Statistica Sinica Preprint No: SS-2024-0358	
Title	Estimation and Model Selection Procedures in Generalized Functional Partially Additive Hybrid Model with Diverging Number of Covariates
Manuscript ID	SS-2024-0358
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0358
Complete List of Authors	Yanxia Liu, Zhihao Wang, Yu Zhen, Wolfgang K. Härdle and Maozai Tian
Corresponding Authors	Maozai Tian
E-mails	mztian@ruc.edu.cn
Notice: Accepted author version.	

ESTIMATION AND MODEL SELECTION PROCEDURES IN GENERALIZED FUNCTIONAL PARTIALLY ADDITIVE HYBRID MODEL WITH DIVERGING NUMBER OF COVARIATES

Yanxia Liu[1], Zhihao Wang[2], Yu Zheng[1] , Wolfgang K. Härdle[3], Maozai Tian[4]

1. *College of Science, North China Institute of Science and Technology, Hebei, 065201, China*

2. *Institute of Statistics and Data Science, Xinjiang University of Finance and Economics, Xinjiang, 830012, China*

3. *School of Business and Economics, Humboldt-Universität Zu Berlin, Berlin, D-10178, Germany; Institute Digital Assets, Academy of Economic Sciences in Bucharest Romania*

3. *Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China*

Abstract: **We study generalized functional partially additive hybrid model (GFPAHM) where the explanatory variables involve both infinite dimensional predictor processes viewed as functional data with measurement errors, and high-dimensional scalar covariates whose impact on the response is nonlinear. Despite extensive work focusing on functional linear models, little effort has been devoted to estimate coefficients and selecting the important additive components for the GFPAHM, which is complicated by the infinite-dimensional functional predictor. We investigate a nonconvex penalized likelihood estimator for simultaneous variable selection and estimation. The proposed method and theoretical development are quite challenging since the numbers of nonlinear components increase as the sample size increases. Asymptotic properties of the proposed shrinkage estimators are investigated. Extensive Monte Carlo simulations have been conducted and show that the proposed**

procedure works effectively even with moderate sample sizes, and analyze the biscuit dough data set as an illustration.

Key words and phrases: nonparametric function estimation; measurement error; principal components analysis; B-spline; group SCAD.

1. Introduction

As new technology being increasingly used in data collection and storage, many variables are continuously monitored over time and become functional data (Ramsay (1982)). Extracting useful information from such data for further regression analysis has become a challenging statistical problem. The most popular model is the functional linear model (FLM), see for example, Cardot et al. (2003); Reiss and Ogden (2007); Delaigle and Hall (2012) and many others. In many cases, the predictor contains high-dimensional scalar covariates and functional predictors, which are usually called mixed data model. For example, the partial functional linear regression model(PFLRM), proposed by Zhang et al. (2007), has attracted a lot of interests in the literature. This framework focuses on inferring the effect of important non-functional predictors while simultaneously accounting for additional information from functional predictors. It has the advantage of leading to more interpretable results than a purely functional linear model. Kong et al. (2016) considered the PFLRM in the framework of conditional mean

*Corresponding author. E-mail: mztian@ruc.edu.cn

function of the response and used regularization methods for selecting important predictors and estimating slope functions or coefficients. Cao et al. (2020) considered the generalized partially functional linear regression model (GPFLRM) via a known link function. Ding et al. (2018) considered a hybrid of functional and varying-coefficient regression models for the analysis of mixed functional data. Du et al. (2019) proposed the generalized functional partially additive hybrid model (GFPAHM), which is a powerful and outstanding tool for analysing the relationship between a discrete response variable and multiple explanatory variables. It is developed to model partial functional linear components while the remaining components are modeled nonparametrically to combine the strengths of both the GPFLM and the GAM for interpretability and flexibility, while circumventing the curse of dimensionality.

Our research is motivated by the chemical composition prediction of biscuit dough pieces. Getting timely and reliable predictions on the chemical composition of biscuit dough pieces is crucial for producers and merchants to create appropriate strategies for the storage and trade of food. It is well known that chemical composition has a significant impact on the food safety, and statistical models can be used to relate chemical composition forecast to food safety prediction. The biscuit dough data was obtained from an experiment done to test the feasibility of near-infrared (NIR) spectroscopy. An NIR reflectance spectrum is recorded for each dough piece. This data set contains 700 points in total measured from 1100 to 2498 nanometers(nm) in steps of 2 nm, it

is natural to treat them as functional predictors, which exhibit non-negligible correlations. Besides the functional predictors, scalar predictors, such as sucrose, dry flour, and water, also have a great impact on chemical composition and need to be included in the prediction model.

To address this issue, this paper develops a generalized functional partially additive hybrid model using both scalar and functional predictors. We use such a model to predict chemical composition and assist producers and stake holders to better predict the future prices of biscuit commodity products and plan their actions accordingly. To our knowledge, there has been no work covering robust estimation for this model with high dimensional scalar covariates, while Du et al. (2019) studied this type of model with a finite number of scalar covariates that does not require penalization. In this paper, we propose to use the group SCAD for variable selection based on a spline approximation to the additive components and functional principal component basis function approximation to the slope function. With this approximation, each nonparametric component is represented by a linear combination of basis functions. Consequently, the problem of additive component selection becomes that of selecting the groups of coefficients in the linear combinations. This may be associated with several technical challenges. Functional data are usually not directly observed but rather have only intermittent noisy measurements, especially when the number p_n of scalar covariates is permitted to increase with sample size n. In an attempt to select variables and estimate parameters

simultaneously, we extend the application of the SCAD penalty to a nonparametric setting, so the asymptotic properties of the penalized estimators seem more difficult with quasi-likelihood as we need to simultaneously deal with the nonconvex penalty function, approximation of nonlinear functions and very high dimensionality. Besides developing a computational algorithm, asymptotic properties of the proposed shrinkage estimators are investigated. We also establish the simulation results of GFPAHM when the functional data are sparse and irregular data with measurement error. Such results are new in nonparametric settings.

The rest of the article is organized as follows. We propose estimation and a class of variable selection procedures for the GFPAHM when the dimension of the nonparametric functions diverges in Section Section 2. In Section 3, asymptotic properties including estimation consistency and the oracle property results are explored. Simulation studies and real data analysis are presented in Section 4 and 5. Discussion and concluding remarks can be found in Section 6. Appendix gives the proof of the theorem.

2. Model and estimation

2.1 Model with measurement errors

Suppose that we observe a random sample $\{X_i(t), \mathbf{Z}_i, Y_i, i = 1, \cdots n\}$ be independent and identically distributed copies of $\{X(t), \mathbf{Z}, Y\}$, where the functional predictor $\{X_i(t), t \in \mathcal{T}\}$ is a zero mean, second-order stochastic process defined on a probability

space $(\Omega, \mathscr{B}, P)$ (that is $\mathrm{E}|X_i(t)|^2 < \infty, \forall t \in \mathcal{T}$) with sample paths in $L^2(\mathcal{T})$, Without loss of generality, we suppose throughout the paper that $\mathcal{T} = [0, 1]$. The response Y_i is a scalar and $\mathbf{Z_i} = (Z_{i1}, \ldots, Z_{ip_n})^\top \in R^{p_n}$ is a p_n-dimensional covariate vector.

One of the advantages of marginal approaches is that we only need to specify the first two moments by $E(Y_i|\mathbf{Z}_i, X_i) = \mu_i(\mathbf{Z}_i, X_i)$, and the conditional variance function $\mathrm{Var}(Y_i|\mathbf{Z}_i, X_i) = \sigma^2 V(\mu_i(\mathbf{Z}_i, X_i))$ for some known positive function $V(\cdot)$. In this article, the known link function $g(\cdot)$ is modeled as follows

$$g(\mu(\mathbf{Z}_i, X_i)) = h(\mathbf{Z}_i) + \int_0^1 \alpha(t) X_i(t) \mathrm{d}t. \tag{2.1}$$

Let $\mathcal{G}_n$ is the collection of functions $h(\cdot)$ with the additive form $h(\mathbf{Z}_i) = \sum_{j=1}^{p_n} h_j(Z_{ij})$, where Z_{ij} is the jth component of $\mathbf{Z}_i$. $\{h_j(\cdot), j = 1, \cdots p_n\}$ are unknown smooth functions and satisfy $E(h_j(Z_{ij})) = 0$ for $1 \le j \le p_n$ for identification purposes. Motivated by real examples, we consider that the dimension p_n is large and varies according to the sample size n, i.e., $p_n = O(n^\alpha)$ for some $\alpha > 0$. Suppose that some of the additive components h_j are zero. Model (2.1) is called the generalized functional partially additive hybrid model (GFPAHM), where $h(\mathbf{Z}_i)$ are the nonparametric components, and $\int_0^1 \alpha(t) X_i(t) \mathrm{d}t$ is the functional linear components, as an important extension of generalized regression model.

It often happens that the functional data are not fully observed but rather, for each subject i, we only get intermittent noisy measurements, i.e., $W_{ij} = X_i(t_{ij}) + \varepsilon_{ij}$, where the measurement errors ε_{ij} are independent of X_i, $E(\varepsilon_{ij}) = 0$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma_\varepsilon^2$ for

all $j \in \{1, \cdots, s_i\}$. We assume that W_{ij} is recorded on a dense grid of points and this allows us to run a statistical smoother through (t_{ij}, W_{ij}), for definiteness, we consider the regression splines smoother for each set of subjects, and denote the smoothed trajectories by $\{\hat{X}_i, i = 1, \cdots, n\}$. The latter is then used to analysis GFPAHM.

The first goal of this article is to provide a simple method of estimating slope functions $\alpha(\cdot)$ and the additive components $\{h_j(\cdot), j = 1, \cdots, p_n\}$ in model (2.1) based on a quasi-likelihood procedure (refer to Severini and Staniswalis (1994) for more details). The second goal is to distinguish the nonzero additive components from the zero components in this generalized hybrid model framework. We apply the group SCAD to select nonzero components and reduce the dimension of the problem.

2.2 Quasi-likelihood estimation and variable selection via group SCAD

In this paper, we will use polynomial splines to approximate the nonparametric functions $\{h_j(\cdot), j = 1, \cdots, p_n\}$ in model (2.1), which is computationally convenient and often accurate since it directly convert a problem with infinite-dimensional parameters to one with a finite number of parameters. The use of polynomial spline smoothing in generalized nonparametric model can go back to Stone (1986), who first obtained the rate of convergence of the polynomial spline estimates for the generalized additive model.

As in most work on nonparametric smoothing, estimation of the functions $\{h_j(\cdot), j = 1, \cdots, p_n\}$ is conducted on compact sets. Without loss of generality, let the compact

set be $\mathcal{X} = [0, 1]$. Let $\mathcal{S}_n$ be the space of polynomial splines of order $h \geq 1$. There exists a B-spline basis $\{b_{k,j}(z_j), k = 1, \cdots, J_n + h, j = 1, \cdots, p_n\}$ for $\mathcal{S}_n$, where h is the order of basis functions and J_n is the number of the interior knots. Equally spaced knots are adopted in the following for simplicity. Let $K_n = J_n + h - 1$. For ease of simplicity of asymptotic analysis and computation, we adopt the normalized B-spline space $\mathcal{S}_n^0$ introduced in Wang et al. (2011) with the following normalized basis

$$B_{k,j}(z_j) = \sqrt{K_n}\left\{b_{k+1,j}(z_j) - \frac{E\left(b_{k+1,j}(z_j)\right)}{E\left(b_{1,j}(z_j)\right)} b_{1,j}(z_j)\right\}, \quad j = 1, \cdots, p_n, \\ k = 1, \cdots, K_n,$$

then $E(B_{k,j}(z_j)) = 0$. The subspace is K_n dimensional due to the empirical version of the constraint $\sum_{i=1}^{n} h_j(Z_{ij}) = 0$. Denote $\boldsymbol{B}(\boldsymbol{z}) = (\boldsymbol{B}_1^\top(z_1), \cdots, \boldsymbol{B}_{p_n}^\top(z_{p_n}))^\top$, and $\boldsymbol{B}_j(z_j) = (B_{1,j}(z_j), \cdots, B_{K_n,j}(z_j))^\top$. Under suitable smoothness assumptions, for any $h_j(z_j)$, write

$$h_j(z_j) \approx \boldsymbol{B}_j^\top(z_j)\boldsymbol{\zeta}_j = \sum_{k=1}^{K_n} B_{k,j}(z_j)\zeta_{jk}, \quad j = 1, \cdots, p_n,$$

where $\boldsymbol{\zeta}_j = (\zeta_{j1}, \cdots, \zeta_{jK_n})^\top$ is the spline coefficient vector.

To estimate the function $\alpha(t)$ and the regression coefficients $\{\boldsymbol{\zeta}_j, j = 1, \cdots, p_n\}$, we consider maximizing the following quasi-likelihood function, which couples $\alpha(t) = \sum_{j=1}^{\infty} \gamma_j \hat{\phi}_j(t)$ with $X_i(t) = \sum_{j=1}^{\infty} \hat{\xi}_{ij}\hat{\phi}_j(t)$, for each $i = 1, \cdots, n$ given the complete orthonormal basis series $\{\hat{\phi}_j(t), j = 1, \cdots, \infty\}$,

$$L(\boldsymbol{\zeta}, \boldsymbol{\gamma}) \approx \sum_{i=1}^{n} Q[g^{-1}(\sum_{j=1}^{p_n}\sum_{k=1}^{K_n} B_{k,j}(Z_{ij})\zeta_{jk} + \sum_{j=1}^{\infty}\hat{\xi}_{ij}\gamma_j), Y_i], \tag{2.2}$$

where $\hat{\xi}_{ij} = \int_0^1 X_i(t)\hat{\phi}_j(t)\mathrm{d}t$ are FPC scores, they are uncorrelated with mean zero and $\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{ij}\hat{\xi}_{ik} = I(j=k)\hat{\nu}_j$, $\hat{\nu}_1 > \hat{\nu}_2 > \cdots \geq 0$ are the ordered eigenvalue sequences, and quasi-likelihood function $Q(m, y)$ satisfies

$$\frac{\partial}{\partial m}Q(m,y) = \frac{y-m}{V(m)}.$$

Since X_i contains measurement errors, $\hat{X}_i$ is used in the subsequent FPC analysis to get empirical estimates for $\hat{\nu}_j, \hat{\phi}_j$ and $\hat{\xi}_{ij}$.

Although the number of principal components is infinite, it is generally assumed that the information related to the response variable is mainly captured by the first m_n FPC scores. The truncated number satisfies $m_n \to \infty$ as the sample size $n \to \infty$. Then we can obtain $\int_0^1 \alpha(t)X_i(t)\mathrm{d}t \approx \sum_{j=1}^{m_n}\hat{\xi}_{ij}\gamma_j \triangleq \hat{\boldsymbol{\xi}}_i^\top\boldsymbol{\gamma}$, where $\hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i1}, \cdots, \hat{\xi}_{im_n})^\top$, $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_{m_n})^\top$. So (2.2) is rewritten as

$$L(\boldsymbol{\zeta}, \boldsymbol{\gamma}) \triangleq \sum_{i=1}^{n} Q[g^{-1}(\boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta} + \hat{\boldsymbol{\xi}}_i^\top\boldsymbol{\gamma}), Y_i], \tag{2.3}$$

where $\boldsymbol{B}(\boldsymbol{Z}_i) = (\boldsymbol{B}_1^\top(Z_{i1}), \cdots, \boldsymbol{B}_{p_n}^\top(Z_{ip_n}))^\top, \boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^\top, \cdots, \boldsymbol{\zeta}_{p_n}^\top)^\top$.

In order to remove all the non-significant covariates in $\{Z_{ij}, j = 1, \cdots, p_n\}$, we call on the modern shrinkage penalty function $p_\lambda(\cdot)$. The L_1 penalty or Lasso (Tibshirani (1996)) is a popular choice for penalized estimation. However, the L_1 penalty is known to over-penalize large coefficients, tends to be biased and requires strong conditions on the design matrix to achieve selection consistency. This is usually not a concern for prediction, but can be undesirable if the goal is to identify the underlying model. In

this paper, we consider the SCAD penalty function with oracle properties proposed by Fan and Li (2001). The specific expression of the penalty function is as follows

$$p_\lambda(\theta) = \begin{cases} \lambda\theta, & 0 \le \theta \le \lambda, \\ -\left(\theta^2 - 2a\lambda\theta + \lambda^2\right) / \left\{2\left(a-1\right)\right\}, & \lambda < \theta \le a\lambda, \\ \left(a+1\right)\lambda^2/2, & \theta > a\lambda, \end{cases}$$

where $a > 2$, $p_\lambda(0) = 0$ and λ is a nonnegative penalty parameter and governs sparsity of the model. Arguing from a Bayesian statistical point of view, Fan and Li (2001) suggested using $a = 3.7$. This value will be used in Section 4. Building upon the quasi-likelihood given in (2.3), We adopt the group descent algorithms introduced in Breheny and Huang (2015). Therefore, the goals of the variable selection and estimation can be achieved by maximizing

$$\arg\max_{\boldsymbol{\zeta},\boldsymbol{\gamma}} PL(\boldsymbol{\zeta},\boldsymbol{\gamma}) = \sum_{i=1}^{n} Q\left[g^{-1}\left(\boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta} + \hat{\boldsymbol{\xi}}_i^\top\boldsymbol{\gamma}\right), Y_i\right] - n\sum_{j=1}^{p_n} p_{\lambda_n}\left(\|\boldsymbol{\zeta}_j\|\right), \tag{2.4}$$

where $\|\boldsymbol{\zeta}_j\|$ is the Euclidean norm invoking a group penalty. The resulting estimators for $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$ obtained from (2.4) are denoted by $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\gamma}}$. Then, the estimator of $\alpha(t)$ is denoted by $\hat{\alpha}(t) = \sum_{j=1}^{m_n} \hat{\gamma}_j\hat{\phi}_j(t)$, the spline estimator of $\hat{h}(\boldsymbol{z}) = \boldsymbol{B}^\top(\boldsymbol{z})\hat{\boldsymbol{\zeta}}$, and the centered spline estimators of the component functions are

$$\hat{h}_j(z_j) = \sum_{k=2}^{K_n} B_{k,j}\left(z_j\right)\hat{\zeta}_{jk} - n^{-1}\sum_{i=1}^{n}\sum_{k=2}^{K_n} B_{k,j}\left(Z_{ij}\right)\hat{\zeta}_{jk}, \quad j = 1, \cdots, p_n.$$

3. Asymptotic properties

In this section, we establish the asymptotic properties of the estimators of our regularized generalized functional partially additive hybrid model. Denote $\boldsymbol{\mathcal{F}} = (\boldsymbol{Z}, X), m_0(\boldsymbol{\mathcal{F}}) = h_0(\boldsymbol{Z}) + \int_0^1 \alpha_0(t)X(t)\mathrm{d}t$, where $\alpha_0(t)$ and $h_0(\boldsymbol{Z})$ are the true values of $\alpha(t)$ and $h(\boldsymbol{Z})$. Without loss of generality, we assume that in the true model, only the first s additive components are nonzero, where s is a fixed and finite integer. The remaining additive components are all zeros. Let $h_0(\mathbf{Z}_i) = \sum_{j=1}^{p_n} h_{0j}(Z_{ij}) = \sum_{j=1}^{s} h_{0j}(Z_{ij}) + \sum_{j=s+1}^{p_n} h_{0j}(Z_{ij})$, with $h_{0j} = 0$ almost surely for $j = s+1, \cdots, p_n$. We write the collections of all components, $\mathbf{Z}_i = \left(\mathbf{Z}_{i\mathcal{S}}^\top, \mathbf{Z}_{i\mathcal{N}}^\top\right)^\top$, where $\mathbf{Z}_{i\mathcal{S}} = (Z_{i1}, \cdots, Z_{is})^\top, \mathbf{Z}_{i\mathcal{N}} = (Z_{i(s+1)}, \cdots, Z_{ip_n})^\top$. Similarly, we can define $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_{\mathcal{S}}^\top, \boldsymbol{\zeta}_{\mathcal{N}}^\top)^\top$, where $\boldsymbol{\zeta}_{\mathcal{S}} = (\boldsymbol{\zeta}_1^\top, \cdots, \boldsymbol{\zeta}_s^\top)^\top$ and $\boldsymbol{\zeta}_{\mathcal{N}} = (\boldsymbol{\zeta}_{s+1}^\top, \cdots, \boldsymbol{\zeta}_{p_n}^\top)^\top$. Denote $a_n = \max_j\{|p'_{\lambda_n}(\|\boldsymbol{\zeta}_j\|)|, \boldsymbol{\zeta}_j \neq \mathbf{0}\}$, $b_n = \max_j\{|p''_{\lambda_n}(\|\boldsymbol{\zeta}_j\|)|, \boldsymbol{\zeta}_j \neq \mathbf{0}\}$. Similar to the notation of Wang et al. (2011), let

$$q_l(m, y) = \frac{\partial^l}{\partial m^l} Q\{g^{-1}(m), y\}, \quad \rho_l(m) = \{\frac{\mathrm{d}g^{-1}(m)}{\mathrm{d}m}\}^l / V\{g^{-1}(m)\}.$$

In particular, there is the following formula

$$\begin{aligned} q_1(m, y) &= \{y - g^{-1}(m)\}\rho_1(m), \\ q_2(m, y) &= \{y - g^{-1}(m)\}\rho_1'(m) - \rho_2(m). \end{aligned}$$

In the theoretical analysis, we begin by making several critical regularity conditions. The next three conditions are fundamental conditions in FLM, which are adopted by Yao et al. (2017) and Kong et al. (2016).

C1: The functional data $X(\cdot)$ satisfies $\mathrm{E}\|X(\cdot)\|^4 \leqslant C < \infty$, and for each scores ξ_j, $\mathrm{E}(\xi_j^4) \leqslant C\nu_j^2$ for $j \geqslant 1$.

C2: Regarding the eigenvalues ν_j, we impose $\nu_j - \nu_{j+1} \geqslant C^{-1}j^{-a-1} (j \geqslant 1)$, $\quad C^{-1}j^{-a} \leqslant \nu_j \leqslant Cj^{-a}$, and the Fourier coefficients γ_j satisfies $|\gamma_j| \leqslant Cj^{-b}$, $\quad j \geqslant 1$, where $a > 1$, $b > a/2 + 1$.

C3: For the truncation parameter m_n satisfies $m_n = O(n^{1/(a+2b)})$ as $n \to \infty$.

C4: The function $q_2(m, y) < 0$ and $C_1 < |q_2^{\iota}(m, y)| < C_2 (\iota = 0, 1)$, where $m \in R$. This ensures the uniqueness of the solution, which is the same as Condition 1a in Carroll et al. (1997).

Similar to the conditions in Wang et al. (2011), we further impose the conditions C5-C7, where C5 is often assumed in asymptotic analysis of nonparametric regression problems which requires a boundedness condition on the covariates.

C5: The distribution of $\boldsymbol{Z}$ is absolutely continuous and its density f is bounded away from zero and infinity on $[0, 1]^{p_n}$.

C6: $|\rho_\ell(m_0)| \leqslant C_\rho$, $|\rho_\ell(m) - \rho_\ell(m_0)| \leqslant C_\rho^* |m - m_0|, \ell = 1, 2$, for all $m - m_0 < C_m$, where C_ρ, C_ρ^* and C_m are positive constants.

C7: $E\left[\{Y - g^{-1}(m_0(\boldsymbol{\mathcal{F}}))\}^2 \mid \boldsymbol{\mathcal{F}}\right] \leqslant C$, where C represents a generic positive constant.

C8: $p_n = O(n^C)$ for some $C < \frac{1}{3}$. Such p_n is allowed to diverge with the sample size n, which is adopted by Sherwood and Wang (2016).

C9: For any unit vector $\omega \in R^{p_n}$, $c \leqslant \omega^\top E(\boldsymbol{Z}\boldsymbol{Z}^\top|U=u)\omega \leqslant C$.

This condition guarantees that the eigenvalues of $E(\boldsymbol{Z}\boldsymbol{Z}^\top|U=u)$ are bounded away from 0 and $\propto$.

For the identifiability of additive components in model (2.1), we impose that

C10: Let q be a positive integer and $0 < v \leqslant 1$ such that $r = q+v > 2$, $\mathcal{H}(r)$ be the collection of functions $h(\cdot)$ on $[0,1]$ whose qth derivative, $h^{(q)}(\cdot)$, exists and satisfies:

$$\|h^{(q)}(s) - h^{(q)}(t)\| \leqslant C|s-t|^v, \quad 0 \leqslant s,t \leqslant 1,$$

where C is a positive constant.

The next condtion gives the rate of growth of the dimension of the spline spaces relative to the sample size, which is same as that in Wang et al. (2011).

C11: The number of knots $n^{1/(2r)} \leq K_n \leq n^{1/4}$.

To this end, we also need the density function of z_j to be bounded below by a positive constant and we require that

C12: For every K_n there is a nonsingular matrix M such that

i) The smallest eigenvalue of $E[M(B_{k,j}(z_j) - E[B_{k,j}(z_j)])]^{\otimes 2}$ is bounded away from zero uniformly in K_n , where $A^{\otimes 2} = AA^\top$.

ii) There is a sequence of constants $C_n(K_n)$ satisfying $\|\sup_x(MB_{k,j}(x))\| \leq C_n(K_n)$ and $(C_n(K_n))^2K_n/n \to 0$ as $n \to \infty$, where for a matrix A, $\|A\| = \mathrm{tr}(AA^\top)$ denotes the Euclidean norm of A.

Theorem 1. *Assume that conditions C1-C12 are satisfied, and the dimension of the*

spline space $\mathcal{S}_n^0$ *fulfils* $K_n = O_p(n^{1/(1+2r)})$. *Then, if* $\frac{a}{4b} < r < \frac{a+2b-1}{2}$ *and* $a_n \to 0$, $b_n \to 0$ *as* $n \to \infty$, *we have*

$$\|\hat{\alpha}(t) - \alpha_0(t)\| = O_p(\delta_n),$$

$$\|\hat{h}(\boldsymbol{Z}) - h_0(\boldsymbol{Z})\| = O_p(\delta_n).$$

where $\delta_n = \sqrt{(m_n + K_n)n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})}} + a_n$.

Remark 1. Theorem 1 demonstrates that the rate of convergence of the estimators $\hat{\alpha}$ and $\hat{h}$ is $\left((m_n + K_n)n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})}\right)^{-\frac{1}{2}}$ if $\lambda_n \to 0$.

Next, we prove the asymptotic oracle properties of the proposed penalized estimator. The main complicated issue comes from the dependence between $\boldsymbol{Z}_{\mathcal{S}}$ and $\hat{\boldsymbol{\xi}}$. Similar to Wang et al. (2011), let $\boldsymbol{\mathcal{T}}_{\mathcal{S}} = (\hat{\boldsymbol{\xi}}, \boldsymbol{Z}_{\mathcal{S}})$, $\tilde{m}(\boldsymbol{\mathcal{T}}_{\mathcal{S}}) = \hat{\boldsymbol{\xi}}^{\top}\boldsymbol{\gamma} + \boldsymbol{B}^{\top}(\boldsymbol{Z}_{\mathcal{S}})\boldsymbol{\zeta}_{\mathcal{S}}$ and

$$\boldsymbol{\Gamma}_1(\hat{\boldsymbol{\xi}}) = \frac{E[\boldsymbol{Z}_{\mathcal{S}}\rho_2\{\tilde{m}(\boldsymbol{\mathcal{T}}_{\mathcal{S}})\}|\hat{\boldsymbol{\xi}}]}{E[\rho_2\{\tilde{m}(\boldsymbol{\mathcal{T}}_{\mathcal{S}})\}|\hat{\boldsymbol{\xi}}]}, \quad \tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}}) = \boldsymbol{B}(\boldsymbol{Z}_{\mathcal{S}}) - \boldsymbol{\Gamma}_1(\hat{\boldsymbol{\xi}}).$$

Theorem 2. *Assume that conditions C1-C12 are satisfied, and* $J_n = O_p(n^{1/(1+2r)})$. *If* $\lambda_n \to 0$ *and* $\lambda_n(n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})})^{-\frac{1}{2}}/\sqrt{m_n + K_n} \to \infty$ *as* $n \to \infty$, *then the estimator* $\hat{\boldsymbol{\zeta}} = (\hat{\boldsymbol{\zeta}}_{\mathcal{S}}^{\top}, \hat{\boldsymbol{\zeta}}_{\mathcal{N}}^{\top})^{\top}$ *satisfies:*

(a) sparsity, that is, $\hat{\boldsymbol{\zeta}}_{\mathcal{N}} = \boldsymbol{0}$ *with probability tending to one;*

(b) $\sqrt{n}(\boldsymbol{\Sigma}_{\lambda_n} + \Omega_{\mathcal{S}})\{\hat{\boldsymbol{\zeta}}_{\mathcal{S}} - \boldsymbol{\zeta}_{0\mathcal{S}} + (\boldsymbol{\Sigma}_{\lambda_n} + \Omega_{\mathcal{S}})^{-1}\boldsymbol{c}_n\} \xrightarrow{d} N(0, E(\Omega_{\mathcal{S}}))$,

where $\Omega_{\mathcal{S}} = \rho_2\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}})\right)\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{\mathcal{S}})$, $\boldsymbol{c}_n = \left\{p'_{\lambda_n}\left(\|\boldsymbol{\zeta}_{01}\|\right)\frac{\boldsymbol{\zeta}_{01}^{\top}}{\|\boldsymbol{\zeta}_{01}\|}, \ldots, p'_{\lambda_n}\left(\|\boldsymbol{\zeta}_{0s}\|\right)\frac{\boldsymbol{\zeta}_{0s}^{\top}}{\|\boldsymbol{\zeta}_{0s}\|}\right\}^{\top}$, *and* $\boldsymbol{\Sigma}_{\lambda_n} = \mathrm{diag}\left\{p''_{\lambda_n}\left(\|\boldsymbol{\zeta}_{01}\|\right), \ldots, p''_{\lambda_n}\left(\|\boldsymbol{\zeta}_{0s}\|\right)\right\}$.

Remark 2. Theorem 2 shows that the proposed variable selection procedure can identify the true model consistently.

4. Simulation study

In this section, we assess the numerical performance of our proposed method against suitable competitors through Monte Carlo studies. We use 200 simulation runs. The simulated data sets are generated as follows

$$g(\mu) = \sum_{j=1}^{p_n} h_j(Z_j) + \int_0^1 \alpha(t)X(t)\mathrm{d}t, \tag{4.5}$$

where $h_1(x) = 2\sin(2\pi x)$, $h_2(x) = 4x(1-x^2)$, $h_3(x) = 2\log(x+1)$ and $h_j(x) = 0$ for $j = 4, \cdots, p_n$. Therefore, only the first three variables in Z_j are relevant and the rest are null variables. The covariates $(Z_1, \cdots, Z_{p_n})^\top$ are generated from Uniform $([0,1]^{p_n})$ with correlations given by $\mathrm{cov}(Z_j, Z_k) = (1/2)^{|j-k|}$ for $j, k = 1, \cdots, p_n$. Following from Shin (2009), we take $\alpha(t) = 2\sqrt{2}\sin(\pi t/2) + 4\sqrt{2}\sin(3\pi t/2)$ and $X(t) = \sum_{j=1}^{100} \xi_j \phi_j(t)$, where the uncorrelated FPC scores $\xi_j's$ are distributed as independent normal with mean 0 and variance $\nu_j = ((j-0.5)\pi)^{-2}$ and $\phi_j(t) = \sqrt{2}\sin((j-0.5)\pi t)$. For the actual functional predictor trajectories were sampled densely over 100 equally spaced time points $t_{ij} \in [0,1]$ with independent and identically distributed noise $\varepsilon_{ij} \sim N(0, 0.1^2)$. In this paper, the log link function is considered , i.e. $\log(\mu) = \sum_{j=1}^{p_n} h_j(Z_j) + \int_0^1 \alpha(t)X(t)\mathrm{d}t$, and response variable distribution follows Poisson(μ) and $\mathrm{Var}(Y|\mathbf{Z}, X) = \mu(\mathbf{Z}, X)$.

In our numerical examples, we use equally spaced knots with the number of inte-

rior knots $J_n = [n^{1/5}]$, where $[n]$ represents the largest integer no larger than n, and we take the spline order to be $h = 3$. It is well known that tuning parameter selection plays an important role in regularization methods. To implement the proposed method, the group SCAD penalty parameter λ_n and the truncated parameter m_n for the corresponding nonparametric predictors should be chosen appropriately. Specific in practical operation, we use the **select** function provided in the **grpreg** package for computation. Refer to Liu et al. (2022) for detailed steps.

To assess the accuracy of variable selection and the resulting estimates $h_j(\cdot)$ (for $j = 1, 2, 3$) in different methods, we adopt the following criteria:

1. Integrated squared bias (ISB):

$$\mathrm{ISB}(\hat{h}_j(z_j)) = \int_0^1 [E\hat{h}_j(z_j) - h_j(z_j)]^2 \mathrm{d}z_j, \quad j = 1, 2, 3.$$

2. Integrated mean squared error (IMSE):

$$\mathrm{IMSE}(\hat{h}_j(z_j)) = \int_0^1 E[\hat{h}_j(z_j) - h_j(z_j)]^2 \mathrm{d}z_j, \quad j = 1, 2, 3,$$

where the expectation is calculated by the empirical mean of the relevant estimators obtained from the 200 data sets. And the definition of ISB and IMSE for function $\alpha(\cdot)$ is similar.

3. The sample Mean, Median, and Variance of the square root of average square errors (RASE):

$$\mathrm{RASE}_1 = (\frac{1}{T_1}\sum_{j=1}^{T_1}(\hat{\alpha}(t_j) - \alpha(t_j))^2)^{\frac{1}{2}},$$

and

$$\text{RASE}_{j+1} = (\frac{1}{T_2}\sum_{i=1}^{T_2}(\hat{h}_j(z_{ij}) - h_j(z_{ij}))^2)^{\frac{1}{2}}, j = 1, 2, 3,$$

where $\{t_j, j = 1, \cdots, T_1\}$, and $\{z_{ij}, i = 1, \cdots, T_2, j = 1, 2, 3\}$ are grid points chosen to be equally spaced in the domains of function $\alpha(\cdot)$ and $h_j(\cdot)$, respectively. In our simulation, $T_1 = T_2 = 100$ are used.

4. C: the average number of zero coefficients that are correctly estimated to be zero.

5. IC: the average number of the true nonzero coefficients that are incorrectly estimated to be zero.

6. CF: the proportion of times the correct model is selected.

Based on 200 repetitions, we applied the proposed variable selection procedures to the model (4.5), the ISB, IMSE, and the sample means, medians and variances of $\text{RASE}_j (j = 1, 2, 3, 4)$ with different sample sizes for Poisson regression model are listed in Table 1.

Table 1: Simulation results over 200 repetitions

	n, p_n	200,5			400,6			600,7		
	Method	gSCAD	gLASSO	gMCP	gSCAD	gLASSO	gMCP	gSCAD	gLASSO	gMCP
$\hat{h}_1$	ISB	0.007	0.066	0.003	0.002	0.047	0.002	0.001	0.051	0.001
	IMSE	0.128	0.186	0.119	0.090	0.114	0.089	0.035	0.085	0.036
	Mean	0.288	0.365	0.289	0.238	0.302	0.239	0.166	0.275	0.167
	Median	0.222	0.296	0.256	0.197	0.270	0.199	0.146	0.260	0.146
	Var	0.045	0.053	0.036	0.033	0.023	0.033	0.007	0.010	0.008
$\hat{h}_2$	ISB	0.001	0.018	0.001	0.003	0.023	0.004	0.004	0.032	0.004
	IMSE	0.114	0.107	0.124	0.101	0.077	0.101	0.053	0.066	0.052
	Mean	0.278	0.282	0.284	0.248	0.238	0.249	0.196	0.225	0.194
	Median	0.222	0.252	0.219	0.177	0.211	0.183	0.154	0.196	0.153
	Var	0.037	0.027	0.043	0.040	0.021	0.039	0.015	0.016	0.015
$\hat{h}_3$	ISB	0.000	0.019	0.001	0.008	0.027	0.009	0.003	0.030	0.003
	IMSE	0.099	0.088	0.111	0.066	0.072	0.065	0.032	0.053	0.033
	Mean	0.258	0.256	0.260	0.204	0.228	0.203	0.150	0.202	0.152
	Median	0.216	0.222	0.199	0.157	0.192	0.155	0.119	0.181	0.121
	Var	0.032	0.023	0.044	0.024	0.020	0.024	0.010	0.012	0.010
$\hat{\alpha}$	ISB	0.002	0.023	0.002	0.001	0.009	0.001	0.002	0.009	0.002
	IMSE	0.214	0.242	0.195	0.109	0.117	0.110	0.062	0.072	0.062
	Mean	0.442	0.471	0.418	0.312	0.325	0.313	0.239	0.256	0.239
	Median	0.413	0.444	0.388	0.294	0.300	0.295	0.225	0.245	0.225
	Var	0.018	0.021	0.021	0.012	0.012	0.012	0.005	0.007	0.005

In addition, in Table 2, we also investigate the model identification results for our estimator, in which the last column "Time" is recorded average running time. Rows refer to methods, where gSCAD, gLASSO and gMCP represent the variable selection methods via group SCAD, group LASSO and group MCP, respectively. From the simulation results in Tables 1 and 2, we can see that the performance of the variable selection method based on gSCAD for the important additive components become

Table 2: Simulation results over different sample sizes comparing gSCAD, gLASSO and gMCP.

n, p_n	Method	C	IC	CF	Time
200,5	Oracle	2	0	1	1.4510
	gSCAD	1.9600	0.0200	0.9450	1.8859
	gLASSO	0.6950	0.0000	0.1700	2.1478
	gMCP	1.9800	0.0350	0.9550	1.8894
400,6	Oracle	3	0	1	2.7272
	gSCAD	2.9900	0.0300	0.9700	4.6746
	gLASSO	1.2850	0.0000	0.1750	5.2271
	gMCP	2.9900	0.0300	0.9700	4.7387
600,7	Oracle	4	0	1	3.6045
	gSCAD	4.0000	0.0000	1.0000	7.9187
	gLASSO	2.3800	0.0000	0.2600	9.6041
	gMCP	4.0000	0.0000	1.0000	8.2049

increasingly better in terms of model complexity as the increase of sample size n and p_n in each case. The performance of the proposed variable selection method gSCAD has smaller IMSE, ISB and RASE as the sample size increases and the value of ISB is smaller than that of IMSE. We also can see that the performances of both gSCAD and gMCP procedures become better in terms of correctly identifying zero coefficients and gSCAD runs a little faster than gMCP according to the Table 2, and with the increase of sample size n, the final model is more and more close to the real model. Correct estimation model of the number of the proportion of (CF) along with the augment of sample size n also increased steadily, and adopt gSCAD variable selection method of CF were higher than gLASSO method under corresponding CF, this shows that our proposed variable selection method can improve the accuracy of estimates.

For the functional predictors, we further consider the sparse variants with high sparsity, that is, $85\% - 95\%$ missings as well as i.i.d. measure error $\varepsilon \sim 0.05N(0,1)$. The sparsification mechanism is the same as Yao et al. (2005) and is applied to each observation and element separately. We present the simulation results of GFPAHM when the functional data are sparse data with measurement error. Figures 1 and 2 display the estimates and the empirical 95% point-by-point confidence intervals for the $h_1(\cdot)$, $h_2(\cdot)$, $h_3(\cdot)$ and $\alpha(\cdot)$ under the log link function when $n = 200$ and 800, respectively. The true curves are well recovered and the confidence intervals have stable performance under different settings. Though the functional data are sparse data with measurement error, the overall performance is satisfactory.

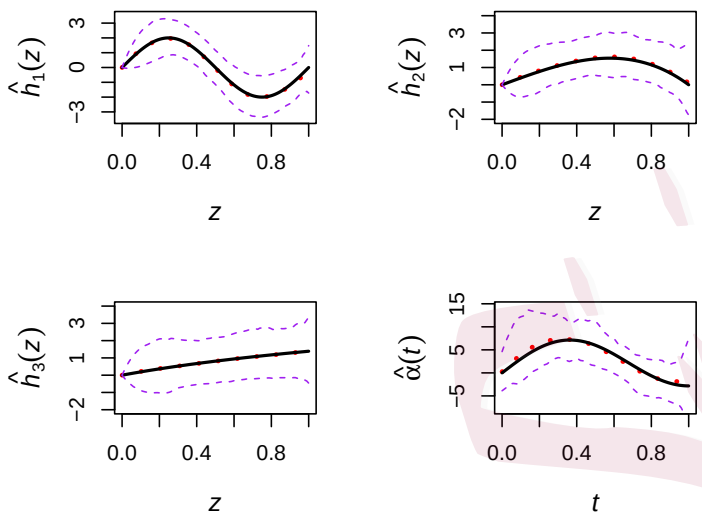

Figure 1: Estimated results for $h_1(\cdot), h_2(\cdot), h_3(\cdot)$ and $\alpha(\cdot)$ with $n = 200$ under sparse functional data. The black solid lines are the true functions, the red dotted lines represent the estimated functions and the purple dashed lines are the 95% point-by-point confidence intervals.

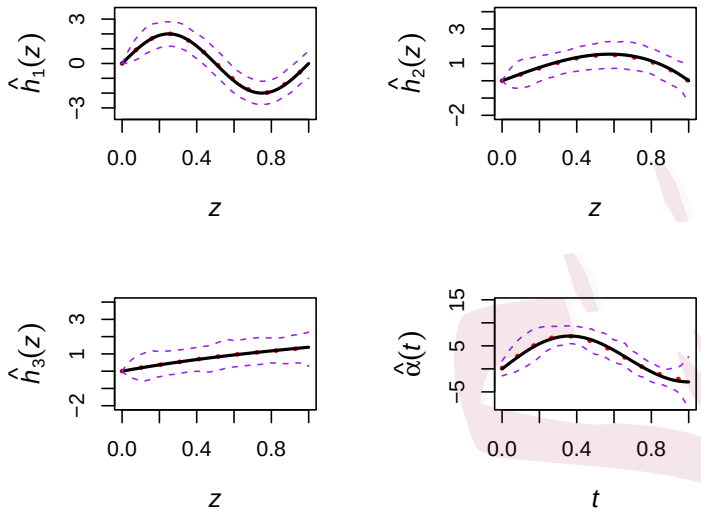

Figure 2: Estimated results for $h_1(\cdot), h_2(\cdot), h_3(\cdot)$ and $\alpha(\cdot)$ with $n = 800$ under sparse functional data. The black solid lines are the true functions, the red dotted lines represent the estimated functions and the purple dashed lines are the 95% point-by-point confidence intervals.

5. An application to the biscuit dough data

We evaluate the performance of our approach on a data set, which was discussed in details by Brown et al. (2001) and Du et al. (2018). Consider generalized functional partially additive hybrid model(GFPAHM) in (2.1) for the data, where we take the water, dry flour and sucrose to be the standardized covariates variable Z_1(water), Z_2(dry flour) and Z_3(sucrose) separately, and the spectra to be $X(t)$. We use identity link with

the fat as the response. The study focuses on studying the relationship between the fat and other ingredient in biscuit dough and examining whether there are any quadratic effects and interaction effects from these covariates. Thus, we add the quadratic terms and interaction term of Z_1, Z_2 and Z_3 to the initial full model, which is adopted by Fan and Li (2004), and consider the following model:

$$\begin{aligned} g(\mu(\mathbf{Z}, X)) = \quad & h_1(Z_1) + h_2(Z_2) + h_3(Z_3) + h_4(Z_1^2) + h_5(Z_2^2) + h_6(Z_3^2) \\ & + h_7(Z_1Z_2) + h_8(Z_1Z_3) + h_9(Z_2Z_3) + \int_{1100}^{2498} \alpha(t)X(t)\mathrm{d}t. \end{aligned} \tag{5.6}$$

We consider the GFPAHM with the group SCAD(gSCAD) and group LASSO(gLASSO) penalty functions. We also use BFPCA and BSpline to analyze the data, where BFPCA stands for the B-spline basis and FPC basis functions are used to approximate $\{h_i(\cdot), i = 1, 2, 3\}$ and $\alpha(\cdot)$, respectively, and BSpline represents that we use the B-spline basis functions for all nonlinear components in mode (5.6). We first partition the data randomly into a training set and a test set with the ratio $4 : 1$. Next, means, medians and standard deviations (SD) of MSEP under four different estimation methods over 500 replications are given in Table 3, where the MSEP represents the mean square error of prediction. We can see that BFPCA-gSCAD has smaller MSEP than the other three methods.

Table 3: Means, medians and standard deviations (SD) of MSEP under different estimation methods

Method	MSEP		
	Mean	Median	SD
BFPCA-gSCAD	0.0214	0.0044	0.0489
BFPCA-gLASSO	0.0408	0.0038	0.0880
BSpline-gSCAD	0.0231	0.0048	0.0518
BSpline-gLASSO	0.0248	0.0032	0.0640

At the same time, the frequency of a single variable being selected in 500 replications is shown in Figure 3. It shows that the percentage of water*sucrose, water*dry flour, sucroseˆ2, dry flour*sucrose and water selected in 500 replications is higher than 0.80. So the proposed variable selection procedures to the model (5.6) suggest that $h_1(Z_1)$, $h_6(Z_3^2)$, $h_7(Z_1Z_2)$, $h_8(Z_1Z_3)$ and $h_9(Z_2Z_3)$ should enter the model, whereas $h_2(Z_2)$, $h_3(Z_3)$, $h_4(Z_1^2)$ and $h_5(Z_2^2)$ are suggested not to enter. That is, according to the results of variable selection, the chemical composition of the cookie dough contains the interaction term of the water, dry flour and sucrose, quadratic terms of sucrose, and also contains water. These findings will help the researchers to build a more comprehensive model and predict the chemical composition of the sample.

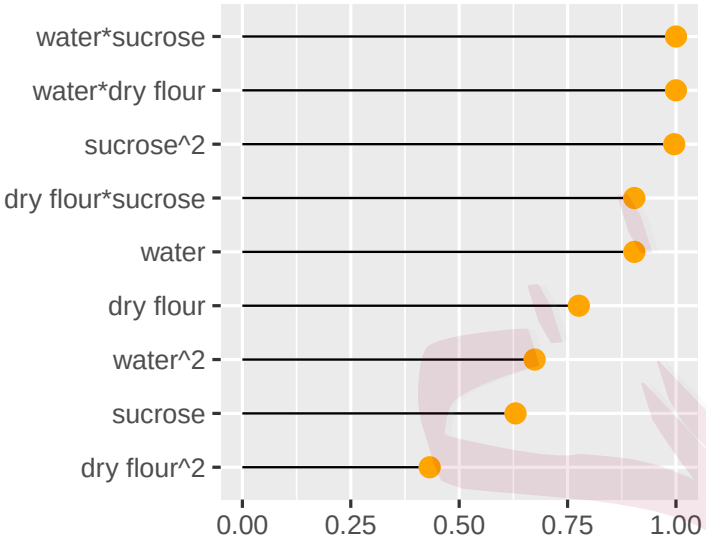

Figure 3: The proportion of each variable selected in 500 replications

6. Discussion and concluding remarks

In this paper, we investigate a generalized functional partially additive hybrid model where the explanatory variables involve both infinite dimensional predictor processes viewed as functional data with measurement errors, and high-dimensional scalar covariates whose impact on the response is nonlinear. We investigate a nonconvex penalized likelihood estimator for simultaneous variable selection and estimation. We apply the proposed estimation method to the biscuit dough data analysis. In the biscuit dough data application, we have shown that the proposed estimation method allows the researchers to build a more concise model and improve explanatory power, especially

when there are multiple scalar covariates in the data distribution, BFPCA-gSCAD method is significantly superior to other methods.

In the future research work, we can consider extending functional data to multivariate functional data, and apply Wong et al. (2018)'s idea of building an additive model for the scores of multivariate functional data to the model of this paper. At the same time, the two-step estimation method proposed by Tang et al. (2023) is worthy of investigation within the proposed framework, so that the asymptotic distribution and inference of additive functions can also be developed.

Supplementary Materials

The online supplementary material contains the technical proofs for the theorems for the proposed methodology.

Acknowledgements

The work was partially supported by the Beijing Natural Science Foundation (No.1242005), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (25XNN015), the Fundamental Research Funds for the Central Universities of China(3142023039); the Science and Technology Support Project of Langfang(2024011081, 2024011084); Key Discipline of Applied Statistics of North China Institute of Science and Technology; Natural Science Foundation of Xin-

jiang Uygur Autonomous Region(2023D01A74); the Social Science Fund of Xinjiang Uygur Autonomous Region (2023BTJ048); National Social Science Fund (22XTJ006); High-level Talent Special Project of Xinjiang University of Finance and Economics (2024XGC033).

Professor Wolfgang Härdle's work is supported through the European Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry; the project "IDA Institute of Digital Assets", CF166/15.11.2022, contract number CN760046/23.05.2024 financed under the Romania's National Recovery and Resilience Plan, Apel nr. PNRR-III-C9-2022-I8; and the Marie Sklodowska-Curie Actions under the European Union's Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance, acronym DIGITAL, Project No. 101119635.; the Yushan Fellowships, TW.

References

Aneiros, P. G. and P. Vieu (2006). Semi-functional partial linear regression. *Statistics and Probability Letters 76*, 1102–1110.

Breheny, P. and J. Huang (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing 25*, 173–187.

Brown, P. J., T. Fearn, and M. Vannucci (2001). Bayesian wavelet regression on curves with application to a spectroscopic calibration problem. *Journal of the American Statistical Association 96*, 398–408.

REFERENCES

Cao, R. Y., J. Du, J. J. Zhou, and T. F. Xie (2020). FPCA-based estimation for generalized functional partially linear models. *Statistical Papers 61*, 2715–2735.

Cardot, H., F. Ferraty, A. Mas, and P. Sarda (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics 30*, 241–255.

Cardot, H., F. Ferraty, P. Sarda, and Toulouse. (2003). Spline estimators for the functional linear model. *Statistica Sinica 13*, 571–591.

Cardot, H. and P. Sarda (2008). Varying-coefficient functional linear regression models. *Communications in Statistics-Theory and Methods 37*, 3186–3203.

Carroll, R. J.and Fan, J. Q., I. Gijbels, and M. P. Wand (1997). Generalized partially linea single-index models. *Journal of the American Statistical Association 92*, 477–489.

Cuevas, A., M. Febrero, and R. Fraiman (2002). Linear functional regression: The case of fixed design and functional response. *The Canadian Journal of Statistics 30*, 285–300.

de Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer.

Delaigle, A. and P. Hall (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics 40*, 322–352.

Ding, H., R. Q. Zhang, and J. Zhang (2018). Quantile estimation for a hybrid model of functional and varying coefficient regressions. *Journal of Statistical Planning and Inference 196*, 1–18.

Du, J., R. Y. Cao, E. Kwessi, and Z. Z. Zhang (2019). Estimation for generalized partially functional linear additive regression model. *Journal of Applied Statistics 46*, 914–925.

Du, J., D. K. Xu, and R. Y. Cao (2018). Estimation and variable selection for partially functional linear

models. *Journal of the Korean Statistical Society 47*, 436–449.

Fan, J. Q. and R. Z. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

Fan, J. Q. and R. Z. Li (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association 99*, 710–723.

Feng, S. Y. and L. G. Xue (2016). Partially functional linear varying coefficient model. *Statistics 50*, 717–732.

Guo, C. H., H. Yang, and J. Lv (2017). Robust variable selection in high-dimensional varying coefficient models based on weighted composite quantile regression. *Statistical Papers 58*, 1009–1033.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*, 70–91.

Hsing, T. and R. L. Eubank (2015). *Theoretical Foundations of Functional Data Analysis with an Introduction to Linear Operators*. Ltd: John Wiley and Sons.

Huang, J., J. L. Horowitz, and F. R. Wei (2010). Variable selection in nonparametric additive models. *Annals of Statistics 38*, 2282–2313.

Kai, B., R. Z. Li, and H. Zou (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics 39*, 305–332.

Kong, D. H., K. J. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*, 147–159.

Liu, Y. X., Z. H. Wang, M. Z. Tian, and K. Yu (2022). Estimation and variable selection for generalized functional partially varying coefficient hybrid models. *Statistical Papers*.

REFERENCES

Morris and S. Jeffrey (2015). Functional regression. *Annual Review of Statistics and Its Application 2*, 321–359.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*, 186–199.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika 47*, 379–396.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis*. New York: Springer.

Reiss, P. T. and R. T. Ogden (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association 102*, 984–996.

Severini, T. A. and J. G. Staniswalis (1994). Quasi-likelihood estimation in semiparametricmodels. *Journal of the American Statistical Association 89*, 501–511.

Sherwood, B. and L. Wang (2016). Partially linear additive quantile regression in ultra-high dimension. *Annals of Statistics 44*, 288–317.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference 139*, 3405–3418.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Annals of Statistics 14*, 590–606.

Tang, Q. G., W. Tu, and L. L. Kong (2023). Estimation for partial functional partially linear additive model. *Computational Statistics and Data Analysis 177*, 1–15.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B 58*, 267–288.

van Der Vaart, A. W. (1998). Asymptotic statistics.

REFERENCES

Wang, H. S. and Y. C. Xia (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association 104*, 747–757.

Wang, J. L., J. M. Chiou, and H. G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Wang, L., X. Liu, H. Liang, and R. J. Carroll (2011). Estimation and variable selection for generalized additive partial linear models. *Annals of Statistics 39*, 1827–1851.

Wei, F., J. Huang, and H. Z. Li (2011). Variable selection and estimation in high-dimensional varying-coefficient models. *Statistica Sinica 21*, 1515–1540.

Wong, R. K. W., Y. H. Li, and Z. Y. Zhu (2018). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association 114*, 406–418.

Wu, Y. C., J. Q. Fan, and H. G. Müller (2010). Varying-coefficient functional linear regression. *Bernoulli 16*(3), 730–758.

Yao, F., H. G. Müller, and J. L. Wang (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics 33*, 2873–2903.

Yao, F., S. Sue Chee, and F. Wang (2017). Regularized partially functional quantile regression. *Journal of Multivariate Analysis 159*, 39–56.

Zhang, D., X. Lin, and M. F. Sowers (2007). Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics 63*, 351–362.

Zhao, P. X. and L. G. Xue (2009). Variable selection for semiparametric varying coefficient partially linear models. *Statistics and Probability Letters 79*, 2148–2157.

REFERENCES

Yanxia Liu, College of Science, North China Institute of Science and Technology, Hebei, 065201, China

E-mail: liuyxbuct@126.com

Zhihao Wang, Institute of Statistics and Data Science, Xinjiang University of Finance and Economics, Urumqi, 830012, China.

E-mail: wangzhihao1109@ruc.edu.cn

Maozai Tian (The corresponding author), 1. Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China. 2. Institute of Statistics and Data Science, Xinjiang University of Finance and Economics, Urumqi, 830012, China.

E-mail: mztian@ruc.edu.cn