf{y} - \mathbf{g(t)})^{\mathbf{T}} \mathbf{W}(\mathbf{y} - \mathbf{g(t)}) + \sum_{r=1}^p \lambda_r \int_{a_r}^{b_r} \left(g_r^{(m)}(t_r)\right)^2 dt_r \tag{31}$$

then the solution that minimizes the following penalized weighted least square (PWLS):

$$N^{-1}(\mathbf{f(t)} - \mathbf{g(t)})^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \mathbf{g(t)}) + \sum_{r=1}^p \lambda_r \int_{a_r}^{b_r} \left(g_r^{(m)}(t_r)\right)^2 dt_r \tag{32}$$

is $\hat{\mathbf{g}}^*(\mathbf{t}) = E\hat{\mathbf{f}}(\mathbf{t})$.

Proof of Theorem 1. In Section 3.1, we obtained the solution to the PWLS in (31), that is, $\hat{\mathbf{f}} = \mathbf{A}\hat{\mathbf{b}} + \mathbf{C}\hat{\mathbf{d}} = \mathbf{Hy,}$ where as provided in (22), $\mathbf{H} = \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W} + \mathbf{CD}^{-1}\mathbf{W}(\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W})$, $\mathbf{D} = \mathbf{WC} + N\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda} = diag\left(\lambda_1 \mathbf{I}_{n_1}, \lambda_2 \mathbf{I}_{n_2}, \dots, \lambda_p \mathbf{I}_{n_p}\right)$, $N = \sum_{r=1}^p n_r$, and $\mathbf{I}$ is an identity matrix. Next, if we substitute $\mathbf{y} = \mathbf{f(t)}$ into Equation (31), we find that the value that minimizes PWLS (32) is

$$\hat{\mathbf{g}}^*(\mathbf{t}) = \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W} + \mathbf{CD}^{-1}\mathbf{W}(\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{WA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{W})\mathbf{y} = E\hat{\mathbf{f}}(\mathbf{t})$$

Thus, Theorem 1 is proved. □

Furthermore, we investigate the asymptotic property of $bias^2(\boldsymbol{\lambda})$. For this purpose, we first provide the following assumptions.

Assumptions (A):
(A1) For every $r = 1,2, \dots, p$, $n_r = n$ and $\lambda_r = \lambda$.
(A2) For every $i = 1,2, \dots, n$, $t_i = \frac{2i-1}{2n}$.
(A3) For any continuous function g and $0 < W_i = W < \infty$, $i = 1,2, \dots, n$, the following statements are satisfied [48,49]:

    (a) $n^{-1} \sum_{i=1}^n g(t_i) \to \int_a^b g(t)\, dt$, as $n \to \infty$.

    (b) $n^{-1} \sum_{i=1}^n W_i g(t_i) \to \int_a^b Wg(t)dt$, as $n \to \infty$.

    (c) $n^{-1} \sum_{i=1}^n \lambda_i \int_{a_i}^{b_i} \left(g_i^{(m)}(t_i)\right)^2 dt_i \to \lambda \int_a^b g^{(m)}(t)dt$, as $n \to \infty$.

Next, given Assumptions (A), the asymptotic property of $bias^2(\boldsymbol{\lambda})$ is provided in the Theorem 2.

Theorem 2. *If the Assumptions in* (A) *hold, then* $Bias^2(\boldsymbol{\lambda}) \le O(\boldsymbol{\lambda})$, *as* $n \to \infty$.

Proof of Theorem 2. Suppose $\hat{\mathbf{g}}(\mathbf{t})$ is the value which minimizes the following penalized weighted least square (PWLS):

$$\int_a^b (\mathbf{f(t)} - \mathbf{g(t)})^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \mathbf{g(t)}) dt + \sum_{r=1}^p \lambda_r \int_{a_r}^{b_r} \left(g_r^{(m)}(t_r) \right)^2 dt_r$$

then, considering the Assumptions (**A**), we have

$$n^{-1}(\mathbf{f(t)} - \mathbf{g(t)})^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \mathbf{g(t)}) \approx \int_a^b (\mathbf{f(t)} - \mathbf{g(t)})^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \mathbf{g(t)}) dt, \text{ as } n \to \infty.$$

Hence, we obtain $\hat{\mathbf{g}}^*(\mathbf{t}) = E\hat{\mathbf{f}}(\mathbf{t}) \approx \hat{\mathbf{g}}(\mathbf{t})$. Thus, for every $\mathbf{g} \in W_2^m[a, b]$ we have

$$Bias^2(\boldsymbol{\lambda}) = \int_a^b E[(\mathbf{f(t)} - E\hat{\mathbf{f}}(\mathbf{t}))^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - E\hat{\mathbf{f}}(\mathbf{t}))] dt$$

$$\leq \int_a^b E[(\mathbf{f(t)} - E\hat{\mathbf{f}}(\mathbf{t}))^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - E\hat{\mathbf{f}}(\mathbf{t}))] dt + \sum_{r=1}^p \lambda_r \int_a^b \hat{g}_r^{(m)}(t_r) dt_r$$

Because $E\hat{\mathbf{f}}(\mathbf{t}) \approx \hat{\mathbf{g}}(\mathbf{t})$, we obtain

$$Bias^2(\boldsymbol{\lambda}) \leq \int_a^b E[(\mathbf{f(t)} - \hat{\mathbf{g}}(\mathbf{t}))^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \hat{\mathbf{g}}(\mathbf{t}))] dt + \sum_{r=1}^p \lambda_r \int_a^b \hat{g}_r^{(m)}(t_r) dt_r$$

Thus, we have the following relationship:

$$Bias^2(\boldsymbol{\lambda}) \leq \int_a^b E[(\mathbf{f(t)} - \mathbf{g(t)})^{\mathbf{T}} \mathbf{W}(\mathbf{f(t)} - \mathbf{g(t)})] dt + \sum_{r=1}^p \lambda_r \int_a^b g_r^{(m)}(t_r) dt_r \quad (33)$$

for every $\mathbf{g} \in W_2^m[a, b]$.

Because every $\mathbf{g} \in W_2^m[a, b]$ satisfies the relationship in (33), by taking $\mathbf{g(t)} = \mathbf{f(t)}$, we have

$$Bias^2(\boldsymbol{\lambda}) \leq \sum_{r=1}^p \lambda_r \int_a^b g_r^{(m)}(t_r) dt_r = O(\boldsymbol{\lambda}) \text{ as } n \to \infty \quad (34)$$

where $O(\cdot)$ represents "big oh". Details about "big oh" can be found in [14,50]. Thus, Theorem 2 is proved.□

Furthermore, the asymptotic property of $Var(\boldsymbol{\lambda})$ is provided in Theorem 3.

Theorem 3. *If Assumptions* (A) *hold and* $r = 1,2, \dots, p$, *then* $Var(\lambda_r) \leq O\left(\dfrac{1}{n \, \lambda_r^{\frac{1}{2m}}}\right)$.

Proof of Theorem 3. To investigate the asymptotic property of $Var(\boldsymbol{\lambda})$, we first define the following function:

$$\boldsymbol{\omega}(\hat{\mathbf{f}}, \mathbf{h}, \alpha) = R(\hat{\mathbf{f}} + \alpha\mathbf{h}) + \sum_{r=1}^p \lambda_r J_r (\hat{\mathbf{f}} + \alpha\mathbf{h})$$

where $R(\mathbf{g}) = n^{-1}(\mathbf{y} - \mathbf{g(t)})^T \mathbf{W}(\mathbf{y} - \mathbf{g(t)})$; $J_r(\mathbf{g}) = \int_a^b [g_r^{(m)}(t_r)]^2 dt_r$; $\mathbf{h} \in W_2^m[a, b]$ and $\alpha \in \mathbb{R}$.

Hence, for any $\mathbf{f}, \mathbf{g} \in W_2^m[a, b]$ we have

$$\boldsymbol{\omega}(\hat{\mathbf{f}}, \mathbf{h}, \alpha) = n^{-1}(\mathbf{y} - \mathbf{f(t)} - \alpha\mathbf{g(t)})^T \mathbf{W}(\mathbf{y} - \mathbf{f(t)} - \alpha\mathbf{g(t)}) +$$

$$\sum_{r=1}^p \lambda_r \int_{a_r}^{b_r} \left(f_r^{(m)}(t_r) + \alpha g_r^{(m)}(t_r) \right)^2 dt_r$$

Next, by taking the solution to $\dfrac{\partial \boldsymbol{\omega}(\hat{\mathbf{f}}, \mathbf{h}, \alpha)}{\partial \alpha} = \mathbf{0}$, for $\alpha = 0$ we have

$$n^{-1}(\mathbf{y} - \mathbf{f(t)})^T \mathbf{W}\mathbf{g(t)} = \sum_{r=1}^p \lambda_r \int_{a_r}^{b_r} \left(f_r^{(m)}(t_r) g_r^{(m)}(t_r) \right)^2 dt_r$$

Furthermore, if $\gamma_1, \gamma_2, \dots, \gamma_n$ are the bases for the natural spline and $f_r(t_r) = \sum_{k=1}^n \beta_{rk} \gamma_{rk}(t_r)$, $r = 1,2, \dots, p$, then according to [51] this implies that

$$\sum_{j=1}^n \left(y_{rj} - \sum_{k=1}^n \beta_{rk} \gamma_{rk}(t_{rj}) \right) W_{rj} g_{rj}(t_{rj})$$
$$= n\lambda_r (-1)^m (2m - 1)! \sum_{i=1}^n g_r(t_{ri}) \sum_{k=1}^n \beta_{rk} d_{rik} \quad (35)$$

where d_{rik} is the rth response diagonal element of matrix **H** in (26).

Now, because Equation (35) holds for every $g_r \in W_2^m[a_r, b_r]$, $r = 1,2, \dots, p$, it follows that determining the solution to Equation (35) is equivalent to determining the value of β_{rk} that satisfies the following equation:

$$y_{ri} = \sum_{k=1}^{n}(n\lambda_r(-1)^m(2m-1)!\, W_{ri}^{-1}d_{rik} + \gamma_{rk}(t_{ri}))\beta_{rk} \ , \ i = 1,2,\dots,n \qquad (36)$$

We can express Equation (36) in matrix notation as follows:

$$\mathbf{y} = (n\lambda_r(-1)^m(2m-1)!\,\mathbf{W}^{-1}\mathbf{K} + \boldsymbol{\gamma})\boldsymbol{\beta} \qquad (37)$$

where $= \{d_{rik}\}$, $\boldsymbol{\gamma} = \{\gamma_{rk}(t_{ri})\}$, $r = 1,2,\dots,p$, and $i,k = 1,2,\dots,n$.

Hence, we obtain the estimator for $\boldsymbol{\beta}$ in Equation (37) as follows:

$$\widehat{\boldsymbol{\beta}} = diag\left(\frac{1}{1 + n\,\lambda_r\,_{\theta_1}}, \dots, \frac{1}{1 + n\,\lambda_r\,_{\theta_n}}\right)\boldsymbol{\gamma}^{\mathbf{T}}\mathbf{y}$$

where $\{\theta_1, \dots, \theta_n\} \in \mathcal{H}$ (here, $\mathcal{H}$ is an RKHS) and is perpendicular to $\boldsymbol{\gamma}$.

Thus, the estimator for the regression function $\mathbf{f(t)}$ can be expressed as follows:

$$\hat{\mathbf{f}}(\mathbf{t}) = \boldsymbol{\gamma}(\mathbf{t})\widehat{\boldsymbol{\beta}} = \sum_{j=1}^{n}\frac{1}{1+n\lambda\theta_j}\boldsymbol{\gamma}_j^T\mathbf{y}\,\boldsymbol{\gamma}_j(\mathbf{t}) \qquad (38)$$

Hence, for $r = 1,2,\dots,p$ Equation (38) results in

$$\hat{f}_r(t_r) = \sum_{j=1}^{n}\frac{1}{1+n\lambda_r\theta_{rj}}\boldsymbol{\gamma}_{rj}^T\mathbf{y}\,\boldsymbol{\gamma}_{rj}(t_r)$$

$$= \sum_{j=1}^{n}\frac{1}{1+n\lambda_r\theta_{rj}}\sum_{k=1}^{n}\gamma_{rj}(t_{rk})y_{rk}\,\gamma_{rj}(t_{rk})$$

Thus, for $r = 1,2,\dots,p$ we have

$$Var\big(\hat{f}_r(t_r)\big) = \sigma_r^2\sum_{j=1}^{n}\frac{\gamma_{rj}^2(t_r)}{\left(1+n\,\lambda_r\,\theta_{rj}\right)^2}\sum_{k=1}^{n}\gamma_{rj}(t_{rk})\,\gamma_{rj}(t_{rk})W_{rk}^{-1}$$

$$\leq \left(\max_{1\leq i\leq n}\{W_{ri}^{-1}\}\right)\sigma_r^2\sum_{j=1}^{n}\frac{\gamma_{rj}^2(t_r)}{\left(1+n\,\lambda_r\,\theta_{rj}\right)^2}$$

From this step, for $r = 1,2,\dots,p$ we have

$$Var(\lambda_r) = \int_{a_r}^{b_r} E\left[\left(E\hat{f}_r(t_r) - \hat{f}_r(t_r)\right)^T W_r\left(E\hat{f}_r(t_r) - \hat{f}_r(t_r)\right)\right]dt_r$$

$$\leq \left(\max_{1\leq i\leq n}\{W_{ri}^{-1}\}\right)\sigma_r^2\sum_{j=1}^{n}\frac{\gamma_{rj}^2(t_r)}{\left(1+n\,\lambda_r\,\theta_{rj}\right)^2}\int_{a_r}^{b_r}\gamma_{rj}^2(t_r)W_r dt_r$$

In the next step, Refs. [51,52] provide an approximation for $n \to \infty$ as follows:

$$n^{-1} = n^{-1}\sum_{j=1}^{n}W_{rj}\gamma_{rj}^2(t_{rj}) \approx \int_{a_r}^{b_r}\gamma_r^2(t_r)W_r dt_r, \text{ and}$$

$$Var(\lambda_r) \leq \left(\max_{1\leq i\leq n}\{W_{ri}^{-1}\}\right)\sigma_r^2 n^{-1}\sum_{j=1}^{n}\frac{1}{\left(1+\lambda_r\,\gamma_j\right)^2}.$$

Furthermore, Ref. [51] leads to the following result:

$$Var(\lambda_r) \leq \left(\max_{1\leq i\leq n}\{W_{ri}^{-1}\}\right)\sigma_r^2 n^{-1}\sum_{j=1}^{n}\frac{1}{(1+\lambda_r\,(\pi j)^{2m})^2}$$

Next, using integral approximation [51], for $r = 1,2,\dots,p$ we have

$$Var(\lambda_r) \leq \left(\max_{1\leq i\leq n}\{W_{ri}^{-1}\}\right)\frac{\sigma_r^2}{n\,\pi\,\lambda_r^{\frac{1}{2m}}}\int_0^{\infty}\frac{dx}{(1+x^{2m})^2} = \frac{1}{n\,\lambda_r^{\frac{1}{2m}}}K(m,\sigma) = O\left(\frac{1}{n\,\lambda_r^{\frac{1}{2m}}}\right)$$

where $(m,\sigma) = \frac{\sigma_r^2}{\pi}\left(\max_{1\leq i\leq n}\{W_{ri}^{-1}\}\right)\int_0^{\infty}\frac{dx}{(1+x^{2m})^2}$. Thus, Theorem 3 is proved. □

Here, based on Theorems 2 and 3, we obtain the asymptotic property of the smoothing spline regression function estimator of the MNR model presented in (1) based on the integrated mean square error (IMSE) criterion, as follows:

$$IMSE(\boldsymbol{\lambda}) = Bias^2(\boldsymbol{\lambda}) + Var(\boldsymbol{\lambda}) \leq O(\boldsymbol{\lambda}) + O(\mathbf{z}) \text{ as } n \to \infty \qquad (39)$$

where $= (\lambda_1, \lambda_2, \dots, \lambda_p)^{\mathrm{T}}$ and $\mathbf{z} = \begin{pmatrix} \dfrac{1}{n\,\lambda_1^{\frac{1}{2m}}} & \dfrac{1}{n\,\lambda_2^{\frac{1}{2m}}} & \cdots & \dfrac{1}{n\,\lambda_p^{\frac{1}{2m}}} \end{pmatrix}^{T}$.

The consistency of the smoothing spline regression function estimator of the MNR model presented in (1) is provided by the following theorem.

Theorem 4. *If* $\hat{\mathbf{f}}(\mathbf{t})$ *is a smoothing spline estimator for regression function* $\mathbf{f}(\mathbf{t})$ *of the MNR model presented in (1), then* $\hat{\mathbf{f}}(\mathbf{t})$ *is a consistent estimator for* $\mathbf{f}(\mathbf{t})$ *based on the integrated mean square error (IMSE) criterion.*

Proof of Theorem 4. Based on Equation (39), we have the following relationship:

$$IMSE(\boldsymbol{\lambda}) = Bias^2(\boldsymbol{\lambda}) + Var(\boldsymbol{\lambda}) \le O(\boldsymbol{\lambda}) + O(\mathbf{z}) \le O(n\boldsymbol{\lambda}) \text{ as } n \to \infty.$$

Hence, according to [48], for any small positive number $\delta > 0$ we have

$$\lim_{n \to \infty} P\left(\left| \frac{IMSE(\boldsymbol{\lambda})}{n\boldsymbol{\lambda}} \right| > \delta \right) \le \lim_{n \to \infty} P(|IMSE(\boldsymbol{\lambda})| > \delta) = 0. \tag{40}$$

Because $P(|IMSE(\boldsymbol{\lambda})| > \delta) = 1 - P(|IMSE(\boldsymbol{\lambda})| \le \delta)$, based on Equation (40) and applying the probability properties we have

$$\lim_{n \to \infty}(1 - P(|IMSE(\boldsymbol{\lambda})| \le \delta)) = 0 \iff 1 - \lim_{n \to \infty} P(|IMSE(\boldsymbol{\lambda})| \le \delta) = 0$$

$$\iff \lim_{n \to \infty} P(|IMSE(\boldsymbol{\lambda})| \le \delta) = 1. \tag{41}$$

Equation (41) means that the smoothing spline regression function estimator of the multiresponse nonparametric regression (MNR) model is a consistent estimator based on the integrated mean square error (IMSE) criterion. Thus, Theorem 4 is proved. □

3.6. Illustration of Theorems

Suppose a paired observation (y_{ri}, t_{ri}) follows the multiresponse nonparametric regression (MNR) model:

$$y_{ri} = f_r(t_{ri}) + \varepsilon_{ri}, t_{ri} \in [0,1], r = 1,2,\dots,p, \qquad i = 1,2,\dots,n_r. \tag{42}$$

Next, for every $r = 1,2,\dots,p$, we assume $f_1 = f_2 = \cdots = f_p = f$ and $\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_p = \varepsilon$ such that $y_1 = y_2 = \cdots = y_p = y$. Hence, we have a nonparametric regression model as follows:

$$y_i = f(t_i) + \varepsilon_i, t_i \in [0,1], i = 1,2,\dots,n. \tag{43}$$

Based on the model presented in (43), we present an illustration related to the four theorems in Section 3.5 through a simulation study with sample size of $\boldsymbol{n = 200}$. Based on this model, let $\boldsymbol{f(t) = sin^3(2\pi t^3)}$, where $\boldsymbol{t}$ is generated from a uniform (0,1) distribution and $\boldsymbol{\varepsilon}$ is generated from a standard normal distribution. The first step is to create plots of the observation values $(\boldsymbol{t}, \boldsymbol{y})$ and $\boldsymbol{f(t)}$, as shown in Figure 4.
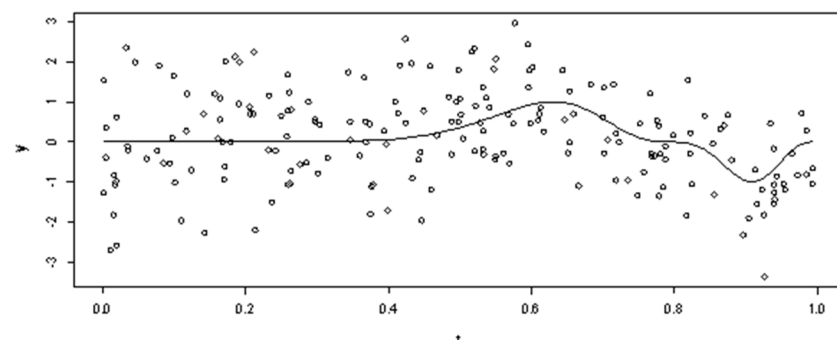
Figure 4. Plots of observation values $(\boldsymbol{t}, \boldsymbol{y})$ and $\boldsymbol{f(t)}$.

It can be seen from Figure 4 that there is a tendency towards y variance inequality. For larger values of t, the y variance tends to be larger. Next, the data model is approximated by a weighted spline with a weight of $w_i = 1/t_i^2$, $i = 1,2,\dots,200$. The next step is to determine the order, number, and location of the knot points. Here, we use a weighted cubic spline model with two knot points, namely, 0.5 and 0.785. This weighted cubic spline model can be written as follows:

$$E(y) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t - 0.5)_+^3 + \beta_5 (t - 0.785)_+^3 \tag{44}$$

The plot of this weighted cubic spline with two knot points is shown in Figure 5. The plots of the residuals and the estimated weighted cubic spline are shown in Figure 6.
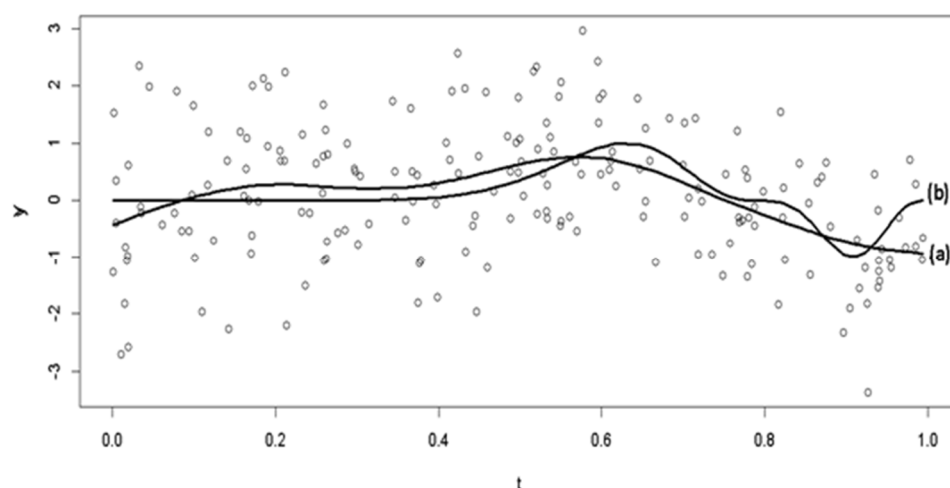
Figure 5. Plots of (**a**) weighted cubic spline with two knot points and (**b**) curve of $f(t)$.
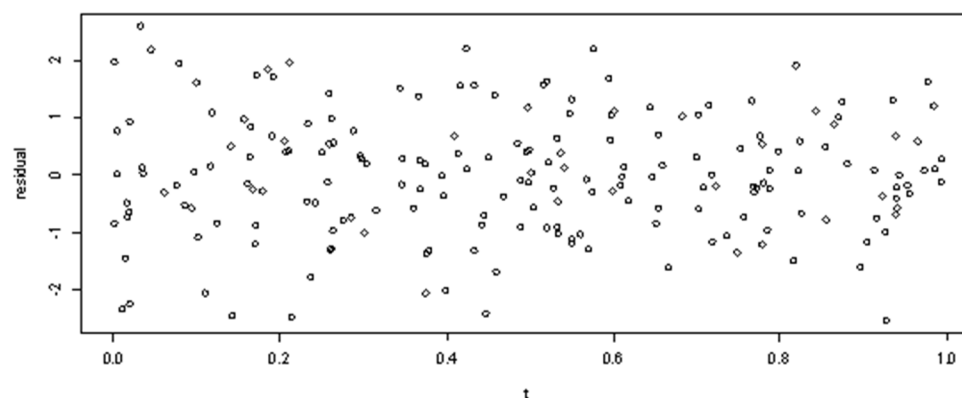
Figure 6. Plots of residuals and estimation values of weighted cubic spline.

Figure 6 shows that with weight $w_i = 1/t_i^2$, the variance of the response variable y tends to be the same. Meanwhile, Figure 7 provides a residual normality plot for the weighted cubic spline model. The plot in Figure 7 shows no indication towards deviation from the normal distribution.

Next, as a comparison, we investigate a weighted cubic polynomial model with weight $w_i = 1/t_i^2$, $i = 1,2,\dots,200$ for fitting the model (43). The fitting of the weighted cubic polynomial model is shown in Figure 8. From the visualization in Figure 8, it can be seen that this weighted cubic polynomial approach tends to approach the function $f(t)$ very globally. This is in contrast to the weighted cubic spline with two knot points in (44), which approaches the function $f(t)$ more locally. Thus, the weighted cubic spline model with two knot points is adequate as an approximation model for the model presented in (43).
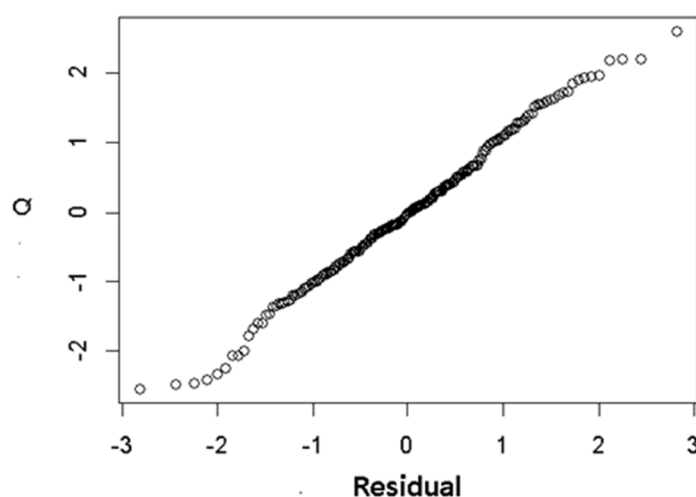
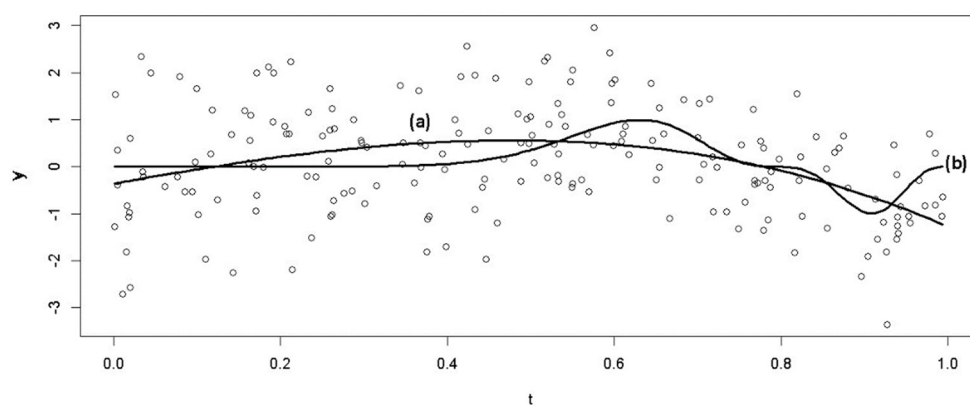**Figure 7.** Plot of weighted cubic spline residual normality.

Figure 8. Plots of (**a**) weighted cubic polynomial and (**b**) curve $f(t)$.

Based on the illustration above and Figures 4–8, if $\hat{f}(t)$ is an estimator for model (43), that is, if $\hat{f}(t)$ is the Penalized Weighted Least Squares (PWLS) solution, then $\hat{g}^*(t) = E(y)$, from Equation (44) is an estimator for model (43) as well, such that $\hat{g}^*(t) = E[\hat{f}(t)]$, as provided by Theorem 1.

The plots of the asymptotic curves of $\boldsymbol{IMSE}(\boldsymbol{\lambda})$, $\boldsymbol{Bias}^2(\boldsymbol{\lambda})$, and $\boldsymbol{Var}(\boldsymbol{\lambda})$, where $\boldsymbol{IMSE}(\boldsymbol{\lambda}) = \boldsymbol{Bias}^2(\boldsymbol{\lambda}) + \boldsymbol{Var}(\boldsymbol{\lambda})$, are shown in Figure 9 [53].
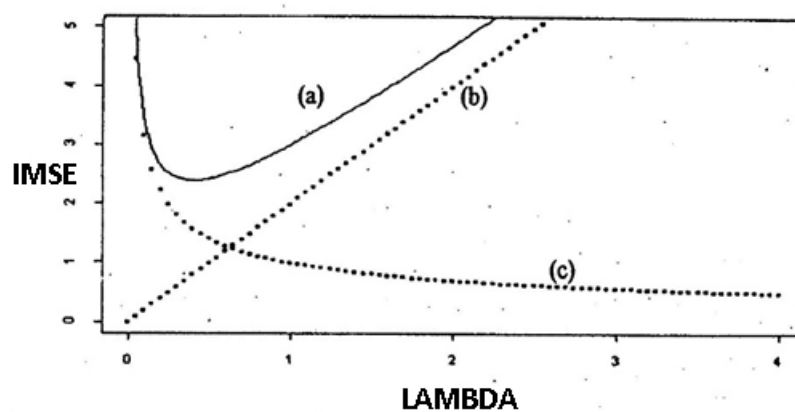
Figure 9. Plots of asymptotic curves of (**a**) $\boldsymbol{IMSE}(\boldsymbol{\lambda})$, (**b**) $\boldsymbol{Bias}^2(\boldsymbol{\lambda})$, and (**c**) $\boldsymbol{Var}(\boldsymbol{\lambda})$.

Figure 9 shows that the Integrated Mean Square Error ($IMSE(\lambda)$) curve represented by curve (a) is the sum of the quadratic bias ($\textbf{Bias}^2(\boldsymbol{\lambda})$) curve represented by curve (b) and the variance curve ($\textbf{Var}(\boldsymbol{\lambda})$) represented by curve (c). It can be seen from Figure 9b that $Bias^2(\boldsymbol{\lambda}) \leq O(\boldsymbol{\lambda})$, as $n \to \infty$, that is, $\underset{n\to\infty}{\text{Lim Sup}} \left| \frac{Bias^2(\boldsymbol{\lambda})}{\boldsymbol{\lambda}} \right| < \infty$ [14,50,54]. This means that $\left| \frac{Bias^2(\boldsymbol{\lambda})}{\boldsymbol{\lambda}} \right|$ remains bounded, as $n \to \infty$, which is provided by Theorem 2. Furthermore, Figure 9c shows that $Var(\boldsymbol{\lambda}) \leq O\left(\frac{1}{n\lambda^{\frac{1}{2m}}} \right)$ (that is, $\underset{n\to\infty}{\text{Lim Sup}} \left| \left(n\lambda^{\frac{1}{2m}} \right)^{-1} / \boldsymbol{\lambda} \right| < \infty$ [14,50,54]. This means that $\left| \left(n\lambda^{\frac{1}{2m}} \right)^{-1} / \boldsymbol{\lambda} \right|$ remains bounded, as $n \to \infty$, which is provided by Theorem 3. Furthermore, Figure 9 shows that $IMSE(\boldsymbol{\lambda}) = Bias^2(\boldsymbol{\lambda}) + Var(\boldsymbol{\lambda}) \leq O(n\lambda)$. According to [14,50,54], this means that $\underset{n\to\infty}{\text{Lim Sup}} \left| \frac{IMSE(\boldsymbol{\lambda})}{n\lambda} \right| < \infty$. In other words, $IMSE(\boldsymbol{\lambda}) \leq O(n\lambda)$ if $\left| \frac{IMSE(\boldsymbol{\lambda})}{n\lambda} \right|$ remains bounded, as $n \to \infty$. According to [14,50,54], for any small positive number $\delta > 0$ we have $\underset{n\to\infty}{\text{Lim}} P\left(\left| \frac{IMSE(\boldsymbol{\lambda})}{n\lambda} \right| > \delta \right) \leq \underset{n\to\infty}{\text{Lim}} P(|IMSE(\boldsymbol{\lambda})| > \delta) = \mathbf{0}$; hence, $\underset{n\to\infty}{\text{Lim}} P(|IMSE(\boldsymbol{\lambda})| \leq \delta) = 1$ is consistent.

4. Conclusions

The smoothing spline estimator with the RKHS approach has good ability to estimate the MNR model, which is a nonparametric regression model where the responses are correlated with each other, because the goodness of fit and smoothness of the estimation curve is controlled by the smoothing parameter, making the estimator very suitable for prediction purposes. Therefore, selection of the optimal smoothing parameter value cannot be omitted, and is crucial to good regression function fitting of the MNR model based on smoothing spline estimator using the RKHS approach. The estimator of the smoothing spline regression function of the MNR model that we obtained is linear with respect to the observations in Equation (22), and is a consistent estimator based on the integrated mean square error (IMSE) criterion. The main influence of this study is lies in the easier estimation of a multiresponse nonparametric regression model where there is a correlation between responses using the RKHS approach based on a smoothing spline estimator. This approach is easier, and faster, and more optimal, as estimation is carried out simultaneously instead of response by response for each observation. In addition, the theory generated in this study can be used to estimate the nonparametric component of the multiresponse semiparametric regression model used to model Indonesian toddlers' standard growth charts. In the future, the results of this study can be further developed within the scope of statistical inference, especially for the purpose of testing hypotheses involving multiresponse nonparametric regression models and multiresponse semiparametric regression models.

References

1. Wahba, G. *Spline Models for Observational Data*; SIAM: Philadelphia, Pennsylvania, 1990.
2. Kimeldorf, G.; Wahba, G. Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **1971**, *33*, 82–95.
3. Cox, D.D. Asymptotics for M-type smoothing spline. *Ann. Stat.* **1983**, *11*, 530–551.
4. Oehlert, G.W. Relaxed boundary smoothing spline. *Ann. Stat.* **1992**, *20*, 146–160.
5. Ana, E.; Chamidah, N.; Andriani, P.; Lestari, B. Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression. *J. Phys. Conf. Ser.* **2019**, *1397*, 012067.
6. Chamidah, N.; Yonani, Y.S.; Ana, E.; Lestari, B. Identification the number of mycobacterium tuberculosis based on sputum image using local linear estimator. *Bullet. Elect. Eng. Inform.* **2020**, *9*, 2109–2116.
7. Cheruiyot, L.R. Local linear regression estimator on the boundary correction in nonparametric regression estimation. *J. Stat. Theory Appl.* **2020**, *19*, 460–471.
8. Cheng, M.-Y.; Huang, T.; Liu, P.; Peng, H. Bias reduction for nonparametric and semiparametric regression models. *Stat. Sin.* **2018**, *28*, 2749–2770.
9. Chamidah, N.; Zaman, B.; Muniroh, L.; Lestari, B. Designing local standard growth charts of children in East Java province using a local linear estimator. *Int. J. Innov. Creat. Chang.* **2020**, *13*, 45–67.
10. Delaigle, A.; Fan, J.; Carroll, R.J. A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Amer. Stat. Assoc.* **2009**, *104*, 348–359.
11. Francisco-Fernandez, M.; Vilar-Fernandez, J.M. Local polynomial regression estimation with correlated errors. *Comm. Statist. Theory Methods* **2001**, *30*, 1271–1293.
12. Benhenni, K.; Degras, D. Local polynomial estimation of the mean function and its derivatives based on functional data and regular designs. *ESAIM Probab. Stat.* **2014**, *18*, 881–809.
13. Kikechi, C.B. On local polynomial regression estimators in finite populations. *Int. J. Stat. Appl. Math.* **2020**, *5*, 58–63.
14. Wand, M.P.; Jones, M.C. *Kernel Smoothing*; Chapman & Hall: London, UK, 1995.
15. Cui, W.; Wei, M. Strong consistency of kernel regression estimate. *Open J. Stats.* **2013**, *3*, 179–182.
16. De Brabanter, K.; De Brabanter, J.; Suykens, J.A.K.; De Moor, B. Kernel regression in the presence of correlated errors. *J. Mach. Learn. Res.* **2011**, *12*, 1955–1976.
17. Chamidah, N.; Lestari, B. Estimating of covariance matrix using multi-response local polynomial estimator for designing children growth charts: A theoretically discussion. *J. Phy. Conf. Ser.* **2019**, *1397*, 012072.
18. Aydin, D.; Güneri, Ö.I.; Fit, A. Choice of bandwidth for nonparametric regression models using kernel smoothing: A simulation study. *Int. J. Sci. Basic Appl. Res.* **2016**, *26*, 47–61.
19. Eubank, R.L. *Nonparametric Regression and Spline Smoothing*, 2nd ed.; Marcel Dekker: New York, NY, USA, 1999.
20. Wang, Y. *Smoothing Splines: Methods and Applications*; Taylor & Francis Group: Boca Raton, FL, USA, 2011.
21. Liu, A.; Qin, L.; Staudenmayer, J. M-type smoothing spline ANOVA for correlated data. *J. Multivar. Anal.* **2010**, *101*, 2282–2296.
22. Gao, J.; Shi, P. M-Type smoothing splines in nonparametric and semiparametric regression models. *Stat. Sin.* **1997**, *7*, 1155–1169.
23. Chamidah, N.; Lestari, B.; Massaid, A.; Saifudin, T. Estimating mean arterial pressure affected by stress scores using spline nonparametric regression model approach. *Commun. Math. Biol. Neurosci.* **2020**, *2020*, 1–12.
24. Fatmawati, I.N.; Budiantara, B.L. Comparison of smoothing and truncated spline estimators in estimating blood pressures models. *Int. J. Innov. Creat. Change* **2019**, *5*, 1177–1199.
25. Chamidah, N.; Lestari, B.; Budiantara, I.N.; Saifudin, T.; Rulaningtyas, R.; Aryati, A.; Wardani, P.; Aydin, D. Consistency and asymptotic normality of estimator for parameters in multiresponse multipredictor semiparametric regression model. *Symmetry* **2022**, *14*, 336.
26. Eilers, P.H.C.; Marx, B.D. Flexible smoothing with B-splines and penalties. *Statist. Sci.* **1996**, *11*, 86–121.
27. Lu, M.; Liu, Y.; Li, C.-S. Efficient estimation of a linear transformation model for current status data via penalized splines. *Stat. Meth. Medic. Res.* **2020**, *29*, 3–14.

28. Wang, Y.; Guo, W.; Brown, M.B. Spline smoothing for bivariate data with applications to association between hormones. *Stat. Sin.* **2000**, *10*, 377–397.

29. Yilmaz, E.; Ahmed, S.E.; Aydin, D. A-Spline regression for fitting a nonparametric regression function with censored data. *Stats* **2020**, *3*, 120–136.

30. Aydin, D. A comparison of the nonparametric regression models using smoothing spline and kernel regression. *World Acad. Sci. Eng. Technol.* **2007**, *36*, 253–257.

31. Lestari, B; Budiantara, I.N.; Chamidah, N. Smoothing parameter selection method for multiresponse nonparametric regression model using spline and kernel estimators approaches. *J. Phy. Conf. Ser.* **2019**, *1397*, 012064.

32. Lestari, B; Budiantara, I.N.; Chamidah, N. Estimation of regression function in multiresponse nonparametric regression model using smoothing spline and kernel estimators. *J. Phy. Conf. Ser.* **2018**, *1097*, 012091.

33. Osmani, F.; Hajizadeh, E.; Mansouri, P. Kernel and regression spline smoothing techniques to estimate coefficient in rates model and its application in psoriasis. *Med. J. Islam. Repub. Iran* **2019**, *33*, 1–5.

34. Lestari, B; Budiantara, I.N. Spline estimator and its asymptotic properties in multiresponse nonparametric regression model. *Songklanakarin J. Sci. Technol.* **2020**, *42*, 533–548.

35. Mariati, M.P.A.M.; Budiantara, I.N.; Ratnasari, V. The application of mixed smoothing spline and Fourier series model in nonparametric regression. *Symmetry* **2021**, *13*, 2094.

36. Wang, Y.; Ke, C. Smoothing spline semiparametric nonlinear regression models. *J. Comp. Graphical Stats.* **2009**, *18*, 165–183.

37. Lestari, B.; Chamidah, N. Estimating regression function of multiresponse semiparametric regression model using smoothing spline. *J. Southwest Jiaotong Univ.* **2020**, *55*, 1–9.

38. Gu, C. *Smoothing Spline ANOVA Models*; Springer: New York, NY, USA, 2002.

39. Aronszajn, N. Theory of reproducing kernels. *Transact. Amer. Math. Soc.* **1950**, *68*, 337–404.

40. Berlinet, A.; Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*; Kluwer Academic: Norwell, MA., USA, 2004.

41. Wahba, G. Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. *Adv. Kernel Methods-Support Vector Learn.* **1999**, *6*, 69–87.

42. Scholkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA., USA, 2002.

43. Zeng, X.; Xia, Y. Asymptotic distribution for regression in a symmetric periodic Gaussian kernel Hilbert space. *Stat. Sin.* **2019**, *29*, 1007–1024.

44. Hofmann, T.; Scholkopf, B.; Smola, A.J. Kernel methods in machine learning. *Ann. Stat.* **2008**, *36*, 1171–1220.

45. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice Hall: New York, NY, USA, 1982.

46. Li, J.; Zhang, R. Penalized spline varying-coefficient single-index model. *Commun. Stat.-Simul. Comp.* **2010**, *39*, 221–239.

47. Ruppert, D.; Carroll, R. *Penalized Regression Splines, Working Paper*; School of Operation Research and Industrial Engineering, Cornell University: New York, NY, USA, 1997.

48. Tunç, C.; Tunç, O. On the stability, integrability and boundedness analyses of systems of integro-differential equations with time-delay retardation. *RACSAM* **2021**, *115*, 115.

49. Aydin, A.; Korkmaz, E. Introduce Gâteaux and Frêchet derivatives in Riesz spaces. *Appl. Appl. Math. Int. J.* **2020**, *15*, 16.

50. Sen, P.K.; Singer, J.M. *Large Sample in Statistics: An Introduction with Applications*; Chapman & Hall: London, UK, 1993.

51. Eubank, R.L. *Spline Smoothing and Nonparametric Regression*; Marcel Dekker, Inc.: New York, NY, USA, 1988.

52. Speckman, P. Spline smoothing and optimal rate of convergence in nonparametric regression models. *Ann. Stat.* **1985**, *13*, 970–983.

53. Budiantara, I.N. *Estimator Spline dalam Regresi Nonparametrik dan Semiparametrik*; Disertasi, Universitas Gadjah Mada: Yogyakarta, Indonesia, 2000.

54. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*; John Wiley: New York, NY, USA, 1980.