

Article



Detection of Interaction Effects in a Nonparametric Concurrent Regression Model

Rui Pan¹, Zhanfeng Wang^{2,*} and Yaohua Wu²

- ¹ School of Data Science, University of Science and Technology of China, Hefei 230026, China; kevinpan@mail.ustc.edu.cn
- ² Department of Statistics and Finance, Management School, University of Science and Technology of China, Hefei 230026, China; wyh@ustc.edu.cn
- * Correspondence: zfw@ustc.edu.cn

Abstract: Many methods have been developed to study nonparametric function-on-function regression models. Nevertheless, there is a lack of model selection approach to the regression function as a functional function with functional covariate inputs. To study interaction effects among these functional covariates, in this article, we first construct a tensor product space of reproducing kernel Hilbert spaces and build an analysis of variance (ANOVA) decomposition of the tensor product space. We then use a model selection method with the L_1 criterion to estimate the functional function with functional covariate inputs and detect interaction effects among the functional covariates. The proposed method is evaluated using simulations and stroke rehabilitation data.

Keywords: model selection; L1 criterion; reproducing kernel Hilbert space; smoothing spline

1. Introduction

Functional data can be found in various fields, such as biology, economics, engineering, geology, medicine and psychology. Recently, statistical methods and theories on functional data were widely studied ([1-6]). Functional data sometimes have more complicated structures. For example, the motivation data of this paper, stroke rehabilitation data, utilized a collection of 3D video games known as Circus Challenge to enhance upper limb function in stroke patients ([7–9]). The patients were scheduled to play the movement game over three months at specified times. At each visit time *t*, the level of impairment of stroke subject *i* was measured using the CAHAI (Chedoke Arm and Hand Activity Inventory) score, denoted as $y_i(t)$, and movements of upper limbs of patients, such as forward circle movement and sawing movement, were also recorded. The movement data at time t and frequency s from the *i*th patient were denoted as $x_i(t, s)$. Determining a way to model the relationship of $y_i(t)$ and functional $x_i(t, \cdot)$ is key to studying whether the movements are helpful to the rehabilitation level of the stroke patient or not. Furthermore, there is a question of whether there are interaction effects among the movements on the stroke patient's rehabilitation. Zhai et al. [9] developed a nonparametric concurrent regression model to study the relationship between functional movements and the CAHAI score. However, they did not consider the interaction effects of functional movements on the CAHAI score. We aim to examine the interaction effects of movements on CAHAI scores and predict the rehabilitation level of stroke patients in this paper.

In this paper, we apply the following nonparametric concurrent regression model (NCRM) to model stroke rehabilitation data:

$$y_i(t) = f(t, \mathbf{x}_i(t, \cdot)) + \epsilon_i(t), \quad i = 1, \dots, n,$$
(1)

where *f* is a bivariate functional to be estimated nonparametrically, response $y_i(t)$ is a function of *t*, covariate $x_i(t, \cdot)$ is a vector of functional with length *q*, and $\epsilon_i(t)$ is a random



Citation: Pan, R.; Wang, Z.; Wu, Y. Detection of Interaction Effects in a Nonparametric Concurrent Regression Model. *Entropy* **2023**, 25, 1327. https://doi.org/10.3390/ e25091327

Academic Editors: Udo Von Toussaint and Yuehua Wu

Received: 1 June 2023 Revised: 5 August 2023 Accepted: 8 September 2023 Published: 12 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). error. To explore the interaction effects among components of covariates $x_i(t, \cdot)$, we use the smooth spline analysis of variance (SS ANOVA) method [10,11] to decompose regression function f.

A multivariate function can be decomposed of main effects and interaction effects via the SS ANOVA method ([10,11]). When the dimension of covariates q is large, the decomposed model contains a large number of interaction effects. Even if only the main effects and second-order interaction terms are investigated, the order of the number of decomposition terms is $O(q^2)$, which leads to a highly complicated model. To model stroke rehabilitation data with q = 3, there are 22 terms, including the main effects and interaction effects. This challenges the estimation method for the NCRM model. To avoid this shortcoming, Zhai et al. [9] took all functional covariates as a whole and did not consider interaction effects among covariates. Following [12], this paper conducts a model selection method for the NCRM model with all main effects and interaction effects. In this method, the regression function is estimated and significant components of the decomposition are selected simultaneously.

Model selection is a crucial step in building statistical models that accurately capture the relationships between variables ([13,14]). It can choose the most suitable model from a set of candidate models based on certain criteria such as goodness-of-fit, predictive performance, and interpretability. Based on the SS ANOVA approach, model selection is crucial to determine the contribution of each component of the decomposition to the overall variance of the response variable. Several methods have been proposed for selecting models with SS ANOVA, including forward selection, backward elimination, and stepwise regression ([15–20]). However, these methods are limited in their ability to handle highdimensional data and identify complex interactions among variables. Hence, regularization methods such as the L_1 penalty have gained popularity in recent years ([12,21–25]), which allow for the selection of sparse and robust models. For example, Zhang et al. [23] developed a nonparametric penalized likelihood method with the likelihood basis pursuit and used it for variable selection and model construction. Lin and Zhang [22] proposed a component selection and smoothing method for multivariate nonparametric regression models by penalizing the sum of component norms of SS ANOVA. Furthermore, Wang et al. [12] developed a unified framework for estimation and model selection methods in nonparametric function-on-function regression, which performs well when using L_1 penalty methods for model selection. Dong and Wang [24] proposed a nonparametric method for learning conditional dependence structure in graph models by applying L_1 regularization to detect the neighborhoods of edges, where SS ANOVA decomposition is used to depict interaction effects of edges in the graph model. In this paper, we borrowed the L_1 regularization idea to build model selection by penalizing the sum of norms of the ANOVA decomposition components for the NCRM model. In addition, Bayesian analysis methods can also be used to study interaction effects; for example, Ren et al. [26] proposed a novel semiparametric Bayesian variable selection model for investigating linear and nonlinear gene-environment interactions simultaneously, allowing for structural identification.

This paper proposes an estimation and model selection approach for the NCRM model (1). Following [12,22], the SS ANOVA decomposition for the tensor product space of the reproducing kernel Hilbert spaces (RKHS) is constructed, and the L_1 penalty approach for the components of the decomposition is implemented. We use estimation procedures under either an L_1 or a joint L_1 and L_2 penalty to fit teh NCRM model. We study the interaction effects of the covariate $x_i(t, \cdot)$ in model (1) via ANOVA decomposition of the regression function, where the tensor product RKHS is built based on Gaussian kernels. The decomposition is different from that of Zhai et al. [9], where they took the covariate as a whole variable and did not consider their interaction effects. Based on the decomposition, model selection with the tensor product RKHS is conducted using the L_1 penalty method. With regards to the covariate $x_i(t, \cdot)$, the models from Wang et al. [12] are not suitable to analyze the stoke data. In this paper, we apply the proposed method to stroke rehabilitation

data and study the relationship of the movements and the patient's CAHAI score. Besides the main effects, the interaction effect of the movements is also detected.

The remainder of the article is organized as follows. In Section 2, we present the tensor product RKHS with the Gaussian kernel and the SS ANOVA decomposition of the regression function. In Section 3, we show model selection and estimation procedures. The simulation study and application of stroke rehabilitation data are presented in Sections 4 and 5. We conclude in Section 6.

2. Nonparametric Concurrent Regression Model

For the NCRM model (1), we consider $x_i(t, \cdot) = (x_{i1}(t, \cdot), \dots, x_{iq}(t, \cdot))$, where $x_{ij}(t,s) : S \to \mathcal{R}$ for any fixed time $t \in \mathcal{T}$ is a function of s within a space denoted by \mathcal{X}_j , $j \in 1, \dots, q$. Generally, t and s can be transformed into [0, 1]. For simplicity, we let $\mathcal{T} = [0, 1]$ and $\mathcal{S} = [0, 1]$ and let $\mathcal{X}_j \subset L^2[0, 1]$, $j = 1, \dots, q$, which are independent of t. Furthermore, we assume that $y_i(t) \in \mathcal{Y} \subset L^2[0, 1]$ and $\epsilon_i(t)$ for $i = 1, \dots, n$ are identically and independently distributed in $L^2[0, 1]$ with mean zero and $\int_0^1 E[\epsilon_i(t)^2]dt < \infty$. It is shown that the regression function f is a functional function with an independent covariate $x_i(t, \cdot)$. To provide a nonparametric estimation of f, the SS ANOVA decomposition method is used to construct a tensor product space of RKHS to which f belongs.

When *f* is treated as a function with respect to the first augment $t \in T$, we consider the Sobolev space [10],

$$\mathcal{H}^{(1)} = \left\{ f: f \text{ and } f' \text{ absolutely continuous, } \int_0^1 (f'')^2 dt < \infty \right\},$$
(2)

where $\mathcal{H}^{(1)}$ can be rewritten as

$$\mathcal{H}^{(1)} = \{1\} \oplus \{t\} \oplus \mathcal{H}^{(1)}_2,$$

where {1} is a constant space, {*t*} is a linear function space with *t* as an independent variable. $\mathcal{H}_2^{(1)}$ is a smooth function space orthogonal to the constant space and the linear function space. Reproducing kernels (RK) for these three subspaces are $\mathcal{K}_0^{(1)}(t,t') = 1$, $\mathcal{K}_1^{(1)}(t,t') = k_1(t)k_1(t')$, and $\mathcal{K}_2^{(1)}(t,t') = k_2(t)k_2(t') - k_4(|t-t'|)$, where k_1 , k_2 and k_4 are defined as

$$k_1(x) = x - 0.5,$$

$$k_2(x) = \frac{1}{2} \left\{ k_1^2(x) - \frac{1}{12} \right\},$$

$$k_4(x) = \frac{1}{24} \left\{ k_1^4(x) - \frac{1}{2} k_1^2(x) + \frac{7}{240} \right\}$$

For functional augments $x(t, \cdot)$, RK and its corresponding RKHS for f as a function of functions in $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_q$ are constructed as follows. For any $u_j, u'_j \in \mathcal{X}_j$, we construct a Gaussian kernel as

$$\mathcal{K}_{2,j}^{(2)}(u_j, u_j') = \exp\left\{-\frac{\left\|u_j - u_j'\right\|^2}{2}\right\},\tag{3}$$

where $||u_j||^2 = \int_0^1 u_j^2(s) ds$. We can show that when the space \mathcal{X}_j is a complete space, $\mathcal{K}_{2,j}^{(2)}$ is a symmetric and strictly positive definite. The unique RKHS $\mathcal{H}_{2,j}^{(2)}$ derived from $\mathcal{K}_{2,j}^{(2)}$ is separable and does not contain any non-zero constants. To construct an SS ANOVA

decomposition, we let $\mathcal{H}_{j}^{(2)} = \{1\} \oplus \mathcal{H}_{2,j}^{(2)}$. Then, the tensor product space in this paper is $\mathcal{H}^{(2)} = \mathcal{H}_{1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{q}^{(2)}$ with the following decomposition:

$$\mathcal{H}^{(2)} = \mathcal{H}_{1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{q}^{(2)}$$

$$= \{1\} \oplus \mathcal{H}_{2,1}^{(2)} \oplus \cdots \oplus \mathcal{H}_{2,q}^{(2)} \oplus \left\{\mathcal{H}_{2,1}^{(2)} \otimes \mathcal{H}_{2,2}^{(2)}\right\} \oplus \cdots \oplus$$

$$\left\{\mathcal{H}_{2,q-1}^{(2)} \otimes \mathcal{H}_{2,1}^{(2)}\right\} \oplus \cdots \oplus \left\{\mathcal{H}_{2,1-1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{2,q}^{(2)}\right\}.$$

$$(4)$$

Decomposition (4) is different from that of Zhai et al. [9] where $\mathcal{H}^{(2)}$ is decomposed of constant space {1} and another RKHS not considering interaction among $x(t, \cdot)$.

Next, we consider the tensor product space $\mathcal{H} = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ which has the following decomposition:

$$\begin{aligned} \mathcal{H} = & \left(\{1\} \oplus \{t\} \oplus \mathcal{H}_{2}^{(1)}\right) \otimes \left(\{1\} \oplus \{\oplus_{j=1}^{q} \mathcal{H}_{2,j}^{(2)}\} \oplus \cdots \oplus \left\{\mathcal{H}_{2,1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{2,q}^{(2)}\right\}\right) \\ = & \left\{1\} \oplus \{t\} \oplus \mathcal{H}_{2}^{(1)} \oplus \{\oplus_{j=1}^{q} \mathcal{H}_{2,j}^{(2)}\} \oplus \left\{\oplus_{j=1}^{q} \{\mathcal{H}_{2,j}^{(2)} \times \{t\}\}\right\} \oplus \\ & \left\{\oplus_{j=1}^{q} \{\mathcal{H}_{2,j}^{(2)} \times \mathcal{H}_{2}^{(1)}\}\} \oplus \cdots \oplus \left\{\mathcal{H}_{2,1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{2,q}^{(2)}\right\} \oplus \\ & \left\{\mathcal{H}_{2,1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{2,q}^{(2)} \times \{t\}\right\} \oplus \left\{\mathcal{H}_{2,1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{2,q}^{(2)} \otimes \mathcal{H}_{2}^{(1)}\right\}.\end{aligned}$$

There, the null space $\{1\} \oplus \{t\}$ stands for the main effect of the parametric form of t, $\mathcal{H}_{2}^{(1)}$ is the main effect of the non-parametric form of t, $\mathcal{H}_{2,j}^{(2)}$ is the main effect of the non-parametric form of u_j , $\{t\} \otimes \mathcal{H}_{2,j}^{(2)}$ is the linear nonparametric interaction between t and u_j , $\mathcal{H}_{2}^{(1)} \otimes \mathcal{H}_{2,j}^{(2)}$ is the nonparametric nonparametric interaction between t and u_j , and so on, $\mathcal{H}_{2}^{(1)} \otimes \mathcal{H}_{2,1}^{(2)} \otimes \cdots \otimes \mathcal{H}_{2,q}^{(2)}$ is the nonparametric nonparametric interaction between t and u, where $u = (u_1, \ldots, u_q)$. We denote $\phi_1(t, u) = 1$ and $\phi_2(t, u) = k_1(t)$ as the basis functions of \mathcal{H}_0 . For example, with q = 3, the RKs corresponding to the above sub-RKHS are

$$\begin{split} \mathcal{H}_{0} &\coloneqq \{1\} \oplus \{t\} \iff \mathcal{K}_{0}((t,u),(t',u')) = 1 + k_{1}(t)k_{1}(t'), \\ \mathcal{H}_{1} &\coloneqq \mathcal{H}_{2}^{(1)} \iff \mathcal{K}_{1}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t'), \\ \mathcal{H}_{1+j} &\coloneqq \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{1+j}((t,u),(t',u')) = \mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{4+j} &\coloneqq \{t\} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{4+j}((t,u),(t',u')) = k_{1}(t)k_{1}(t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{7+j} &\coloneqq \mathcal{H}_{2}^{(1)} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{7+j}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{7+j} &\coloneqq \mathcal{H}_{2,j}^{(2)} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{7+j}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{7+j} &\coloneqq \mathcal{H}_{2,j}^{(2)} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{7+j}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{7+j} &\coloneqq \mathcal{H}_{2,j}^{(2)} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{7+j}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{7+j} &\coloneqq \mathcal{H}_{2,j}^{(2)} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{7+j}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{1+j+l} &\coloneqq \mathcal{H}_{2,j}^{(2)} \otimes \mathcal{H}_{2,l}^{(2)} \iff \mathcal{K}_{1+j+l}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}')\mathcal{K}_{2,l}^{(2)}(u_{l},u_{l}'), \\ \mathcal{H}_{1+j+l} &\coloneqq \mathcal{H}_{2}^{(1)} \otimes \mathcal{H}_{2,j}^{(2)} \otimes \mathcal{H}_{2,j}^{(2)} \iff \mathcal{K}_{20}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}')\mathcal{K}_{2,l}^{(2)}(u_{l},u_{l}'), \\ \mathcal{H}_{20} &\coloneqq \mathcal{H}_{2,1}^{(2)} \otimes \mathcal{H}_{2,3}^{(2)} \otimes \mathcal{H}_{2,3}^{(2)} \iff \mathcal{K}_{20}((t,u),(t',u')) = k_{1}(t)k_{1}(t')\prod_{j=1}^{3}\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{21} &\coloneqq \{t\} \otimes \mathcal{H}_{2,1}^{(2)} \otimes \mathcal{H}_{2,3}^{(2)} \otimes \mathcal{H}_{2,3}^{(2)} \iff \mathcal{K}_{22}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\prod_{j=1}^{3}\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{22} &\coloneqq \mathcal{H}_{2}^{(1)} \otimes \mathcal{H}_{2,1}^{(2)} \otimes \mathcal{H}_{2,3}^{(2)} \iff \mathcal{K}_{2,3}^{(2)} \iff \mathcal{K}_{22}((t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\prod_{j=1}^{3}\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{H}_{22} &\coloneqq \mathcal{H}_{2}^{(1)} \otimes \mathcal{H}_{2,1}^{(2)} \otimes \mathcal{H}_{2,3}^{(2)} \iff \mathcal{K}_{2,2}^{(2)}(t,u),(t',u')) = \mathcal{K}_{2}^{(1)}(t,t')\prod_{j=1}^{3}\mathcal{K}_{2,j}^{(2)}(u$$

for j, l = 1, 2, 3 and j < l, where the left and right parts stand for the tensor product spaces and their corresponding RKs, respectively.

3. Model Selection and Estimation

We let the projection of f onto \mathcal{H}_0 be $\sum_{k=1}^2 d_k \phi_k(t, u)$, $u = (u_1, \ldots, u_q)$, and $\{\mathcal{H}_1, \cdots, \mathcal{H}_Q\}$ be the sub-RKHS generated by the tensor product method in Section 2, where Q is a number of sub-RKHS. L_1 penalties are applied to coefficients d_k for the space \mathcal{H}_0 and components of the decomposition of f (projections of f onto \mathcal{H}_j , $j = 1, \cdots, Q$). We estimate f by minimizing the following penalized least squares:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\int_{0}^{1}(y_{i}(t)-f(t,\boldsymbol{x}_{i}(t,\cdot)))^{2}dt\right\}+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\lambda_{2}\sum_{v=1}^{Q}w_{2,v}\|P_{v}f\|_{\mathcal{H}},$$
(5)

where $f \in \mathcal{H}$, P_v is the projection operator onto \mathcal{H}_j , $\|\cdot\|_{\mathcal{H}}$ is a norm induced from \mathcal{H} , λ_1 and λ_2 are tuning parameters, and $0 \le w_{1k}, w_{2,v} < \infty$ are pre-specified weights. We may set $w_{11} = 0$ when $\phi_1 = 1$ to avoid penalty to the constant function.

Since the response function is a stochastic process in the $L^2[0, 1]$ space, there exists a set of orthogonal basis functions { $\eta_k(t), k = 1, 2, ...$ } in $L^2[0, 1]$, where { $\eta_k(t), k = 1, 2, ..., n$ } is an empirical functional principal component (EFPC) of { $y_1(t), ..., y_n(t)$ } ([27]). We let $v_{ik} = \langle y_i(t), \eta_k(t) \rangle$ and $L_{ik}f = \int_0^1 f(t, x_i(t, \cdot))\eta_k(t)dt$ for i = 1, 2, ..., n and k = 1, ..., n. We assume that { L_{ik} } are bounded linear functionals. With EFPC, functional data can be transformed to scalar data such that modeling and analysis can be conducted by using traditional statistical methods. It can show that the PLS (5) based on functional data $y_i(t)$ reduces to the following PLS based on scalar data { v_{ik} }:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{\left(\nu_{ik}-L_{ik}f\right)^{2}\right\}+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\lambda_{2}\sum_{v=1}^{Q}w_{2,v}\|P_{v}f\|_{\mathcal{H}}.$$
(6)

By Lemma 3.1 in Wang et al. [12], minimizing the PLS (6) is equivalent to minimizing the following PLS:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{\left(\nu_{ik}-L_{ik}f\right)^{2}\right\}+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v}^{-1}\|P_{v}f\|_{\mathcal{H}}^{2}+\tau_{1}\sum_{v=1}^{Q}w_{2,v}\theta_{v},\quad(7)$$

subject to $\theta_v \ge 0$ for $1 \le v \le Q$, where λ_1 , τ_0 , τ_1 are tuning parameters.

We let $\mathcal{H}^* = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_Q$. To provide an RK with linear combination of its subspaces RK for the space of \mathcal{H}^* , we define a new inner product in \mathcal{H}^* ,

$$< f,g>_{*} = \sum_{v=1}^{Q} w_{2,v} \theta_{v}^{-1} < P_{v}f, P_{v}g>,$$
(8)

where $\langle \cdot, \cdot \rangle$ is the inner product in \mathcal{H} . Under the new inner product, the RK of \mathcal{H}_1^* is

$$\mathcal{K}^{*}((t, u), (t', u')) = \sum_{v=1}^{Q} w_{2,v}^{-1} \theta_{v} \mathcal{K}_{v}((t, u), (t', u'))$$

where coefficient θ_v can measure the contribution of the components of the decomposition to the model. Next, we use the reproducing property of the kernel function to transform the infinite-dimensional optimization problem (7) into a finite-dimensional solution problem. We let $\mathcal{H}_{1n} = \text{span}\{\int_0^1 \mathcal{K}^*((t, \mathbf{x}(t, \cdot)), (t', \mathbf{x}_i(t', \cdot)))\eta_k(t')dt', i = 1, 2, ..., n, k = 1, 2, ..., n\}$, which is a subspace of \mathcal{H}^* . Then, any $f \in \mathcal{H}^*$ can be decomposed as

$$f = f_0 + f_{1n} + \rho,$$

where $f_0 \in \mathcal{H}_0$, $f_{1n} \in \mathcal{H}_{1n}$, and $\rho \in \mathcal{H}^* \ominus \mathcal{H}_{1n}$. We denote

$$\mathcal{K}^*_{(t',\boldsymbol{X}_i(t',\cdot))}(t,\boldsymbol{x}(t,\cdot)) = \mathcal{K}^*((t',\boldsymbol{x}_i(t',\cdot)),(t,\boldsymbol{x}(t,\cdot)))$$

as the evaluation function of point $(t', x_i(t', \cdot))$, and $f_1 = f_{1n} + \rho$. Then, we can rewrite the PLS (7) as

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{v_{ik}-u_{ik}-\langle f_{1}(t',\mathbf{x}_{i}(t',\cdot)),\eta_{k}(t')\rangle\right\}^{2}+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v}^{-1}\|P_{v}f\|_{\mathcal{H}^{*}}^{2}+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\tau_{1}\sum_{v=1}^{Q}w_{2,v}\theta_{v}\\
=\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{v_{ik}-u_{ik}-\langle f_{1},\mathcal{K}_{(t',\mathbf{x}_{i}(t',\cdot))}^{*}\rangle\right\}^{2}+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v}^{-1}\|P_{v}f\|_{\mathcal{H}^{*}}^{2}+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\tau_{1}\sum_{v=1}^{Q}w_{2,v}\theta_{v}\\
=\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{v_{ik}-u_{ik}-\langle f_{1},\int_{0}^{1}\mathcal{K}_{(t',\mathbf{x}_{i}(t',\cdot))}^{*}\eta_{k}(t')dt'\rangle_{\mathcal{H}^{*}}\right\}^{2}+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v}^{-1}\|P_{v}f\|_{\mathcal{H}^{*}}^{2}+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\tau_{1}\sum_{v=1}^{Q}w_{2,v}\theta_{v}\\
=\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\left\{v_{ik}-u_{ik}-\langle f_{1n},\int_{0}^{1}\mathcal{K}_{(t',\mathbf{x}_{i}(t',\cdot))}^{*}\eta_{k}(t')dt'\rangle_{\mathcal{H}^{*}}\right\}^{2}+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v}^{-1}\|P_{v}f\|_{\mathcal{H}^{*}}^{2}+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v}^{-1}\|P_{v}f_{n}\|_{\mathcal{H}^{*}}^{2}+\tau_{0}\sum_{v=1}^{Q}w_{2,v}\theta_{v},$$
(9)

where $u_{ik} = \int_0^1 f_0(t', \mathbf{x}_i(t', \cdot))\eta_k(t')dt'$. The first equality uses the reproducing property, and the third equality uses the fact that ρ is orthogonal to \mathcal{H}_{1n} . Minimizing (9) must have $\rho = 0$, and we obtain the following representer theorem:

Theorem 1 (Representer Theorem). *The solution to PLS* (9) *is*

$$f(t, \mathbf{x}(t, \cdot)) = \sum_{j=1}^{2} d_{j} \varphi_{j}(t) + \sum_{v=1}^{Q} w_{2,v}^{-1} \theta_{v} \sum_{i=1}^{n} \sum_{k=1}^{n} c_{ik} \xi_{ik}(t, \mathbf{x}(t, \cdot)),$$
(10)

where $\varphi_1(t) = 1$, $\varphi_2(t) = k_1(t)$, and $\xi_{ik}(t, \mathbf{x}(t, \cdot)) = \int_0^1 \mathcal{K}_v((t, \mathbf{x}(t, \cdot)), (t', \mathbf{x}_i(t', \cdot)))\eta_k(t')dt'$.

From this representer theorem, the PLS (9) reduces to

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(\nu_{ik}-\sum_{j=1}^{2}a_{ikj}d_{j}-\sum_{v=1}^{Q}w_{2,v}^{-1}\theta_{v}\sum_{j=1}^{n}\sum_{l=1}^{n}c_{jl}b_{ikjl})^{2}$$
$$+\lambda_{1}\sum_{k=1}^{2}w_{1k}|d_{k}|+\tau_{0}\sum_{v=1}^{Q}w_{2,v}^{-1}\theta_{v}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}c_{ik}b_{ikjl}c_{jl}+\tau_{1}\sum_{v=1}^{Q}w_{2,v}\theta_{v},$$
(11)

where $a_{ikj} = \int_0^1 \varphi_j(t)\eta_k(t)dt$, $b_{ikjl} = \sum_{v=1}^Q w_{2,v}^{-1}\theta_v b_{ikjl}^v$, $b_{ikjl}^v = \int_0^1 \xi_{jl}(t, \mathbf{x}_i(t, \cdot))\eta_k(t)dt$. We let $\mathbf{\Sigma} = \sum_{v=1}^Q w_{2,v}^{-1}\theta_v \Sigma_v$, the (i + (k - 1)n, j + (l - 1)n)th element of Σ_v is b_{ikjl}^v . We let $\mathbf{Y}_k = (v_{1k}, \ldots, v_{nk})^\top$, $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$, $\mathbf{c} = (c_{11}, c_{21}, \ldots, c_{nn})^\top$, $\mathbf{d} = (d_1, d_2)^\top$, $w_2 = (w_{2,1}, \ldots, w_{2,Q})^\top$, $\mathbf{\Sigma}$ be an $n^2 \times n^2$ matrix with b_{ikjl} as the (i + (k - 1)n, j + (l - 1)n) element. Then, the PLS (11) reduces to

$$\frac{1}{n} \|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{d} - \boldsymbol{\Sigma}\boldsymbol{c}\|^2 + \lambda_1 \sum_{k=1}^2 w_{1k} |d_k| + \tau_0 \boldsymbol{c}^\top \boldsymbol{\Sigma}\boldsymbol{c} + \tau_1 \boldsymbol{w}_2^\top \boldsymbol{\theta},$$
(12)

subject to $\theta_v \ge 0$, $v = 1, 2, \ldots, Q$.

The backfitting algorithm in Wang et al. [12] is applied to solve the PLS (12) as follows (Algorithm 1):

Algorithm 1 Model Selection AlgorithmSet initial value $d = d^0$, $\theta = \theta^0$.repeatUpdate c by minimizing $\frac{1}{n} || Y - Td - \Sigma c ||^2 + \tau_0 c^\top \Sigma c$ Calculate $Y^* = Y - R\theta$, where R is an $n \times Q$ matrix with the v-th column being $w_{2,v}^{-1} \Sigma_v c$ Update d by minimizing $\frac{1}{n} || Y^* - Td ||^2 + \lambda_1 \sum_{k=1}^2 w_{1k} |d_k|$ select tuning parameter M by the k-fold cross-validation or BIC methodUpdate θ by minimizing $\frac{1}{n} || Y - Td - R\theta ||^2 + \tau_0 c^\top R\theta$ subject to $\theta_v \ge 0$ for $1 \le v \le Q$ and $w_2^\top \theta \le M$ until c, d and θ converge

4. Statistical Properties

In this section, we assume that \mathcal{X} and \mathcal{Y} are complete measurable spaces. We let P be a probability measure on $\mathcal{X}^q \times L^2(\mathcal{T})$ and $\mathcal{M} = \mathcal{T} \times \mathcal{X}^q$. Without the loss of generality, we let the terms $\lambda_1 \sum_{k=1}^2 w_{1k} |d_k|$ and $\lambda_2 \sum_{v=1}^Q w_{2,v} ||P_v f||_{\mathcal{H}}$ in (6) be combined into $||f||_{\mathcal{H}}$.

We define a loss function,

$$L(f; \mathbf{x}, \mathbf{y}) = \int_0^1 (y(t) - f(t, \mathbf{x}(t, \cdot)))^2 dt,$$

where $y(t) \in \mathcal{Y}$ and $x \in \mathcal{X}^q$. The corresponding L-risk function (Steinwart and Christmann [28]) is

 \mathbf{T} $[\mathbf{T}/\mathbf{C}]$

$$R_{L,P}(f) = E_P[L(f; \mathbf{x}, \mathbf{y})].$$
We let $f^* = \underset{f \in \mathcal{H}}{\arg\min} R_{L,P}(f), R^*_{L,P,\mathcal{H}} = R_{L,P}(f^*), \text{ and}$

$$f_{P,\lambda} = \underset{f \in \mathcal{H}}{\arg\min} \{R_{L,P}(f) + \lambda \|f\|_{\mathcal{H}}\}$$

Obviously, $\hat{f} = f_{D,\lambda}$. We state convergence properties in the following theorem and show its proofs in Appendix B.

Theorem 2. Assume that $f : \mathcal{M} \to \mathbb{R}$ is measurable for any $f \in \mathcal{H}$, \mathcal{M} is a complete measurable space, and $|P|_2 = \int_{\mathcal{X}^q \times L^2(\mathcal{T})} ||y(t)||_2^2 dP(\mathbf{x}, y) < \infty$. When $\lambda \to 0$ and $\lambda^6 n \to \infty$ as $n \to \infty$, we have

$$|R_{L,P}(\hat{f}) - R^*_{L,P,H}| = O_p(\lambda).$$

Theorem 2 states that as λ tends to 0 and $\lambda^6 n$ tends to infinity as *n* tends to infinity, the function estimate \hat{f} is L-risk consistent (Steinwart and Christmann [28]).

5. Simulation

In this section, numerical experiments are studied to evaluate the performance of the proposed model selection approach. Functional covariate take $\mathbf{x}_i(t, \cdot) = (x_{i1}(t, \cdot), x_{i2}(t, \cdot))$, where $x_{ij}(t, \cdot) = \cos(2\pi(x_{ij}^*(t, \cdot)))$, and $x_{ij}^*(t, \cdot)$ follows a Gaussian process with mean function $\mu(t) = t$. Kernel function for the GP takes the RBF kernel $k_g(s_1, s_2) = exp(-(s_1 - s_2)^2/2)$ for j = 1 and the rational quadratic kernel $k_l(s_1, s_2) = 1/(1 + (s_1 - s_2)^2)$ for j = 2. Three functions for $f(t, \mathbf{x}(t, \cdot))$ are presented as follows: for $t \in [0, 1]$,

$$\begin{split} M_1 &: f(t, \mathbf{x}(t, \cdot)) = 1 + 5\cos(2\pi t)^3, \\ M_2 &: f(t, \mathbf{x}(t, \cdot)) = 1 + 0.5t + 10\int_0^1 x_1^3(t, s)ds + 5\int_0^1 x_2^3(t, s)ds + 10\int_0^1 x_1^3(t, s)ds\int_0^1 x_2^3(t, s)ds \\ M_3 &: f(t, \mathbf{x}(t, \cdot)) = 1 + 5\cos(2\pi t) + 10\int_0^1 x_1(t, s)x_2(t, s)ds. \end{split}$$

We see that M_1 has the main effect of t, M_2 consists of three main effects and the nonparametric nonparametric interaction of x_1 and x_2 , M_3 consists of the main effect of t, and the nonparametric nonparametric interaction of x_1 and x_2 . Random error $\epsilon_i(t)$ follows $N(0, 0.2^2)$ and $N(0, 0.5^2)$. All simulations are repeated 200 times.

We generate *n* samples $\{y_i(t), x_i(t, \cdot) : i = 1, ..., n\}$ as training data, and $n_t = 50$ samples $\{\tilde{y}_i(t), \tilde{x}_i(t, \cdot) : i = 1, ..., n\}$ as test data. For comparison, we evaluate the performance using the following root mean square error (RMSE) on the test data:

$$\text{RMSE} = \sqrt{\frac{1}{n_t} \sum_{i=1}^{n_t} \left\| f(t, \widetilde{x}_i(t, \cdot)) - \widehat{f}(t, \widetilde{x}_i(t, \cdot)) \right\|_2^2},$$

where $\|\cdot\|_2$ is the norm of $L^2(\mathcal{T})$.

The proposed model selection method is used to train the model and predict the test data, denoted by L_1 . Not considering model selection, we use the L_2 penalty method to estimate the NCRM model denoted by L_2 , which minimizes the following objective function,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\int_{0}^{1}(y_{i}(t)-f(t,\boldsymbol{x}_{i}(t,\cdot)))^{2}dt\right\}+\lambda\sum_{v=1}^{Q}\|P_{v}f\|_{\mathcal{H}}^{2},$$
(13)

where λ is the tuning parameter. After model selection, the selected model is estimated with the L_2 penalty method, which is denoted by $L_1 + L_2$. Table 1 shows the average RMSE and standard deviation in parentheses for these three estimation methods, L_1 , L_2 and $L_1 + L_2$. We see that under models M1 and M3, $L_1 + L_2$ have the smallest RMSEs among these three estimation methods. Under model M2, L_1 has better performance than L_2 and comparable results with those of $L_1 + L_2$. In addition, for the three different methods, prediction performance improves as the σ decreases and the training sample size increases.

To evaluate the performance of model selection by the L_1 penalty method, we take three measurement indices in Wang et al. [12], specificity (SPE), sensitivity (SEN) and F_1 scores,

$$SPE = \frac{TN}{TN + FP}, \quad SEN = \frac{TP}{TP + FN}, \quad F_1 = \frac{2TP}{2TP + FN + FP},$$

where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives, respectively. The non-zero components of the decomposition of the regression function are considered as positive samples. For $\theta_v > 0$, its estimated value is larger than 0, which is considered a true positive.

Table 2 shows the sensitivities, specificities, and F1 scores. Overall, the L_1 penalty method performs well in different simulation settings. In addition, model selection becomes better with decreasing σ and increasing training sample size.

n	σ	Model	L_1	<i>L</i> ₂	$L_1 + L_2$
20	0.2	M1	0.2906 (0.1386)	0.3010 (0.0287)	0.0981 (0.0420)
		M2	0.8876 (0.1147)	0.9577 (0.0782)	0.9493 (0.0707)
		M3	1.2065 (0.8556)	0.9423 (0.1713)	0.7108 (0.6317)
	0.5	M1	0.3027 (0.1323)	0.4238 (0.0654)	0.1205 (0.0591)
		M2	0.9486 (0.1166)	1.0039 (0.0760)	0.9996 (0.0794)
		M3	1.3906 (0.9134)	1.0044 (0.1952)	0.8008 (0.5486)
40	0.2	M1	0.1655 (0.0127)	0.2390 (0.0416)	0.0792 (0.0113)
		M2	0.7722 (0.0517)	0.8088 (0.0877)	0.7928 (0.0630)
		M3	0.6664 (0.2255)	0.5773 (0.0456)	0.3968 (0.0423)
	0.5	M1	0.1913 (0.0944)	0.3257 (0.0860)	0.0913 (0.0179)
		M2	0.8605 (0.0678)	0.8918 (0.0856)	0.8819 (0.0746)
		M3	0.7824 (0.2060)	0.6869 (0.0420)	0.4952 (0.0764)
80	0.2	M1	0.1358 (0.0860)	0.1416 (0.0238)	0.0737 (0.0076)
		M2	0.6040 (0.0620)	0.6599 (0.0894)	0.6595 (0.0708)
		M3	0.3227 (0.0218)	0.3793 (0.0237)	0.2635 (0.0248)
	0.5	M1	0.1635 (0.1412)	0.2419 (0.0704)	0.0844 (0.0206)
		M2	0.7338 (0.0552)	0.7589 (0.0696)	0.7578 (0.0605)
		M3	0.4080 (0.1488)	0.5511 (0.0475)	0.3566 (0.0470)

Table 1. Average values and standard deviations of RMSEs. (Methods corresponding to bold numbers perform best).

Table 2. Average values and standard deviations of SPE, SEN, F1 scores under models M1, M2, M3.

n	σ	Model	SPE	SEN	F1
20	0.2	M1	0.9956 (0.0291)	0.9800 (0.1404)	0.9783 (0.1421)
		M2	0.9700 (0.0601)	0.7017 (0.1472)	0.7883 (0.1166)
		M3	0.9906 (0.0366)	0.9850 (0.1219)	0.9592 (0.1527)
	0.5	M1	0.9961 (0.0233)	0.9850 (0.1219)	0.9800 (0.1279)
		M2	0.9786 (0.0511)	0.6917 (0.1529)	0.7881 (0.1201)
		M3	0.9889 (0.0386)	0.9850 (0.1219)	0.9553 (0.1543)
40	0.2	M1	1.0000 (0.0000)	1.0000 (0.0000)	1.0000 (0.0000)
		M2	0.9614 (0.0667)	0.9817 (0.0762)	0.9512 (0.0872)
		M3	0.9833 (0.0428)	1.0000 (0.0000)	0.9517 (0.1212)
	0.5	M1	0.9989 (0.0111)	0.9950 (0.0707)	0.9933 (0.0744)
		M2	0.9407 (0.0720)	0.9600 (0.1086)	0.9177 (0.1060)
		M3	0.9844 (0.0417)	1.0000 (0.0000)	0.9550 (0.1178)
80	0.2	M1	0.9983 (0.0175)	1.0000 (0.0000)	0.9958 (0.0424)
		M2	0.9743 (0.0737)	1.0000 (0.0000)	0.9635 (0.0701)
		M3	1.0000 (0.0000)	1.0000 (0.0000)	1.0000 (0.0000)
	0.5	M1	0.9956 (0.0291)	0.9850 (0.1219)	0.9817 (0.1259)
		M2	0.9546 (0.0853)	1.0000 (0.0000)	0.9317 (0.1031)
		M3	0.9983 (0.0135)	1.0000 (0.0000)	0.9950 (0.0406)

6. Application

The proposed model selection approach is applied to analyze stroke rehabilitation data with 70 stroke survivors ([7]).

The data consist of 34 acute patients with an incidence of stroke less than a month ago and 36 chronic patients with an incidence of stroke more than six months ago. To improve upper limb functions for stroke patients, a convenient home-based rehabilitation system via action video games with 3D-position movement behaviors has been developed [7,8]. The patients played the movement game at scheduled times. For each visit time *t*, the impairment level of subject *i* was assessed using a measure called CAHAI (Chedoke Arm and Hand Activity Inventory) score, denoted as $y_i(t)$, and movements such as forward circular movement and sawing movement were recorded. In this paper, three movements, forward circular movement of the parental limb from the *x*-axis ($x_{i1} = LA05.lx$), sawing movement of the parental limb from the *y*-axis ($x_{i2} = LA09.ly$), and the movement of the non-parental limb from the direction of the *x*-axis ($x_{i3} = LA28.rqx$) are taken as functional covariates. For the purpose of illustrating the proposed method, we use the data from acute patients. During the three-month study period, each acute patient received up to seven assessments, which resulted in 173 observations. CAHAI scores were normalized before analysis.

In this paper, we focus on the interaction effect upon the order of two, and take the following decomposition:

$$\begin{split} \mathcal{K}_{0}((t,\boldsymbol{u}),(t',\boldsymbol{u}')) &= 1 + k_{1}(t)k_{1}(t'), \\ \mathcal{K}_{1}((t,\boldsymbol{u}),(t',\boldsymbol{u}')) &= \mathcal{K}_{2}^{(1)}(t,t'), \\ \mathcal{K}_{1+j}((t,\boldsymbol{u}),(t',\boldsymbol{u}')) &= \mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{K}_{4+j}((t,\boldsymbol{u}),(t',\boldsymbol{u}')) &= k_{1}(t)k_{1}(t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}') + \mathcal{K}_{2}^{(1)}(t,t')\mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}'), \\ \mathcal{K}_{5+j+l}((t,\boldsymbol{u}),(t',\boldsymbol{u}')) &= \mathcal{K}_{2,j}^{(2)}(u_{j},u_{j}')\mathcal{K}_{2,l}^{(2)}(u_{l},u_{l}'), \end{split}$$

for j, l = 1, 2, 3 and j < l. Readers can also choose various kinds of SS ANOVA decomposition by merging kernel functions according to their own needs. From Section 3, we have

$$\mathcal{K}^*((t, u), (t', u')) = \sum_{v=1}^{10} w_{2,v}^{-1} \theta_v \mathcal{K}_v((t, u), (t', u')),$$

where coefficient θ_v for kernel function \mathcal{K}_v provides levels of contribution of \mathcal{K}_v to the overall model.

The penalty method with L_1 regularization for model selection is applied to stroke rehabilitation data. Parameters $\{\theta_v\}$ are computed, and values larger than 0 are $\theta_2 = 4.157$, $\theta_3 = 0.819$, $\theta_4 = 0.636$, $\theta_7 = 0.592$ and $\theta_{10} = 0.741$. This shows that the main effects of $x_{i1}(t, \cdot)$, $x_{i2}(t, \cdot)$ and $x_{i3}(t, \cdot)$, the linear nonparametric interaction of t and $x_3(t, \cdot)$ and the nonparametric-nonparametric interaction of $x_2(t, \cdot)$ and $x_3(t, \cdot)$ have nonzero contributions to the CAHAI score. Thus, the three movements, forward circular movements of the parental limb, awing movements of the parental limb and of the non-parental limb, may be helpful to the recovery of stroke patients. In addition, the interaction of awing movements of the parental limb and the non-parental limb, may contribute to the level of daily life dependence or upper limb function impairment. Figure 1 plots the estimates of nonparametric regression functions for four stroke patients. We can see that the regression function in the NCRM model has the same trend as the scores of CAHAI, and on the whole, they all show an increasing trend along with fluctuating trends, which shows that movements may improve upper limb function for stroke patients.

Prediction performance of the proposed method is evaluated using a tenfold cross-validation,

$$\text{RPE} = \frac{1}{10} \sum_{i=1}^{10} \sqrt{\frac{1}{n_j}} \sum_{i \in jth \ fold} \left\| Y_i(t) - \hat{Y}_i^{(-j)}(t) \right\|_{2'}^2}$$

where $\hat{Y}_i^{(-j)}(t)$ is the predicted value of $Y_i(t)$ based on the fitted selected model with the $L_1 + L_2$ penalty and the data excluding the *j*th fold. The RPE for Stoke data is 1.0690, which is smaller than 1.1700 from the method of Zhai et al. [9].



Figure 1. CAHAI scores (red line) and their corresponding fitted values (blue line) for 4 patients.

7. Conclusions

For functional data with functional covariate inputs, this paper applies the Gaussian kernel function to construct the tensor product RKHS to model the regression function. This leads to a nonparametric concurrent regression model. The L_1 penalty method is used to detect components of the SS ANOVA decomposition of the regression function, which has nonzero contribution to model fit. The backfitting algorithm is developed to estimate the model selection. The proposed method is applied to stroke rehabilitation data, and the results show that besides the main effects, there are interaction effects of the movements on the CAHAI score. This indicates that movements may help improve the level of daily life dependence or impairment of upper limb function of a stroke patient.

Author Contributions: Conceptualization, Z.W. and Y.W.; methodology, Z.W. and R.P.; data curation, R.P.; writing draft preparation, R.P.; writing and editing, Z.W. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China grant number 11971457, 12201601 and Anhui Provincial Natural Science Foundation grant number 2208085.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SS ANOVA	Smooth spline analysis of variance
RKHS	Reproducing Kernel Hilbert Space
NCRM	Nonparametric concurrent regression model

Appendix A. Tensor Product and Reproducing Kernel Hilbert Space

In this section, we provide a brief description of tensor product space, reproducing kernel Hilbert space (RKHS) and SS ANOVA.

Tensor Product Space. A tensor product space refers to the direct product of multiple vector spaces. For two vector spaces, denoted by *V* and *W*, the tensor product space is $V \otimes W = \{(v, w) | v \in V, w \in W\}$. For details, please see Lin [29].

Reproducing Kernel Hilbert Space (RKHS). Reproducing kernel is a kernel function with the property of reproduction, and a reproducing kernel Hilbert space is a type of Hilbert space that possesses the property of a reproducing kernel. Mathematically, for a reproducing kernel *K* and its deduced RKHS \mathcal{H} , the reproduction property is $f(x) = \langle f, K(x, \cdot) \rangle$, where *f* is in \mathcal{H} and *x* is an input variable. RKHS provides an effective tool for modeling nonlinear relationships and handling high-dimensional data. In the context of regression, an RKHS is utilized as the foundation for model selection and estimation. For details, please see Wainwright [30].

Smoothing spline analysis of variance. Smoothing spline analysis of variance (SS ANOVA) is a powerful tool which combines the strengths of smoothing splines and analysis of variance to facilitate the simultaneous exploration of main effects and interactions among variables. SS ANOVA is an important and useful method to model nonlinear relationships within the regression framework [31–36]. For example, Wahba [31] presented theory and applications of smoothing spline models, with a special focus on function estimation from functional data with noise, where it includes univariate smoothing spline, multidimensional thin plate spline, splines on the sphere, additive spline, and interacting spline. Furthermore, Wahba et al. [32] extended the SS ANOVA model to the exponential family distribution and used the developed method to estimate the risk of diabetic retinopathy progression. In addition, Gao et al. [34] combined an SS ANOVA model with a log-linear model to fit multivariate Bernoulli data.

To illustrate the SS ANOVA approach, we consider a nonparametric model represented as follows:

$$y = f(x_1, \dots, x_p) + \epsilon, \tag{A1}$$

where *f* is the unknown smoothing function and ϵ is an error term. Applying SS ANOVA to model (A1), we decompose *f* as follows:

$$f(x_1, \dots, x_p) = \mu + \sum_{i=1}^p f_i(x_i) + \sum_{i < j} f_{i,j}(x_i, x_j) + \dots + f_{1,\dots,p}(x_1, \dots, x_p).$$
(A2)

For this decomposition, μ denotes the overall mean, functions $f_1(x_1), f_2(x_2), \ldots$, and $f_p(x_p)$ capture the main effects inherent in Model (A1), and functions $f_{i,j}(x_i, x_j)$ provide the interactions between variables x_i and x_j , and so forth. One way to model these functions, $f_1(x_1), f_2(x_2), \ldots, f_p(x_p)$, and $f_{1,\ldots,p}(x_1, \ldots, x_p)$, is to use the smoothing splines, such as the cubic splines.

Appendix B. Proof of Theorem 2

From the triangle inequality, we have

$$|R_{L,P}(\hat{f}) - R^*_{L,P,\mathcal{H}}| \le |R_{L,P}(\hat{f}) - R_{L,P}(f_{P,\lambda})| + |R_{L,P}(f_{P,\lambda}) - R^*_{L,P,\mathcal{H}}|.$$

Hence, we separately calculate the convergence rates of $|R_{L,P}(\hat{f}) - R_{L,P}(f_{P,\lambda})|$ and $|R_{L,P}(f_{P,\lambda}) - R^*_{L,P,\mathcal{H}}|$.

Since $|P|_2$ and $||K_H||_H$ are bounded, without loss of generality, we assume that q = 1, $|P|_2 = 1$, and $||K_H|| = 1$, where K_H is the RK of H and then $R_{L,P}(0) \le |P|_2 = 1$. From the proof of Theorem 3 in Zhai et al. [9], we have

$$\left|R_{L,P}(\hat{f}) - R_{L,P}(f_{P,\lambda})\right| \le c \left(2 + \|\hat{f}\|_{\infty} + \|f_{P,\lambda}\|_{\infty}\right) \left\|\hat{f} - f_{P,\lambda}\right\|_{\infty},$$

where *c* is an indeterminate constant depending on *L*.

We know that

$$\lambda \|f_{P,\lambda}\|_{\mathcal{H}} \leq \operatorname*{arg\,inf}_{f\in\mathcal{H}} R_{L,P}(f) + \lambda \|f\|_{\mathcal{H}} \leq R_{L,P}(0) \leq 1.$$

Hence, when $\left\|f_{P,\lambda} - \hat{f}\right\|_{\mathcal{H}} \leq 1$, we have $\|f_{P,\lambda}\|_{\infty} \leq \|K_{\mathcal{H}}\|_{\infty} \|f_{P,\lambda}\|_{\mathcal{H}} \leq c\lambda^{-1}$ and $\|\hat{f}\|_{\infty} \leq \|f_{P,\lambda}\|_{\infty} + \|f_{P,\lambda} - \hat{f}\|_{\infty} \leq c\lambda^{-1} + 1$. Therefore, we have

$$\left|R_{L,P}(\hat{f}) - R_{L,P}(f_{P,\lambda})\right| \leq c\lambda^{-1} ||f_{P,\lambda} - \hat{f}||_{\mathcal{H}}.$$

Meanwhile,

$$\begin{split} \lambda \|f_{P,\lambda}\|_{\mathcal{H}} + R_{L,P}(f_{P,\lambda}) - R_{L,P,H}^* &= \inf_{f \in \mathcal{H}} \lambda \|f\|_{\mathcal{H}} + R_{L,P}(f) - R_{L,P,H}^* \\ &\leq \lambda \|f^*\|_{\mathcal{H}} + R_{L,P}(f^*) - R_{L,P,H}^*, \end{split}$$

which shows that

$$R_{L,P}(f_{P,\lambda}) - R^*_{L,P,H} \leq \lambda(\|f^*\|_{\mathcal{H}} - \|f_{P,\lambda}\|_{\mathcal{H}}) \leq c_2\lambda,$$

where $c_2 > 0$ is a constant. Taking the Fréchet derivative of $\mathcal{R}_{L,P}(f) + \lambda ||f||_{\mathcal{H}}$ with respect to f, setting it to zero, we have

$$\lambda - E_P(y - f_{P,\lambda}(x))\Phi = 0, \tag{A3}$$

where $\Phi((t, x(t, \cdot))) = K_{\mathcal{H}}(\cdot, (t, x(t, \cdot)))$ is a canonical map. We let $h(x, y) = -2(y - f_{P,\lambda}(x))$. Following the proof of Theorem 5.9 in [28], we show that

$$\langle f_{\bar{P},\lambda} - f_{P,\lambda}, E_{\bar{P}}h\Phi - E_{P}h\Phi \rangle + \lambda \|f_{P,\lambda} - f_{\bar{P},\lambda}\|_{\mathcal{H}}^{2} \leq 0,$$

where \overline{P} is any distribution defined on $\mathcal{X}^q \times L^2(\mathcal{T})$. According to (A3), we know that

$$\lambda \| f_{P,\lambda} - f_{\bar{P},\lambda} \|_{\mathcal{H}}^2 \leq \langle f_{P,\lambda} - f_{\bar{P},\lambda}, E_{\bar{P}}h\Phi - E_{P}h\Phi \rangle$$
$$\leq \| f_{P,\lambda} - f_{\bar{P},\lambda} \|_{\mathcal{H}} \cdot \| E_{\bar{P}}h\Phi - E_{P}h\Phi \|_{\mathcal{H}},$$

which indicates that

$$\left\|f_{P,\lambda}-f_{\bar{P},\lambda}\right\|_{\mathcal{H}}\leq\frac{1}{\lambda}\left\|E_{\bar{P}}h\Phi-E_{P}h\Phi\right\|_{\mathcal{H}}.$$

Let $\bar{P} = D$, and from Lemma 9.2 of Steinwart and Christmann [28], we have

$$\begin{split} & P\Big(\Big|R_{L,P}(\hat{f}) - R^*_{L,P,\mathcal{H}}\Big| \geq \epsilon\Big) \\ & \leq P\Big(c\lambda^{-2}\|E_Ph\Phi - E_Dh\Phi\|_{\mathcal{H}} + c_2\lambda > \epsilon\Big) \\ & \leq O\Big(n^{-1}\lambda^{-6}\Big), \end{split}$$

with $\epsilon = O(\lambda)$. Thence, we obtain the order of $\left| R_{L,P}(\hat{f}) - R_{L,P,\mathcal{H}}^* \right|$ as $O_p(\lambda)$.

References

- 1. Aue, A.; Rice, G.; Sönmez, O. Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2018**, *80*, 509–529. [CrossRef]
- 2. Aristizabal, J.P.; Giraldo, R.; Mateu, J. Analysis of variance for spatially correlated functional data: Application to brain data. *Spat. Stat.* **2019**, *32*, 100381. [CrossRef]
- 3. Slaoui, Y. Recursive nonparametric regression estimation for independent functional data. Stat. Sin. 2020, 30, 417–437. [CrossRef]
- 4. Yao, F.; Yang, Y. Online estimation for functional data. J. Am. Stat. Assoc. 2021, 1–15.
- 5. Smida, Z.; Cucala, L.; Gannoun, A.; Durif, G. A median test for functional data. J. Nonparametric Stat. 2022, 34, 520–553. [CrossRef]
- 6. De Silva, J.; Abeysundara, S. Functional Data Analysis on Global COVID-19 Data. Asian J. Probab. Stat. 2023, 21, 12–28. [CrossRef]
- Serradilla, J.; Shi, J.; Cheng, Y.; Morgan, G.; Lambden, C.; Eyre, J.A. Automatic assessment of upper limb function during play of the action video game, circus challenge: Validity and sensitivity to change. In Proceedings of the 2014 IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH), IEEE, Rio de Janeiro, Brazil, 14–16 May 2014; pp. 1–7.
- Shi, J.Q.; Cheng, Y.; Serradilla, J.; Morgan, G.; Lambden, C.; Ford, G.A.; Price, C.; Rodgers, H.; Cassidy, T.; Rochester, L.; et al. Evaluating functional ability of upper limbs after stroke using video game data. In Proceedings of the Brain and Health Informatics: International Conference, BHI 2013, Maebashi, Japan, 29–31 October 2013; pp. 181–192.
- 9. Zhai, Y.; Wang, Z.; Wang, Y. A nonparametric concurrent regression model with multivariate functional inputs. *Stat. Its Interface* 2023, *To be appear*.
- 10. Wang, Y. Smoothing Splines: Methods and Applications; Chapman and Hall: New York, NY, USA, 2011.
- 11. Gu, C. Smoothing Spline ANOVA Models, 2nd ed.; Springer: New York, NY, USA, 2013.
- 12. Wang, Z.; Dong, H.; Ma, P.; Wang, Y. Estimation and model selection for nonparametric function-on-function regression. *J. Comput. Graph. Stat.* **2022**, *31*, 835–845. [CrossRef]
- 13. Vapnik, V.; Izmailov, R. Rethinking statistical learning theory: Learning using statistical invariants. *Mach. Learn.* 2019, 108, 381–423. [CrossRef]
- 14. Hsu, H.L.; Ing, C.K.; Tong, H. On model selection from a finite family of possibly misspecified time series models. *Ann. Stat.* **2019**, 47, 1061–1087. [CrossRef]
- 15. Guo, W. Inference in smoothing spline analysis of variance. J. R. Stat. Soc. Ser. B Stat. Methodol. 2002, 64, 887–898. [CrossRef]
- 16. Yuan, M.; Lin, Y. Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 2006, 68, 49–67. [CrossRef]
- 17. Olusegun, A.M.; Dikko, H.G.; Gulumbe, S.U. Identifying the limitation of stepwise selection for variable selection in regression analysis. *Am. J. Theor. Appl. Stat.* **2015**, *4*, 414–419. [CrossRef]
- Malik, N.A.M.; Jamshaid, F.; Yasir, M.; Hussain, A.A.N. Time series Model selection via stepwise regression to predict GDP Growth of Pakistan. *Indian J. Econ. Bus.* 2021, 20, 1881–1894.
- 19. Untadi, A.; Li, L.D.; Li, M.; Dodd, R. Modeling Socioeconomic Determinants of Building Fires through Backward Elimination by Robust Final Prediction Error Criterion. *Axioms* 2023, *12*, 524. [CrossRef]
- Radman, M.; Chaibakhsh, A.; Nariman-zadeh, N.; He, H. Generalized sequential forward selection method for channel selection in EEG signals for classification of left or right hand movement in BCI. In Proceedings of the 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 24–25 October 2019; pp. 137–142.
- 21. Storlie, C.B.; Bondell, H.D.; Reich, B.J.; Zhang, H.H. Surface estimation, variable selection, and the nonparametric oracle property. *Stat. Sin.* **2011**, *21*, 679–705. [CrossRef]
- Lin, Y.; Zhang, H.H. Component selection and smoothing in multivariate nonparametric regression. *Ann. Stat.* 2006, 34, 2272–2297. [CrossRef]
- 23. Zhang, H.H.; Wahba, G.; Lin, Y.; Voelker, M.; Ferris, M.; Klein, R.; Klein, B. Variable selection and model building via likelihood basis pursuit. *J. Am. Stat. Assoc.* 2004, *99*, 659–672. [CrossRef]
- 24. Dong, H.; Wang, Y. Nonparametric Neighborhood Selection in Graphical Models. J. Mach. Learn. Res. 2022, 23, 1–36.
- 25. Dong, H. Nonparametric Learning Methods for Graphical Models; University of California: Santa Barbara, CA, USA, 2022.
- Ren, J.; Zhou, F.; Li, X.; Chen, Q.; Zhang, H.; Ma, S.; Jiang, Y.; Wu, C. Semiparametric Bayesian variable selection for geneenvironment interactions. *Stat. Med.* 2020, 39, 617–638. [CrossRef]
- 27. Hsing, T.; Eubank, R. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators; John Wiley & Sons: Hoboken, NJ, USA, 2015; p. 997.
- 28. Steinwart, I.; Christmann, A. Support Vector Machines; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.
- 29. Lin, Y. Tensor product space ANOVA models. Ann. Stat. 2000, 28, 734–755. [CrossRef]
- 30. Wainwright, M.J. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint;* Cambridge University Press: Cambridge, UK, 2019; Volume 48.
- 31. Wahba, G. Spline Models for Observational Data; SIAM: Philadelphia, PA, USA, 1990.
- Wahba, G.; Wang, Y.; Gu, C.; Klein, R.; Klein, B. Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy: The 1994 Neyman Memorial Lecture. *Ann. Stat.* 1995, 23, 1865–1895. [CrossRef]
- 33. Gu, C.; Wahba, G. Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". J. Comput. Graph. Stat. 1993, 2, 97–117.

- 34. Gao, F.; Wahba, G.; Klein, R.; Klein, B. Smoothing spline ANOVA for multivariate Bernoulli observations with application to ophthalmology data. *J. Am. Stat. Assoc.* **2001**, *96*, 127–160. [CrossRef]
- 35. Guo, W.; Dai, M.; Ombao, H.C.; Von Sachs, R. Smoothing spline ANOVA for time-dependent spectral analysis. *J. Am. Stat. Assoc.* **2003**, *98*, 643–652. [CrossRef]
- Chiu, C.Y.; Liu, A.; Wang, Y. Smoothing spline mixed-effects density models for clustered data. *Stat. Sin.* 2020, 30, 397–416. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.