

Research Article

Automatic Variable Selection for Partially Linear Functional Additive Model and Its Application to the Tecator Data Set

Yuping Hu ^{1,2}, Sanying Feng ¹, and Liugen Xue ²

¹School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

²College of Applied Sciences, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Liugen Xue; lgxue@bjut.edu.cn

Received 1 November 2017; Revised 12 March 2018; Accepted 27 May 2018; Published 16 August 2018

Academic Editor: J.-C. Cortés

Copyright © 2018 Yuping Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We introduce a new partially linear functional additive model, and we consider the problem of variable selection for this model. Based on the functional principal components method and the centered spline basis function approximation, a new variable selection procedure is proposed by using the smooth-threshold estimating equation (SEE). The proposed procedure automatically eliminates inactive predictors by setting the corresponding parameters to be zero and simultaneously estimates the nonzero regression coefficients by solving the SEE. The approach avoids the convex optimization problem, and it is flexible and easy to implement. We establish the asymptotic properties of the resulting estimators under some regularity conditions. We apply the proposed procedure to analyze a real data set: the Tecator data set.

1. Introduction

Functional data may be viewed as realization of observed stochastic processes, and it is commonly encountered in many fields of applied sciences, such as econometrics, biomedical studies, and physics experiment. The Tecator data set is collected by the Tecator company and is publicly available at <http://lib.stat.cmu.edu/datasets/tecator>. This data set consists of 215 meat samples. The measurements were made through a spectrometer named the Tecator Infratec Food and Feed Analyzer, and the spectral curves were recorded at wavelengths ranging from 850 nm to 1050 nm. For each meat sample, the data consists of a 100 channel spectrum of absorbances as well as the contents of moisture (water), fat, and protein. The three contents of fat, protein, and moisture (water), measured in percentages, are determined by analytic chemistry. We aim to predict the fat content of a meat sample. In this paper, we propose a new partially linear functional additive model and apply the SEE procedure to analyze the Tecator data set.

With the development of computer technology, much progress has been made on developing methodologies for analyzing functional data by many researchers, like Ramsay and Silverman [1], Cardot, Ferraty, and Sarda [2], Lian and

Li [3], Fan, James, and Radchenko [4], Feng and Xue [5], Yu, Zhang, and Du [6], Zhou, Du, and Sun [7], among others. Regression models play a major role in the functional data analysis. The most widely used regression model is the following functional linear model:

$$Y = \int_{\mathcal{T}} \beta(t) X(t) dt + \varepsilon, \quad (1)$$

where Y is a scalar response, functional predictor $X(t)$ is a smooth and square-integrable random function defined on a compact domain $\mathcal{T} = [0, 1]$ for simplicity, $\beta(t)$ is the square-integrable regression parameter function, and ε is a random error, which is independent of $X(t)$. A commonly adopted approach for fitting model is the basis expansion; that is, $X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t)$, where $\mu(t) = E\{X(t)\}$. Model (1) is then transformed to a linear form with the coefficients ξ_k : $E(Y | X) = a_0 + \sum_{k=1}^{\infty} \xi_k b_k$, where $a_0 = \int_0^1 \beta(t) \mu(t) dt$ and $b_k = \int_0^1 \beta(t) \phi_k(t) dt$. The basis function set $\{\phi_k(t)\}$ can be either predetermined (e.g., Fourier basis, wavelets, or B-splines basis) or data-driven. One convenient choice for data-driven basis is the eigenbasis of the autocovariance operator of $X(t)$, in which case the random coefficients $\{\xi_k\}$ are called the functional principal component (FPC) scores. The FPC

scores have zero means and variances equal to the corresponding eigenvalues $\{\lambda_k, k = 1, 2, \dots\}$. We focus on the FPC representation of the functional regression throughout this paper.

Müller and Yao [8] relaxed the linearity assumption and proposed a functional additive model (FAM). This leads to a more widely applicable and flexible framework for the functional regression models. In the case of scalar response, the linear structure is replaced by the sum of nonlinear functional components; that is,

$$E(Y | X) = a_0 + \sum_{k=1}^{\infty} f_k(\xi_k), \quad (2)$$

where $f_k(\xi_k)$ are unknown smooth functions. Model (2) was fitted by estimating $\{\xi_k\}$ by the functional principal component analysis, and by estimating $\{f_k\}$ by the local polynomial smoothing in Müller and Yao [8]. Zhu, Yao, and Zhang [9] proposed a new regularization framework for the structure estimation of FAM in the context of reproducing kernel Hilbert spaces. The selection was achieved by the penalized least squares using a penalty which encourages the sparse structure of the additive components, and the rate of convergence was investigated.

However, in many real world problems, it is common to collect information on a large number of nonfunctional predictors. How to incorporate scalar predictors into the functional regression and perform model selection are important issues. In this paper, we combine the linear model with the functional additive model and introduce a new partially linear functional additive model (PLFAM).

Traditional real-value additive models were studied in Stone [10], Wang and Yang [11], Huang, Horowitz and Wei [12], and Zhao and Xue [13]. When the explanatory variables are of functional nature, Ferraty and Vieu [14] used a two-step procedure to estimate an additive model with two functional predictors. Fan, James, and Radchenko [4] suggested a new penalized least squares method to fit the nonlinear functional additive model. This method can efficiently fit the high-dimensional functional models while simultaneously performing variable selection to identify the relevant predictors. Febrero-Bande and Gonzalez-Manteiga [15] extended the ideas of the generalized additive models with multivariate data to the functional data covariates. The proposed algorithm was a modified version of the local scoring and backfitting algorithm that allows for the nonparametric estimation of the link function.

In the last decades, variable selection has received substantial attention, which has been a very important topic in regression analysis. Generally speaking, most of the variable selection procedures are based on penalized estimation based on some penalty functions, like Lasso penalty [16], SCAD penalty [17], Adaptive Lasso [18], and so on. However, these penalty functions have a singularity at zero such that these penalized estimation procedures require convex optimization, while adding the burden of computation. To overcome this problem, Ueki [19] developed a new variable selection procedure called the smooth-threshold estimating equation that can automatically eliminate irrelevant parameters by

setting them as zero. The method has also been successfully applied to a large class of models; for example, Lai, Wang and Lian [20] explored the generalized the estimation equation (GEE) estimation and the smooth-threshold generalized estimation equation (SGEE) variable selection for single-index model with clustered data. Li et al. [21] considered the SGEE variable selection for the generalized linear model with longitudinal data. Tian, Xue, and Xu [22] proposed a smooth-threshold estimating equation variable selection for varying coefficient models with longitudinal data.

As we know, functional regression models have been widely applied to engineering problems. For example, Escabias, Aguilera, and Valderrama [23] used functional logistic regression to deal with the environmental problem, which is to estimate the risk of drought in a specific zone from time evolution of temperatures. Sonja, Branimir, and DraDen [24] dealt with tool wear in milling process and the prediction of its behaviour by utilizing functional data analysis (FDA) methodology. Pokhrel and Tsokos [25] applied functional data analysis techniques to model age-specific brain cancer mortality trend and forecast entire age-specific functions using exponential smoothing state-space models.

In this article, we propose a new functional regression model and consider the variable selection problem for this model; then we apply the proposed procedure to analyze the Tecator data set. Motivated by the idea in Ueki [19], based on the functional principal components analysis and the centered spline basis function approximation, an automatic variable selection procedure is proposed using the smooth-threshold estimating equation. The proposed procedure automatically eliminates the irrelevant parameters in the model, while estimating the nonzero regression coefficients. Our approach can be easily implemented without solving any convex optimization problems, and it reduces the burden of computation. The proposed method shares some of the desired features including the oracle property. The proposed smooth-threshold estimating equation approach is flexible and easy to implement. Finally, the proposed method is applied to analyze a real data set: the Tecator data set. The validity of the partially linear functional additive model and the SEE method are confirmed.

The rest of this paper is organized as follows. In Section 2, we propose a variable selection procedure for PLFAM and study the asymptotic properties under some regularity conditions. In Section 3, we give the computation of the estimators as well as the choice of the tuning parameters. In Section 4, we apply the proposed method to analyze the Tecator data set. Concluding remarks are presented in Section 5. The technical proofs of all asymptotic results are provided in the Appendix.

2. Methodology and Main Results

Let Y be a real-valued random variable, Z be a p -dimensional vector of random variables, and $X(t)$ be a zero mean and square-integrable random function defined on interval $[0, 1]$. Assume that $\{Y_i, Z_i, X_i(\cdot)\}_{i=1}^n$ are independent identically and distributed realizations of the pair $\{Y, Z, X(\cdot)\}$. Denote by $\xi_{i,\infty} = (\xi_{i1}, \xi_{i2}, \dots)^T$ the FPC scores sequence of $X_i(t)$, which is associated with eigenvalues $\{\lambda_1, \lambda_2, \dots\}$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$.

2.1. Smooth-Threshold Estimating Equation. For the convenience of model regularization, we would like to restrict the predictor variables to take values in $[0, 1]$ without loss of generality. This is achieved by taking a transformation of the FPC scores through a cumulative distribution function (CDF) $\Psi : R \rightarrow [0, 1]$ for all $\{\xi_{ik}\}$. We take the normal CDF denoted by $\Psi(\cdot, \lambda_k)$, with zero mean and variance λ_k . It is easy to see that, if ξ_{ik} s follow normal distribution, the normal CDF leads to uniformly distributed transformed variables on $[0, 1]$.

Denoting the transformed variable of $\{\xi_{ik}\}$ by $\{\zeta_{ik}\}$, i.e., $\zeta_{ik} = \Psi(\xi_{ik}, \lambda_k)$, and denoting $\zeta_{i,\infty} = (\zeta_{i1}, \zeta_{i2}, \dots)^T$, we propose a partially linear functional additive model as follows:

$$Y_i = Z_i^T \beta + \sum_{k=1}^{\infty} f_{0k}(\zeta_{ik}) + \varepsilon_i, \quad (3)$$

where ε_i is independent error with zero mean and variance σ^2 , β is a $p \times 1$ vector of unknown regression coefficients, and $f_0(\zeta_{i,\infty}) = \sum_{k=1}^{\infty} f_{0k}(\zeta_{ik})$ is a smooth function. To ensure identifiability, we assume that $E f_{0k}(\zeta_{ik}) = 0$. In this paper, we assume that the PLFAM is the sparse structure of the underlying true model, and this assumption is critical in the context of functional data analysis. It means that the number of important functional additive components that contribute to the response is finite, but not necessarily restricted to the leading terms. In particular, we denote by \mathcal{S} the index set of the important FPC scores and assume that $|\mathcal{S}| < \infty$, where $|\cdot|$ denotes the cardinality of a set. In other words, there is a sufficiently large q such that $\mathcal{S} \subseteq \{1, \dots, q\}$, which implies that $f_k \equiv 0$ as long as $k > q$. Model (3) is thus equivalent to

$$Y_i = Z_i^T \beta + \sum_{k=1}^q f_{0k}(\zeta_{ik}) + \varepsilon_i. \quad (4)$$

We replace $f_{0k}(\zeta_{ik})$ with its basis function approximations. Let $B(\zeta_{ik}) = (B_1(\zeta_{ik}), \dots, B_L(\zeta_{ik}))^T$ be the centralized B-spline basis functions with the order of M , where $L = K + M$, and K is the number of interior knots. Thus, $f_{0k}(\zeta_{ik})$ can be approximated by

$$f_{0k}(\zeta_{ik}) \approx B(\zeta_{ik})^T \gamma_k, \quad k = 1, \dots, q. \quad (5)$$

Substituting this into model (4), we have

$$Y_i \approx Z_i^T \beta + \sum_{k=1}^q B(\zeta_{ik})^T \gamma_k + \varepsilon_i. \quad (6)$$

Let $P_i = (Z_i^T, W_i^T)$, $W_i^T = (B(\zeta_{i1}), \dots, B(\zeta_{iq}))^T$, and $\theta = (\theta_1, \dots, \theta_s)^T = (\beta^T, \gamma_1^T, \dots, \gamma_q^T)^T$, $s = p + qL$. Due to the unknown ζ_{ik} , we substitute the estimator of ζ_{ik} by $\hat{\zeta}_{ik} = \Psi(\hat{\xi}_{ik}, \hat{\lambda}_k)$, where $\hat{\xi}_{ik}$ and $\hat{\lambda}_k$ are the estimators of ξ_{ik} and λ_k , respectively. Let $\hat{P}_i = (Z_i^T, \hat{W}_i^T)$, $\hat{W}_i^T = (B(\hat{\zeta}_{i1}), \dots, B(\hat{\zeta}_{iq}))^T$, and define the following estimating function of θ

$$U(\theta) = \sum_{i=1}^n \hat{P}_i^T (Y_i - \hat{P}_i \theta). \quad (7)$$

Motivated by the idea of Ueki [19], we propose the following smooth-threshold estimating equation:

$$(I_s - \Delta)U(\theta) + \Delta\theta = 0, \quad (8)$$

where $s = p + qL$, I_s is the s -dimensional identity matrix, and Δ is the diagonal matrix, i.e., $\Delta = \text{diag}(\Delta_1, \Delta_2)$, where $\Delta_1 = \text{diag}\{\delta_{11}, \dots, \delta_{1p}\}$ is a $p \times p$ diagonal matrix, and $\Delta_2 = \text{diag}\{\delta_{21}, \dots, \delta_{21}, \dots, \delta_{2q}, \dots, \delta_{2q}\}$ is a $qL \times qL$ diagonal matrix. Note that $\delta_{1j} = 1$ reduces to $\beta_j = 0$ for $j = 1, \dots, p$, and $\delta_{2k} = 1$ reduces to $\gamma_k = 0$, that is, $f_{0k}(\cdot) = 0$, for $k = 1, \dots, q$. Therefore, (8) can yield a sparse solution. Unfortunately, we cannot directly obtain the estimator of θ by solving the smooth-threshold estimating equation (8). The reason is that (8) involves the unknown parameters δ_{1j} and δ_{2k} , which need to be chosen using some data-driven criteria.

For the choice of δ_{1j} and δ_{2k} , Ueki [19] suggested that δ_{1j} and δ_{2k} may be determined by the data, $\hat{\delta}_{1j} = \min(1, \nu / \|\hat{\beta}_j^{(0)}\|^{1+\tau})$ and $\hat{\delta}_{2k} = \min(1, \nu / \|\hat{\gamma}_k^{(0)}\|^{1+\tau})$, with initial estimators $\hat{\beta}_j^{(0)}$ and $\hat{\gamma}_k^{(0)}$, respectively. The initial estimators $\hat{\beta}_j^{(0)}$ and $\hat{\gamma}_k^{(0)}$ are the solutions of the estimating equation $U(\theta) = 0$. Note that this choice involves two tuning parameters (ν, τ) . In Section 3, we will propose a BIC-type criterion to select the tuning parameters. Replacing Δ in (8) by $\hat{\Delta}$ with diagonal elements $\hat{\delta}_{1j}$ and $\hat{\delta}_{2k}$, the smooth-threshold estimating equation (8) becomes

$$(I_s - \hat{\Delta})U(\theta) + \hat{\Delta}\theta = 0. \quad (9)$$

The solution of (9) denoted by $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p, \hat{\theta}_{p+1}, \dots, \hat{\theta}_s)^T = (\hat{\beta}^T, \hat{\gamma}_1^T, \dots, \hat{\gamma}_q^T)^T$ is called the SEE estimator. Thus, $\hat{\beta}$ is the estimator of β , and $B_k(\hat{\zeta}_k)^T \hat{\gamma}_k$ is the estimator of $f_{0k}(\zeta_k)$. For the convenience of the notations, we define $\mathcal{A}_1 = \{j : \hat{\delta}_{1j} \neq 1\}$ as the set of indices of nonzero regression coefficient estimators, and define $\mathcal{A}_2 = \{k : \hat{\delta}_{2k} \neq 1\}$ as the set of nonzero function estimators. $\mathcal{A}_1^c = \{j : \hat{\delta}_{1j} = 1\}$ and $\mathcal{A}_2^c = \{k : \hat{\delta}_{2k} = 1\}$ are the sets of zero regression coefficient estimators and zero function estimators, respectively. We further denote $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$, and $\mathcal{A}^c = \mathcal{A}_1^c \cup \mathcal{A}_2^c$. The SEE estimator is defined in the following form:

$$\hat{\theta}_{\mathcal{A}} = \left\{ \sum_{i=1}^n \hat{P}_{i,\mathcal{A}}^T \hat{P}_{i,\mathcal{A}} - \hat{G}_{\mathcal{A}} \right\}^{-1} \sum_{i=1}^n \hat{P}_{i,\mathcal{A}}^T Y_i, \quad (10)$$

$$\hat{\theta}_{\mathcal{A}^c} = 0,$$

Where $\hat{P}_{i,\mathcal{A}} = (Z_{i,\mathcal{A}}^T, \hat{W}_{i,\mathcal{A}}^T)$, $\hat{G}_{\mathcal{A}} = (I_{|\mathcal{A}|} - \hat{\Delta}_{\mathcal{A}})^{-1} \hat{\Delta}_{\mathcal{A}}$, $\hat{\Delta}_{\mathcal{A}} = \text{diag}\{\hat{\Delta}_{1,\mathcal{A}_1}, \hat{\Delta}_{2,\mathcal{A}_2}\}$, and $I_{|\mathcal{A}|}$ is the identity matrix with the same dimension of $\hat{\Delta}_{\mathcal{A}}$.

2.2. Asymptotic Properties. We first introduce some notations. Let $f_0(\cdot)$ and β_0 denote the true values of $f(\cdot)$ and β , respectively. γ_0 is the spline coefficient vector from the spline approximation to $f_0(\cdot)$. Denote $\theta_0 = (\beta_0^T, \gamma_0^T)^T$, $\mathcal{A}_{10} = \{j : \beta_{0j} \neq 0\}$, $\mathcal{A}_{10}^c = \{j : \beta_{0j} = 0\}$, $\mathcal{A}_{20} = \{k : f_{0k}(\cdot) \neq 0\}$, and

$\mathcal{A}_{20}^c = \{k : f_{0k}(\cdot) = 0\}$. We also denote $\mathcal{A}_0 = \mathcal{A}_{10} \cup \mathcal{A}_{20}$ and $\mathcal{A}_0^c = \mathcal{A}_{10}^c \cup \mathcal{A}_{20}^c$. The norm of $f(\cdot)$ is defined as $\|f(x)\| = (\int_0^1 f^2(x)dx)^{1/2}$.

For convenience and simplicity, let C denote a positive constant that may be different at each appearance throughout this paper. We list some regularity conditions that are used in this paper.

(C1) The transformation function $\Psi(\xi, \lambda_k)$ is differentiable at ξ and λ_k , and satisfies the fact that $|\partial\Psi(\xi, \lambda_k)/\partial\xi| \leq C\lambda_k^c$ and $|\partial\Psi(\xi, \lambda_k)/\partial\lambda_k| \leq C\lambda_k^c|\xi|$ for some positive constants C and negative constants c_0 .

(C2) The second derivative $X^{(2)}(t)$ is continuous on $[0, 1]$ with probability 1 and $\int E[X^{(k)}(t)]^4 dt < \infty$ with probability 1, for $k = 0, 2$.

(C3) The spline regression parameter is identified; that is, there is a unique $\theta_0 = (\beta_0^T, \gamma_{01}^T, \dots, \gamma_{0q}^T)^T \in S$, where the parameter space S is compact.

(C4) The inner knots $\{c_i, i = 1, \dots, K\}$ satisfy

$$\begin{aligned} \max_{1 \leq i \leq K} |h_{i+1} - h_i| &= o(K^{-1}), \\ \frac{h}{\min_{1 \leq i \leq K} \{h_i\}} &\leq C_0, \end{aligned} \quad (11)$$

where $h_i = c_i - c_{i-1}$, $h = \max_{1 \leq i \leq K} h_i$, $c_0 = 0$, and $c_{K+1} = 1$.

(C5) $f_{0k}(\cdot)$ is the r th continuously differentiable on $(0, 1)$, where $r \geq 2$.

The following theorem gives the consistency of our proposed estimators.

Theorem 1. *Suppose that regularity conditions (C1)-(C5) hold. For any positive ν and τ , $n^{r/(2r+1)}\nu \rightarrow 0$ and $n^{r(1+\tau)/(2r+1)}\nu \rightarrow \infty$, as $n \rightarrow \infty$, and if $K = O(n^{1/(2r+1)})$, one has*

$$\begin{aligned} \|\widehat{f}_{0k}(\widehat{\zeta}_k) - f_{0k}(\zeta_k)\| &= O_p(n^{-r/(2r+1)}), \\ k &= 1, \dots, q. \end{aligned} \quad (12)$$

In the following theorem, we will show that such consistent estimators enjoy the sparsity property.

Theorem 2. *Under the regularity conditions of Theorem 1, one has*

- (i) $P(\mathcal{A}_1^c = \mathcal{A}_{10}^c) \rightarrow 1$,
- (ii) $P(\mathcal{A}_2^c = \mathcal{A}_{20}^c) \rightarrow 1$.

Next, we will show that the estimators of the nonzero coefficients for the parametric components have the same asymptotic distribution as that based on the correct model.

Theorem 3. *Suppose that the regularity conditions of Theorem 1 hold; as $n \rightarrow \infty$, one has,*

$$\{D(\widehat{\beta}_{\mathcal{A}_1})\}^{-1/2} \{\widehat{\beta}_{\mathcal{A}_1} - \beta_{0\mathcal{A}_1}\} \xrightarrow{\mathcal{L}} N(0, I_{|\mathcal{A}_1|}), \quad (13)$$

where $D(\widehat{\beta}_{\mathcal{A}_1})$ is defined in (A.23) in the Appendix, $|\mathcal{A}_1|$ is the number of \mathcal{A}_1 , $I_{|\mathcal{A}_1|}$ is the identity matrix, and " $\xrightarrow{\mathcal{L}}$ " represents the convergence in distribution.

3. Issues in Practical Implementation

3.1. Computational Algorithm. Since the transformed FPC scores $\{\zeta_i\}$ cannot be observed, we first need to estimate the FPC scores before the estimation and selection of $f_{0k}(\cdot)$. In what follows, we propose the algorithm to implement the estimation procedure.

Step 1. Apply the functional principal component analysis (Zhu, Yao, and Zhang, 2014) to estimate the FPC scores $\{\xi_{i1}, \dots, \xi_{iq}\}$ of $X_i(t)$ denoted by $\{\widehat{\xi}_{i1}, \dots, \widehat{\xi}_{iq}\}$. Then we can obtain the transformed variables $\widehat{\zeta}_{ik} = \Psi(\widehat{\xi}_{ik}, \widehat{\lambda}_k)$, where $\widehat{\lambda}_k$ is the estimated eigenvalue, and q is chosen to explain nearly 100% of the total variation.

Step 2. Calculate the initial estimate $\widehat{\theta}^0$ of θ by solving the estimating equation $U(\theta) = 0$.

Step 3. Choose the tuning parameters ν and τ by the BIC-type criterion in the next subsection.

Step 4. Solve the smooth-threshold estimating equations (9) and update the estimator of θ as follows:

$$\begin{aligned} \widehat{\theta}_{\mathcal{A}} &= \left\{ \sum_{i=1}^n \widehat{P}_{i,\mathcal{A}}^T \widehat{P}_{i,\mathcal{A}} - \widehat{G}_{\mathcal{A}} \right\}^{-1} \sum_{i=1}^n \widehat{P}_{i,\mathcal{A}}^T Y_i, \\ \widehat{\theta}_{\mathcal{A}^c} &= 0. \end{aligned} \quad (14)$$

3.2. Selection of the Tuning Parameters. Following Ueki [19], we minimize the following BIC-type criterion to choose the tuning parameters (ν, τ) .

$$\begin{aligned} BIC_{\nu,\tau} &= \sum_{i=1}^n (Y_i - \widehat{P}_i \widehat{\theta}_{\nu,\tau})^T (Y_i - \widehat{P}_i \widehat{\theta}_{\nu,\tau}) + \log(n) \\ &\quad \cdot df_{\nu,\tau}, \end{aligned} \quad (15)$$

where $\widehat{\theta}_{\nu,\tau}$ is the SEE estimator for given (ν, τ) , and $df_{\nu,\tau}$ is the number of nonzero coefficients $\widehat{\theta} = (\widehat{\beta}^T, \|\widehat{\gamma}_1\|, \dots, \|\widehat{\gamma}_q\|)^T$.

4. Application to Real Data

We demonstrate the effectiveness of the proposed method by an application for the Tecator data set. The Tecator data set contains 215 samples; each sample contains the finely chopped pure meat with different moisture, fat, and protein contents, which are measured in percentages and are determined by the analytical chemistry. The functional covariate by $X(t)$ for each food sample consists of a 100-channel spectrum of absorbances recorded on a Tecator Infracore Food and Feed Analyzer working in the wavelength range 850-1050 nm by the near-infrared transmission (NIT) principle. In this analysis, Y is the fat content, $\{\xi_k\}$ are functional principal component (FPC) scores of $X(t)$, and $\zeta_k = \Psi(\xi_k, \lambda_k)$, we take the protein and the moisture content by Z_1 and Z_2 , respectively. In order to predict the fat content of a meat sample, many models and algorithms are proposed to fit the

TABLE 1: The prediction mean squared error (PMSE) for the different models.

Models	PMSE
$Y = \mu + m_1(Z_1) + m_2(X) + \varepsilon,$	0.0242
$Y = \mu + m_3(Z_2) + m_4(X) + \varepsilon,$	0.0368
$Y = \mu + m_5(Z_1) + m_6(Z_2) + m_7(X) + \varepsilon,$	0.0395
$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \sum_{k=1}^4 f_{0k}(\zeta_k) + \varepsilon,$ (SEE Method)	0.0153
$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \sum_{k=1}^4 f_{0k}(\zeta_k) + \varepsilon,$ (SCAD Method)	0.0160
$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \sum_{k=1}^4 f_{0k}(\zeta_k) + \varepsilon,$ (LASSO Method)	0.0234

data; see, for example, Aneiros-Pérez and Vieu [26]. In this paper, to fit the data, we consider the following partially linear functional additive model:

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \sum_{k=1}^q f_{0k}(\zeta_{ik}) + \varepsilon_i. \quad (16)$$

To compare the performance of different models, the sample is divided into two data sets: the training sample $I_1 = \{1, \dots, 165\}$ is used to obtain the estimators of the parameter and the nonparametric function, and the testing sample $I_2 = \{166, \dots, 215\}$ is used to verify the quality of prediction by the following mean squared errors of prediction (PMSE),

$$\text{PMSE} = \frac{1}{50} \sum_{j \in I_2} \frac{(Y_j - \hat{Y}_j)^2}{\text{var}_{I_2}(Y)}. \quad (17)$$

In this section, we consider the variable selection problem for model (16). We apply the proposed smooth-threshold estimating equation approach to eliminate the irrelevant parameters in the model, while estimating the nonzero regression coefficients. Steps are as follows.

Step 1. Estimate the FPC scores $\{\xi_{i1}, \dots, \xi_{iq}\}$ of $X_i(t)$ denoted by $\{\hat{\xi}_{i1}, \dots, \hat{\xi}_{iq}\}$. Then, we can obtain the transformed variables $\hat{\zeta}_{ik} = \Psi(\hat{\xi}_{ik}, \hat{\lambda}_k)$, where $\hat{\lambda}_k$ is the estimated eigenvalue, and $q = 4$ is chosen to explain nearly 100% of the total variation.

Step 2. Calculate the initial estimate $\hat{\theta}^0$ of θ by solving the estimating equation $U(\theta) = 0$.

Step 3. Choose the tuning parameters ν and τ by the BIC-type criterion.

Step 4. Solve the smooth-threshold estimating equations (9).

The estimated values of the $\beta_0, \beta_1,$ and β_2 are $\hat{\beta}_0 = 95.2503, \hat{\beta}_1 = -1.0223,$ and $\hat{\beta}_2 = -0.7093,$ and $f_{02}(\cdot)$ is estimated to be 0. The SEE method selects two parametric components Z_1, Z_2 and three nonparametric components: $f_{01}(\cdot), f_{03}(\cdot),$ and $f_{04}(\cdot)$. Hence, in conclusion, we would say that the strong linear relationship between the fat content and the protein and moisture contents, and $f_{01}(\cdot), f_{03}(\cdot),$ and $f_{04}(\cdot),$ are important functional nonparametric part. Our approach can be easily implemented and reduces the burden of computation.

Aneiros-Pérez and Vieu [26] proposed several additive semifunctional models, $Y = \mu + m_1(Z_1) + m_2(X) + \varepsilon, Y = \mu +$

$m_3(Z_2) + m_4(X) + \varepsilon,$ and $Y = \mu + m_5(Z_1) + m_6(Z_2) + m_7(X) + \varepsilon,$ where $m_i(\cdot)$ are nonparametric function for $i = 1, \dots, 7,$ and $\mu = E(Y)$. To assess the performance of the proposed PLFAM and SEE method, we compare with these additive semifunctional models and the SCAD-penalized method and the Lasso-penalized method. The PMSE results are reported in Table 1. From Table 1 we can see the following.

(1) Real explanatory variables Z_1 and Z_2 can be used to improve the accuracy of the prediction; it is consistent with the conclusion of Aneiros-Pérez and Vieu [26].

(2) Compared with these additive semifunctional models, the PLFAM has a much smaller PMSE. The proposed PLFAM performs better than other models.

(3) Compared to the SCAD-penalized method and the Lasso-penalized method, the PMSE of PLFAM (SEE method) is smaller than all of the results of them.

These conclusions confirm the validity of the proposed PLFAM and SEE method.

5. Concluding Remarks

The article develops a SEE procedure for automatic variable selection in the partially linear functional additive model. The proposed procedure can identify nonzero regression coefficients significant variables from the parametric components and the nonparametric components simultaneously, and it automatically eliminates the irrelevant parameters by setting them as zero and simultaneously estimates the nonzero regression coefficients. It is noteworthy that the proposed procedure avoids the convex optimization problem, and the resulting estimator enjoys the oracle property. The application to the Tecator data set confirms the validity of the proposed PLFAM model and the SEE method.

Appendix

Proofs of the Main Results

In this Appendix, we will prove the main results stated in Section 2.

Proof of Theorem 1. Let $S_n(\theta) = (I_s - \hat{\Delta})U(\theta) + \hat{\Delta}\theta,$ and $\rho = n^{-r/2r+1}$. We will prove that, $\forall \varepsilon > 0,$ there exists a constant $C > 0,$ such that

$$P\left(\sup_{\|u\|=C} n^{-1/2} u^T S_n(\theta_0 + \rho u) > 0\right) \geq 1 - \varepsilon \quad (A.1)$$

for n large enough. This will be sufficient to ensure that there exists a local solution to the equation $S_n(\theta) = 0$ such that $\|\hat{\theta}_{\nu, \tau} - \theta_0\| = O_p(\rho)$ with probability at least $1 - \varepsilon$. That is, $\|\hat{\beta}_{\nu, \tau} - \beta_0\| = O_p(\rho)$ and $\|\hat{\gamma}_{\nu, \tau} - \gamma_0\| = O_p(\rho)$ with probability at least $1 - \varepsilon$. We will evaluate the sign of $n^{-1/2} u^T S_n(\theta_0 + \rho u)$ in the ball $\{\theta_0 + \rho u : \|u\| = C\}$. Note that

$$\begin{aligned} n^{-1/2} u^T S_n(\theta_0 + \rho u) &= n^{-1/2} u^T S_n(\theta_0) \\ &\quad + n^{-1/2} \rho u^T \frac{\partial}{\partial \theta} S_n(\bar{\theta}) u \quad (\text{A.2}) \\ &\equiv I_{n1} + I_{n2}, \end{aligned}$$

where $\bar{\theta}$ lies between θ_0 and $\theta_0 + \rho u$. By some simple calculations, we have

$$\begin{aligned} I_{n1} &= n^{-1/2} u^T (I_S - \hat{\Delta}) U(\theta_0) + n^{-1/2} u^T \hat{\Delta} \theta_0 \\ &\equiv I_{n11} + I_{n12}. \end{aligned} \quad (\text{A.3})$$

By the Cauchy-Schwarz inequality, we obtain that

$$\begin{aligned} |I_{n11}| &\leq n^{-1/2} \|u^T (I_S - \hat{\Delta})\| \|U(\theta_0)\| \leq n^{-1/2} \left(1 \right. \\ &\quad \left. - \min \left[\min_{j \in \mathcal{A}_1} \hat{\delta}_{1j}(\nu, \tau), \min_{k \in \mathcal{A}_2} \hat{\delta}_{2k}(\nu, \tau) \right] \right) \|u\| \\ &\quad \cdot \|U(\theta_0)\|. \end{aligned} \quad (\text{A.4})$$

Since $\min_{j \in \mathcal{A}_1} \hat{\delta}_{1j}(\nu, \tau) \leq \min_{j \in \mathcal{A}_{10}} \hat{\delta}_{1j}(\nu, \tau)$ and $\min_{k \in \mathcal{A}_2} \hat{\delta}_{2k}(\nu, \tau) \leq \min_{k \in \mathcal{A}_{20}} \hat{\delta}_{2k}(\nu, \tau)$, we only need to show the convergence rate of $\min_{j \in \mathcal{A}_{10}} \hat{\delta}_{1j}(\nu, \tau)$ and $\min_{k \in \mathcal{A}_{20}} \hat{\delta}_{2k}(\nu, \tau)$. Given the initial estimator $\hat{\theta}^{(0)}$, and we assume that $\hat{\theta}^{(0)}$ satisfies $\|\hat{\theta}^{(0)} - \theta_0\| = O_p(n^{-r/(2r+1)})$. By $\nu n^{r/(2r+1)} \rightarrow 0$, for any $\varepsilon > 0$ and $j \in \mathcal{A}_{10}$, we have

$$\begin{aligned} P(\hat{\delta}_{1j}(\nu, \tau) > n^{-r/(2r+1)} \varepsilon) &= P\left(\frac{\nu}{|\hat{\beta}_j^{(0)}|^{1+\tau}} \right. \\ &\quad \left. > n^{-r/(2r+1)} \varepsilon\right) = P\left(\left(\frac{\nu n^{r/(2r+1)}}{\varepsilon}\right)^{1/(1+\tau)} \right. \\ &\quad \left. > |\hat{\beta}_j^{(0)}|\right) \leq P\left(\left(\frac{\nu n^{r/(2r+1)}}{\varepsilon}\right)^{1/(1+\tau)} \right. \\ &\quad \left. > \min_{j \in \mathcal{A}_{10}} |\beta_{0j}|\right) \\ &\quad - O_p(n^{-r/(2r+1)}) \rightarrow 0. \end{aligned} \quad (\text{A.5})$$

This means that $\hat{\delta}_{1j}(\nu, \tau) = o_p(n^{-r/(2r+1)})$ for each $j \in \mathcal{A}_{10}$. Similarly, we get $\hat{\delta}_{2k}(\nu, \tau) = o_p(n^{-r/(2r+1)})$ for each $k \in \mathcal{A}_{20}$. Therefore, we have

$$\begin{aligned} \min \left(\min_{j \in \mathcal{A}_1} \hat{\delta}_{1j}(\nu, \tau), \min_{k \in \mathcal{A}_2} \hat{\delta}_{2k}(\nu, \tau) \right) \\ = o_p(n^{-r/(2r+1)}). \end{aligned} \quad (\text{A.6})$$

By this, (A.4), and (A.5), similar to the proof of Theorem 3.6 in Wang [27], we have $\|U(\theta_0)\| = O_p(n^{1/2})$, then $|I_{n11}| = O_p(1) \|u\| - o_p(n^{-r/(2r+1)}) \|u\|$.

For I_{n12} , we have $|I_{n12}| \leq n^{-1/2} \|u\| \|\theta_0\| = O_p(n^{-1/2}) \|u\|$. Hence $|I_{n1}| = O_p(1) \|u\|$. Now consider I_{n2} , we can derive that

$$\begin{aligned} I_{n2} &= n^{-1/2} \rho u^T \frac{\partial}{\partial \theta} S_n(\bar{\theta}) u \\ &= n^{-(4r+1)/(4r+2)} u^T (I_S - \hat{\Delta}) \left(\sum_{i=1}^n P_i^T P_i \right) u \\ &\quad + n^{-(4r+1)/(4r+2)} u^T \hat{\Delta} u \equiv I_{n21} + I_{n22}, \end{aligned} \quad (\text{A.7})$$

where $|I_{n22}| = O_p(n^{-(4r+1)/(4r+2)}) \|u\|^2$. For sufficiently large n , $n^{-1/2} u^T S_n(\theta_0 + \rho u)$ on the ball $\{\theta_0 + \rho u : \|u\| = C\}$ is asymptotically dominated in probability 1 by I_{n21} , which is positive for the sufficiently large C . Thus, (A.1) holds. By some simple calculations, we have

$$\begin{aligned} \|\hat{f}_{0k}(\tilde{\zeta}_k) - f_{0k}(\zeta_k)\| &= \|\hat{f}_{0k}(\tilde{\zeta}_k) - f_{0k}(\tilde{\zeta}_k) \\ &\quad + f_{0k}(\tilde{\zeta}_k) - f_{0k}(\zeta_k)\| \leq \|\hat{f}_{0k}(\tilde{\zeta}_k) - f_{0k}(\tilde{\zeta}_k)\| \\ &\quad + \|f_{0k}(\tilde{\zeta}_k) - f_{0k}(\zeta_k)\| \leq \left\{ \int_0^1 \{B^T(\tilde{\zeta}_k) \hat{\gamma}_k \right. \\ &\quad \left. - B^T(\tilde{\zeta}_k) \gamma_k - \eta_k(\tilde{\zeta}_k)\}^2 d\tilde{\zeta}_k \right\}^{1/2} \\ &\quad + \|f_{0k}(\tilde{\zeta}_k) - f_{0k}(\zeta_k)\| \leq \left\{ 2 \int_0^1 \{B^T(\tilde{\zeta}_k) \hat{\gamma}_k \right. \\ &\quad \left. - B^T(\tilde{\zeta}_k) \gamma_k\}^2 d\tilde{\zeta}_k + 2 \int_0^1 \eta_k^2(\tilde{\zeta}_k) d\tilde{\zeta}_k \right\}^{1/2} \\ &\quad + \|f_{0k}(\tilde{\zeta}_k) - f_{0k}(\zeta_k)\| = \left\{ 2(\hat{\gamma}_k - \gamma_k)^T H(\hat{\gamma}_k \right. \\ &\quad \left. - \gamma_k) + 2 \int_0^1 \eta_k^2(\tilde{\zeta}_k) d\tilde{\zeta}_k \right\}^{1/2} + \|f_{0k}(\tilde{\zeta}_k) \\ &\quad \left. - f_{0k}(\zeta_k)\|, \end{aligned} \quad (\text{A.8})$$

where $\eta_k(\tilde{\zeta}_k) = f_{0k}(\tilde{\zeta}_k) - B(\tilde{\zeta}_k)^T \gamma_k$ for $k = 1, \dots, q$, and $H = (h_{ij})$ is a matrix with $h_{ij} = \int_0^1 B(\tilde{\zeta}_k) B(\tilde{\zeta}_k)^T d\tilde{\zeta}_k$. From conditions (C3)-(C5) and Corollary 6.21 in Schumaker [28], we get that $\|\eta_k(\tilde{\zeta}_k)\| = O(K^{-r})$. Invoking the same arguments, we can get that $\|\hat{\gamma}_k - \gamma_0\| = O_p(\rho)$. By $\|H\| = O(1)$, a simple calculation yields that

$$(\hat{\gamma}_k - \gamma_k)^T H(\hat{\gamma}_k - \gamma_k) = O_p(n^{-2r/(2r+1)}). \quad (\text{A.9})$$

In addition, it is easy to show that

$$\int_0^1 \eta_k^2(\tilde{\zeta}_k) d\tilde{\zeta}_k = O_p(n^{-2r/(2r+1)}). \quad (\text{A.10})$$

By conditions (C1) and (C2), and Lemma 2 in Zhu et al. [9], we get that

$$\|f_{0k}(\tilde{\zeta}_k) - f_{0k}(\zeta_k)\| \leq C \|\tilde{\zeta}_k - \zeta_k\| = O_p(n^{-1/2}). \quad (\text{A.11})$$

Invoking (A.9)-(A.11), we complete the proof of Theorem 1. \square

Proof of Theorem 2. Similar to Theorem 1 in Zhao and Xue [29], it is known that the initial estimator $\hat{\theta}^{(0)} = (\hat{\theta}^{(0)T}, \hat{\gamma}^{(0)T})^T$ obtained by solving the estimating equation $U(\theta) = 0$ is $n^{-r/(2r+1)}$ -consistent. Noting that $n^{r(1+\tau)/(2r+1)} \nu \rightarrow \infty$, we can derive that

$$\begin{aligned} \sum_{j \in \mathcal{A}_{10}^c} P\left(\frac{\nu}{|\hat{\beta}_j^{(0)}|^{1+\tau}} < 1\right) &= \nu^{-1} O(n^{-r(1+\tau)/(2r+1)}) \\ &\rightarrow 0, \end{aligned} \quad (\text{A.12})$$

which implies that $P(\hat{\delta}_{1j} = 1 \text{ for all } j \in \mathcal{A}_{10}^c) \rightarrow 1$. Invoking the same argument, we can derive that $P(\hat{\delta}_{2k} = 1 \text{ for all } k \in \mathcal{A}_{20}^c) \rightarrow 1$. Thus, we finish the proof of Theorem 2. \square

Proof of Theorem 3. As shown in Theorem 2, $\hat{\theta}_j = 0$ for $j \in \mathcal{A}_0^c$ with probability tending to 1. In addition, $\hat{\theta}_{\mathcal{A}_0}$ is the solution of the smooth-threshold estimating equation

$$(I_{|\mathcal{A}_0|} - \hat{\Delta}_{\mathcal{A}_0})U(\hat{\theta}_{\mathcal{A}_0}) + \hat{\Delta}_{\mathcal{A}_0}\hat{\beta}_{\mathcal{A}_0} = 0. \quad (\text{A.13})$$

Applying a Taylor expansion to (A.13), we can show that $\hat{\theta}_{\mathcal{A}_0} - \theta_{\mathcal{A}_0}$ can be approximated by

$$\begin{aligned} &\left[-\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta_{\mathcal{A}_0}} U_i(\theta_{\mathcal{A}_0}) - \frac{1}{n} \hat{G}_{\mathcal{A}_0}\right]^{-1} \\ &\cdot \left[\frac{1}{n} \sum_{i=1}^n U_i(\theta_{\mathcal{A}_0}) + \frac{1}{n} \hat{G}_{\mathcal{A}_0} \theta_{\mathcal{A}_0}\right]. \end{aligned} \quad (\text{A.14})$$

On the other hand,

$$\begin{aligned} \left\| \frac{1}{\sqrt{n}} \hat{G}_{\mathcal{A}_0} \theta_{\mathcal{A}_0} \right\|^2 &\leq \frac{1}{n \{1 - \max_{j \in \mathcal{A}_0} \hat{\delta}_j(\nu, \tau)\}^2} \\ &\cdot \sum_{j \in \mathcal{A}_0} \frac{(\nu \theta_j)^2}{\hat{\theta}_j^{(0)2(1+\tau)}} = \frac{\nu^2}{n \{1 - \max_{j \in \mathcal{A}_0} \hat{\delta}_j(\nu, \tau)\}^2} \\ &\cdot \sum_{j \in \mathcal{A}_0} \left| \hat{\theta}_j^{(0)(-\tau)} + (\theta_j - \hat{\theta}_j^{(0)}) \hat{\theta}_j^{(0)(-\tau-1)} \right|^2 \\ &= O_p(n^{-1} \nu^2) \end{aligned}$$

$$\begin{aligned} &\cdot \sum_{j \in \mathcal{A}_0} \left(2 \left| \hat{\theta}_j^{(0)} \right|^{-2\tau} + 2 \left| (\theta_j - \hat{\theta}_j^{(0)}) \hat{\theta}_j^{(0)(-\tau-1)} \right|^2 \right) \\ &\leq O_p(n^{-1} \nu^2) \left(2 \text{smin}_{j \in \mathcal{A}_0} \left| \hat{\theta}_j^{(0)} \right|^{-2\tau} \right. \\ &\quad \left. + 2 \text{smin}_{j \in \mathcal{A}_0} \left| \hat{\theta}_j^{(0)} \right|^{-2\tau-2} \left\| \theta_{\mathcal{A}_0} - \hat{\theta}_{\mathcal{A}_0}^{(0)} \right\|^2 \right) \\ &= O_p\left(\left(n^{r/(2r+1)} \nu \right)^2 n^{-4r/(2r+1)} \iota^{-2\tau} \right) \left(1 \right. \\ &\quad \left. + O_p\left(\iota^{-2} n^{-(4r+1)/(2r+1)} \right) \right) = o_p\left(n^{-4r/(2r+1)} \right), \end{aligned} \quad (\text{A.15})$$

where $\iota = \min_{j \in \mathcal{A}_0} |\hat{\theta}_j^{(0)}|$. Using the same argument, we obtain that

$$\begin{aligned} \left\| \frac{1}{n} \hat{G}_{\mathcal{A}_0} \right\|^2 &= O_p\left(\left(n^{r/(2r+1)} \nu \right)^2 n^{-(6r+2)/(2r+1)} \iota^{-2\tau-2} \right) \\ &= o_p\left(n^{-(6r+2)/(2r+1)} \right). \end{aligned} \quad (\text{A.16})$$

By the simple calculation, we get

$$\hat{\theta}_{\mathcal{A}_0} = \left[\frac{1}{n} \sum_{i=1}^n \hat{P}_{i, \mathcal{A}_0}^T \hat{P}_{i, \mathcal{A}_0} - \frac{1}{n} \hat{G}_{\mathcal{A}_0} \right]^{-1} \frac{1}{n} \sum_{i=1}^n \hat{P}_{i, \mathcal{A}_0}^T Y_i. \quad (\text{A.17})$$

By the fact $(H_1 + hH_2)^{-1} = H_1^{-1} - hH_1^{-1}H_2H_1^{-1} + O(h^2)$, we have

$$\begin{aligned} &\left[\frac{1}{n} \sum_{i=1}^n \hat{P}_{i, \mathcal{A}_0}^T \hat{P}_{i, \mathcal{A}_0} - \frac{1}{n} \hat{G}_{\mathcal{A}_0} \right]^{-1} \\ &= \left[\frac{1}{n} \sum_{i=1}^n \hat{P}_{i, \mathcal{A}_0}^T \hat{P}_{i, \mathcal{A}_0} \right]^{-1} + O(n^{-2/3}). \end{aligned} \quad (\text{A.18})$$

For simplicity, let

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \hat{P}_{i, \mathcal{A}}^T \hat{P}_{i, \mathcal{A}} \\ &= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n Z_{i, \mathcal{A}_1}^T Z_{i, \mathcal{A}_1} & \frac{1}{n} \sum_{i=1}^n Z_{i, \mathcal{A}_1}^T \widehat{W}_{i, \mathcal{A}_2} \\ \frac{1}{n} \sum_{i=1}^n Z_{i, \mathcal{A}_1}^T \widehat{W}_{i, \mathcal{A}_2} & \frac{1}{n} \sum_{i=1}^n \widehat{W}_{i, \mathcal{A}_2}^T \widehat{W}_{i, \mathcal{A}_2} \end{pmatrix} \\ &= \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}. \end{aligned} \quad (\text{A.19})$$

By the inverse of the block matrix, we have

$$\begin{aligned} &\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}^{-1} = \begin{pmatrix} H^{11} & H^{12} \\ H^{21} & H^{22} \end{pmatrix} \\ &= \begin{pmatrix} H_{11.2}^{-1} & -H_{11.2}^{-1} H_{12} H_{22}^{-1} \\ -H_{22.1}^{-1} H_{21} H_{11}^{-1} & H_{22.1}^{-1} + H_{22}^{-1} H_{21} H_{11.2}^{-1} H_{12} H_{22}^{-1} \end{pmatrix}, \end{aligned} \quad (\text{A.20})$$

where $H_{11,2} = H_{11} - H_{12}H_{22}^{-1}H_{21}$. Consequently, we have

$$\widehat{\beta}_{\mathcal{A}_1} = \frac{1}{n}H^{11} \left\{ \sum_{i=1}^n Z_{i,\mathcal{A}_1}^T Y_i - H_{12}H_{22}^{-1} \sum_{i=1}^n \widehat{W}_{i,\mathcal{A}_2}^T Y_i \right\},$$

$$E(\widehat{\beta}_{\mathcal{A}_1} | Z, \widehat{\zeta}) = \beta_{0,\mathcal{A}_1} + \frac{1}{n}H^{11} \left\{ \sum_{i=1}^n Z_{i,\mathcal{A}_1}^T \eta_{\mathcal{A}_2}(\widehat{\zeta}_i) \right. \quad (\text{A.21})$$

$$\left. - H_{12}H_{22}^{-1} \sum_{i=1}^n \widehat{W}_{i,\mathcal{A}_2}^T \eta_{\mathcal{A}_2}(\widehat{\zeta}_i) \right\},$$

$$\text{var}(\widehat{\beta}_{\mathcal{A}_1} | Z, \widehat{\zeta}) = \frac{1}{n^2}$$

$$\cdot H^{11} \sum_{i=1}^n (Z_{i,\mathcal{A}_1} - \widehat{W}_{i,\mathcal{A}_2} H_{22}^{-1} H_{21})^T \text{var}(Y_i) \quad (\text{A.22})$$

$$\cdot (Z_{i,\mathcal{A}_1} - \widehat{W}_{i,\mathcal{A}_2} H_{22}^{-1} H_{21}) H^{11},$$

where the elements of $\eta_{\mathcal{A}_2}(\widehat{\zeta}_i)$ are $\eta_k = f_{0k}(\widehat{\zeta}_{ik}) - B_k^T(\widehat{\zeta}_{ik})\gamma_k$ for $k \in \mathcal{A}_2$. For convenience, we denote

$$D(\widehat{\beta}_{\mathcal{A}_1}) = \text{var}(\widehat{\beta}_{\mathcal{A}_1} | Z, \widehat{\zeta}). \quad (\text{A.23})$$

By (A.21), we have

$$\widehat{\beta}_{\mathcal{A}_1} - E(\widehat{\beta}_{\mathcal{A}_1} | Z, \widehat{\zeta})$$

$$= \frac{1}{n}H^{11} \left(\sum_{i=1}^n Z_{i,\mathcal{A}_1}^T \varepsilon_i - H_{12}H_{22}^{-1} \sum_{i=1}^n \widehat{W}_{i,\mathcal{A}_2}^T \varepsilon_i \right). \quad (\text{A.24})$$

Similar to the proof of (A.16) in Tian, Xue, and Hu [30], we can show that

$$\{D(\widehat{\beta}_{\mathcal{A}_1})\}^{-1/2} \{\widehat{\beta}_{\mathcal{A}_1} - E(\widehat{\beta}_{\mathcal{A}_1} | Z, \widehat{\zeta})\} \xrightarrow{\mathcal{L}} N(0, I_{|\mathcal{A}_1|}). \quad (\text{A.25})$$

We complete the proof of Theorem 3. \square

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Yuping Hu's research was supported by the National Social Science Foundation of China (18BTJ021). Liugen Xue's research was supported by the National Natural Science Foundation of China (11571025, Key Grant: 11331011) and the Beijing Natural Science Foundation (1182002). Sanying Feng's research was supported by the National Nature Science Foundation of China (11501522) and the Startup Research Fund of Zhengzhou University (1512315004).

References

- [1] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 1997.

- [2] H. Cardot, F. Ferraty, and P. Sarda, "Spline estimators for the functional linear model," *Statistica Sinica*, vol. 13, no. 3, pp. 571–592, 2003.
- [3] H. Lian and G. Li, "Series expansion for functional sufficient dimension reduction," *Journal of Multivariate Analysis*, vol. 124, pp. 150–165, 2014.
- [4] Y. Fan, G. M. James, and P. Radchenko, "Functional additive regression," *The Annals of Statistics*, vol. 43, no. 5, pp. 2296–2325, 2015.
- [5] S. Feng and L. Xue, "Partially functional linear varying coefficient model," *Statistics. A Journal of Theoretical and Applied Statistics*, vol. 50, no. 4, pp. 717–732, 2016.
- [6] P. Yu, Z. Zhang, and J. Du, "A test of linearity in partial functional linear regression," *Metrika. International Journal for Theoretical and Applied Statistics*, vol. 79, no. 8, pp. 953–969, 2016.
- [7] J. Zhou, J. Du, and Z. Sun, "M-estimation for partial functional linear regression model," *Communications in Statistics—Theory and Methods*, vol. 45, no. 21, pp. 6436–6446, 2016.
- [8] H.-G. Müller and F. Yao, "Functional additive models," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1534–1544, 2008.
- [9] H. Zhu, F. Yao, and H. H. Zhang, "Structured functional additive regression in reproducing kernel Hilbert spaces," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 3, pp. 581–603, 2014.
- [10] C. J. Stone, "Additive regression and other nonparametric models," *The Annals of Statistics*, vol. 13, no. 2, pp. 689–705, 1985.
- [11] L. Wang and L. Yang, "Spline-backfitted kernel smoothing of nonlinear additive autoregression model," *The Annals of Statistics*, vol. 35, no. 6, pp. 2474–2503, 2007.
- [12] J. Huang, J. L. Horowitz, and F. Wei, "Variable selection in nonparametric additive models," *Journal of the American Statistical Association*, vol. 38, no. 4, pp. 2282–2313, 2010.
- [13] P. Zhao and L. Xue, "Empirical likelihood for nonparametric components in additive partially linear models," *Communications in Statistics—Simulation and Computation*, vol. 42, no. 9, pp. 1935–1947, 2013.
- [14] F. Ferraty and P. Vieu, "Additive prediction and boosting for functional data," *Computational Statistics & Data Analysis*, vol. 53, no. 4, pp. 1400–1413, 2009.
- [15] M. Febrero-Bande and W. Gonzalez-Manteiga, "Generalized additive models for functional data," *Test*, vol. 22, no. 2, pp. 278–292, 2013.
- [16] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996.
- [17] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [18] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [19] M. Ueki, "A note on automatic variable selection using smooth-threshold estimating equations," *Biometrika*, vol. 96, no. 4, pp. 1005–1011, 2009.
- [20] P. Lai, Q. Wang, and H. Lian, "Bias-corrected GEE estimation and smooth-threshold GEE variable selection for single-index models with clustered data," *Journal of Multivariate Analysis*, vol. 105, no. 1, pp. 422–432, 2012.

- [21] G. R. Li, H. Lian, S. Y. Feng, and L. X. Zhu, "Automatic variable selection for longitudinal generalized linear models," *Computational Statistics & Data Analysis*, vol. 61, pp. 174–186, 2013.
- [22] R. Q. Tian, L. G. Xue, and D. K. Xu, "Automatic variable selection for varying coefficient models with longitudinal data," *Statistics & Probability Letters*, vol. 119, pp. 84–90, 2016.
- [23] M. Escabias, A. M. Aguilera, and M. J. Valderrama, "Modeling environmental data by functional principal component logistic regression," *Environmetrics*, vol. 16, no. 1, pp. 95–107, 2005.
- [24] S. Jozić, B. Lela, and D. Bajić, "A new mathematical model for flank wear prediction using functional data analysis methodology," *Advances in Materials Science and Engineering*, vol. 2014, Article ID 138168, 8 pages, 2014.
- [25] K. P. Pokhrel and C. P. Tsokos, "Forecasting age-specific brain cancer mortality rates using functional data analysis models," *Advances in Epidemiology*, vol. 2015, Article ID 721592, 11 pages, 2015.
- [26] G. Aneiros-Pérez and P. Vieu, "Semi-functional partial linear regression," *Statistics & Probability Letters*, vol. 76, no. 11, pp. 1102–1110, 2006.
- [27] L. Wang, "GEE analysis of clustered binary data with diverging number of covariates," *The Annals of Statistics*, vol. 39, no. 1, pp. 389–417, 2011.
- [28] L. L. Schumaker, *Spline Functions*, Wiley, New York, NY, USA, 1981.
- [29] P. Zhao and L. Xue, "Variable selection for semiparametric varying coefficient partially linear models," *Statistics & Probability Letters*, vol. 79, no. 20, pp. 2148–2157, 2009.
- [30] R. Tian, L. Xue, and Y. Hu, "Smooth-threshold GEE variable selection for varying coefficient partially linear models with longitudinal data," *Journal of the Korean Statistical Society*, vol. 44, no. 3, pp. 419–431, 2015.



Hindawi

Submit your manuscripts at
www.hindawi.com

