Title: Semiparametric Functional Regression Models with Multivariate Functional Predictors

Author: Yehua Li, University of California at Riverside

Abstract:

Motivated by an application on predicting crop yield using temperature trajectories and other scalar predictors, we consider two classes of semiparametric functional regression models, both of which are extensions of the classic functional linear models. We jointly model multiple functional predictors that are cross-correlated using multivariate functional principal component analysis (mFPCA), and use the mFPCA score as extracted features in a second stage semiparametric regression. In the proposed partially linear functional additive models (PLFAM), we predict the scalar response by both the parametric effects of the multivariate predictor and additive nonparametric effects of the mFPCA scores, and adopt the component selection and smoothing operator (COSSO) penalty to select relevant components and regularize the fitting. In the second class of semiparametric functional regression models, we also consider the interactions between the functional and multivariate predictors, where we assume the interaction depends on a nonparametric, single-index structure of the multivariate predictor to avoid the curse of dimensionality. We establish theoretic properties for both models, where we let the number of principal components diverge to infinity with the sample size. A fundamental difference between our framework and the existing high-dimensional semiparametric regression models is that the

principal component scores are estimated with errors, the magnitudes of which increase with the order of FPC. The practical performances of the proposed methods are illustrated through analysis of the motivating crop yield data.

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Partially Linear Functional Additive Models for Multivariate Functional Data

Raymond K. W. Wong, Yehua Li & Zhengyuan Zhu

To cite this article: Raymond K. W. Wong, Yehua Li & Zhengyuan Zhu (2019) Partially Linear Functional Additive Models for Multivariate Functional Data, Journal of the American Statistical Association, 114:525, 406-418, DOI: 10.1080/01621459.2017.1411268

To link to this article: https://doi.org/10.1080/01621459.2017.1411268

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 114, NO. 525, 406–418, Theory and Methods
https://doi.org/10.1080/01621459.2017.1411268

Taylor & Francis
Taylor & Francis Group

Partially Linear Functional Additive Models for Multivariate Functional Data

Raymond K. W. Wong[a], Yehua Li [b], and Zhengyuan Zhu[c]

[a]Department of Statistics, Texas A&M University, College Station, TX; [b]Department of Statistics, University of California, Riverside, CA; [c]Department of Statistics & Statistical Laboratory, Iowa State University, Ames, IA

ABSTRACT

We investigate a class of partially linear functional additive models (PLFAM) that predicts a scalar response by both parametric effects of a multivariate predictor and nonparametric effects of a multivariate functional predictor. We jointly model multiple functional predictors that are cross-correlated using multivariate functional principal component analysis (mFPCA), and model the nonparametric effects of the principal component scores as additive components in the PLFAM. To address the high-dimensional nature of functional data, we let the number of mFPCA components diverge to infinity with the sample size, and adopt the component selection and smoothing operator (COSSO) penalty to select relevant components and regularize the fitting. A fundamental difference between our framework and the existing high-dimensional additive models is that the mFPCA scores are estimated with error, and the magnitude of measurement error increases with the order of mFPCA. We establish the asymptotic convergence rate for our estimator, while allowing the number of components diverge. When the number of additive components is fixed, we also establish the asymptotic distribution for the partially linear coefficients. The practical performance of the proposed methods is illustrated via simulation studies and a crop yield prediction application. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received October 2016
Revised November 2017

KEYWORDS
Additive model; Functional data; Measurement error; Principal component analysis; Reproducing kernel Hilbert space; Spline

1. Introduction

As new technology being increasingly used in data collection and storage, many variables are continuously monitored over time and become multivariate functional data (Ramsay and Silverman 2005; Zhou, Huang, and Carroll 2008; Kowal, Matteson, and Ruppert 2017). Extracting useful information from such data for further regression analysis has become a challenging statistical problem. There has been significant amount of recent work devoted to regression models with functional predictors and the most popular model is the functional linear model (James 2002; Cardot, Ferraty, and Sarda 2003; Müller and Stadtmüller 2005; Cai and Hall 2006; Crainiceanu, Staicu, and Di 2009; Li, Wang, and Carroll 2010; Cai and Yuan 2012), where the scalar response variable is assumed to depend on an L^2 inner product of the functional predictor with an unknown coefficient function.

Functional data are infinite dimensional vectors in a functional space (Hsing and Eubank 2015). Due to the richness of information in such data, a simple linear model is often found inadequate and many researchers have investigated nonlinear functional regression models. The most widely used approach is to project functional data into a low-rank functional subspace and use the projections as predictors in a nonlinear model (James and Silverman 2005; Li and Hsing 2010a; Yao, Lei, and Wu 2015). The most popular and best understood dimension reduction tool for functional data is the functional principal component analysis (FPCA) (Yao, Müller, and Wang 2005; Hall, Müller, and Wang 2006; Li and Hsing 2010b). A recent development in nonlinear functional regression model is the functional additive model (FAM) (Müller and Yao 2008; Zhu, Yao, and Zhang 2014), where FPCA scores are used as predictors in an additive model.

Our research is motivated by a crop yield prediction application in agriculture. Agriculture is a major industry in the United States, the source of livelihood for millions of farmers and a vital contributor to global food security. Getting timely and reliable predictions on crop production is crucial for planners and policy makers to create appropriate strategies for the storage, distribution, and trade of agricultural products. The U.S. National Agricultural Statistical Service is the federal agency responsible for providing such statistics to the public, and their in-season crop yield forecast is primarily based on survey data. It is well known that weather has a significant impact on crop yield, and statistical models can be used to relate weather forecast to crop yield prediction (Cadson, Todey, and Taylor 1996; Hansen 2002; Prasad et al. 2006; Lobell and Burke 2010). Since measurements of meteorological variables, such as maximum and minimum temperatures, are typically available on a daily basis and their effects on yield vary at different growing stage of the crop, it is natural to treat them as functional predictors. Besides the functional predictors, scalar predictors, such as crop management methods, also have a great impact on yield and need to be included in the prediction model.

We propose a partially linear functional additive model (PLFAM) to predict a scalar response variable using both scalar and functional predictors. We use such a model to predict

CONTACT Yehua Li *yehuali@ucr.edu* Department of Statistics, University of California, Riverside, CA 92521.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/JASA*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

crop yield using the temperature trajectories. Such a model is of fundamental importance in plant science and agricultural economics: it advances our understanding of the relationship between weather conditions and crop yield, help to evaluate the impact of climate change on crop production and assist farmers and stake holders to better predict the future prices of agricultural commodity products and plan their actions accordingly. In many applications including our motivating data example, the functional predictors are strongly correlated to each other. To extract information more efficiently, we jointly model these predictors as a multivariate functional predictor, and perform dimension reduction using multivariate FPCA (mFPCA) (Ramsay and Silverman 2005; Chiou et al. 2014). The proposed PLFAM includes the parametric effects of the scalar predictors and additive nonparametric effects of the mFPCA scores. To automatically select significant additive components, we impose COSSO penalties (Lin and Zhang 2006) to the component functions and estimate the model in a reproducing kernel Hilbert space (RKHS) framework.

Our approach is different from that of Zhu, Yao, and Zhang (2014) in a few important perspectives. On the methodology side, we consider multiple functional predictors, extract informative signals from the functional predictors using mFPCA, and we adopt a semiparametric partially linear structure in our model to take into account the effects of scalar predictors. On the theory side, we allow the number of additive components in the model to diverge to infinity with the sample size, to acknowledge the fact that functional data have infinite number of principal components. Our theory is fundamentally different from those in the high-dimensional additive model literature, since our predictors in the additive model are estimated mFPCA scores that are contaminated with measurement errors (Carroll et al. 2006). As we show, the magnitude of measurement error gets higher for higher order principal components. In contrast, Zhu, Yao, and Zhang (2014) only allow finite number of principal components in their model. To bound the effect of measurement errors, they also impose some very restrictive conditions which, in effect, limit their estimator in a finite dimensional subspace of the Sobolev space. Our results, on the other hand, does not rely on such artificial assumptions.

The rest of the article is organized as follows. We describe the model and assumptions in Section 2 and the estimation procedure in Section 3. In Section 4, we investigate the asymptotic properties of the proposed estimator. We illustrate the proposed method with simulation studies in Section 5 and apply it to the motivating data example in Section 6. Some final remarks are collected in Section 7. Technical proofs and additional numerical results are relegated to the supplementary material.

2. Model and Assumptions

Let Y be a scalar random variable associated with a predictor $\boldsymbol{Z} \in \mathbb{R}^p$ and a multivariate functional predictor $\boldsymbol{X} = (X_1, \ldots, X_d)^\intercal$, where p and d are positive integers, and $X_j(t)$ is a stochastic process defined on the time domain $\mathcal{T}_j$ for $j = 1, \ldots, d$. For simplicity, we focus on the case $\mathcal{T}_j \equiv \mathcal{T}$, but having different domains does not affect our methodological nor theoretical developments. Let $\{\boldsymbol{z}_i, \boldsymbol{x}_i\}_{i=1}^n$ be iid copies of $\{\boldsymbol{Z}, \boldsymbol{X}\}$. Their relationship with the response $\{y_i\}_{i=1}^n$ are modeled as

$$y_i = m(\boldsymbol{z}_i, \boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where m is the regression function and ε_i are zero mean errors independent with $\{\boldsymbol{x}_i\}_{i=1}^n$ and $\{\boldsymbol{z}_i\}_{i=1}^n$. Further we assume $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2/\pi_i$, where σ_ε^2 is an unknown variance parameter and π_i's are known positive weights. In our application, the response y_i is the averaged crop yield per acre obtained from a survey, and π_i is proportional to the size of the harvest land.

2.1. Multivariate Functional Principal Component Analysis

We assume that, with probability 1, the trajectory of X_j is contained in a Hilbert space $\mathbb{X}_j$, with inner product $\langle \cdot, \cdot \rangle_{\mathbb{X}_j}$ and norm $\| \cdot \|_{\mathbb{X}_j}$. We will focus on the case that $\mathbb{X}_j$'s are L^2 functional spaces and the inner products are $\langle f, g \rangle_{\mathbb{X}_j} = \int_{\mathcal{T}} f(t)g(t)dt$ for any $f, g \in \mathbb{X}_j$. Let $\mathbb{X} = \bigoplus_{j=1}^d \mathbb{X}_j$ be the direct sum of the functional spaces, which is a bigger Hilbert space equipped with the induced inner product and norm, that is $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle_{\mathbb{X}} = \sum_{j=1}^d \langle x_{1j}, x_{2j} \rangle_{\mathbb{X}_j}$ and $\|\boldsymbol{x}_1\|_{\mathbb{X}} = \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle_{\mathbb{X}}^{1/2}$ for any $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})^\intercal \in \mathbb{X}$, $i = 1, 2$.

Define the mean function of the multivariate functional predictor as $\boldsymbol{\mu}(t) = \mathbb{E}\{\boldsymbol{X}(t)\} = \{\mu_1(t), \ldots, \mu_d(t)\}^\intercal$, where $\mu_j(t) = \mathbb{E}\{X_j(t)\}$. The cross-covariance function between X_j and $X_{j'}$ is $\mathcal{C}_{jj'}(s,t) = \mathbb{E}[\{X_j(s) - \mu_j(s)\}\{X_{j'}(t) - \mu_{j'}(t)\}]$, and the covariance of $\boldsymbol{X}$ is a $d \times d$ matrix of cross-covariance functions

$$\mathcal{C}(s,t) = \mathbb{E}[\{\boldsymbol{X}(s) - \boldsymbol{\mu}(s)\}\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\}^\intercal] = \{\mathcal{C}_{jj'}(s,t)\}_{j,j'=1}^d.$$

We assume that $\mathcal{C}$ defines a bounded, self-adjoint, positive semidefinite integral operator (Hsing and Eubank 2015). Standard operator theory warrants a spectral decomposition

$$\mathcal{C}(s,t) = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\psi}_k(s) \boldsymbol{\psi}_k^\intercal(t),$$

where $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ are the eigenvalues and $\boldsymbol{\psi}_k = (\psi_{k1}, \ldots, \psi_{kd})^\intercal \in \mathbb{X}$ are the corresponding eigenfunctions such that $\langle \boldsymbol{\psi}_k, \boldsymbol{\psi}_{k'} \rangle_{\mathbb{X}} = \int_{\mathcal{T}} \boldsymbol{\psi}_k(t)^\intercal \boldsymbol{\psi}_{k'}(t)dt = I(k = k')$. By a standard Karhunen–Loève expansion

$$\boldsymbol{X}(t) = \boldsymbol{\mu}(t) + \sum_{k=1}^{\infty} \xi_k \boldsymbol{\psi}_k(t),$$

where $\xi_k = \langle \boldsymbol{X} - \boldsymbol{\mu}, \boldsymbol{\psi}_k \rangle_{\mathbb{X}}$ are zero-mean random variables with $\mathbb{E}(\xi_k \xi_{k'}) = \lambda_k I(k = k')$. The variables ξ_k are the mFPCA scores of $\boldsymbol{X}$.

2.2. Partially Linear Functional Additive Model

Direct estimation of Model (1) suffers from the "curse-of-dimensionality" and is unpractical. Many popular alternative approaches are based on dimension reduction through FPCA, and the effects of the functional predictors are modeled through their principal component scores, including the functional linear models (FLM) and the FAM. Our PLFAM model follows a similar strategy and can be considered as a special case of Model (1) with additional structural assumptions.

We denote the sequence of mFPCA scores of $\boldsymbol{x}_i$ by $\boldsymbol{\xi}_{i,\infty} = (\xi_{i1}, \xi_{i2}, \ldots)^\intercal$. Even though in theory there are infinite number of principal components, the number of eigenfunctions estimated from the sample is at most $n - 1$, and

as shown in our theory in Section 4.1 even fewer of eigenfunctions are estimated consistently. For these practical reasons, it is a common practice to only use the low-order FPCA scores as predictors in a regression. Denote the truncated mFPCA scores as $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{is})^\intercal$, with a positive integer s. To avoid possible scale issues, we instead use the standardized version $\zeta_{ik} = \Phi(\lambda_k^{-1/2}\xi_{ik})$, where $\Phi(\cdot)$ is a continuously differentiable map from $\mathbb{R}$ to $[0, 1]$. We let $\Phi(\cdot)$ be the standard Gaussian cumulative distribution function (CDF) in all of our numerical studies. When the distribution of ξ is close to Gaussian, ζ is approximately uniform in $[0, 1]$, which is convenient for nonparametric modeling on the effect of ζ. Other continuous CDFs can also be used as $\Phi(\cdot)$, such as the logistic function. Write $\boldsymbol{\zeta}_{i,\infty} = (\zeta_{i1}, \zeta_{i2}, \ldots)$ and $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{is})$.

Assuming that all useful information in the multivariate functional predictor is contained in the first s principal components, which are related to the response in an additive form, and the covariate effect is linear, then model (1) becomes the following PLFAM

$$\begin{aligned} y_i &= m_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i = \boldsymbol{u}_i^\intercal\boldsymbol{\theta}_0 + f_0(\boldsymbol{\zeta}_i) + \varepsilon_i \\ &= \boldsymbol{u}_i^\intercal\boldsymbol{\theta}_0 + \sum_{k=1}^{s} f_{0k}(\zeta_{ik}) + \varepsilon_i, \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}_0 \in \mathbb{R}^{p+1}$ and $\boldsymbol{u}_i = (1, \boldsymbol{z}_i^\intercal)^\intercal$. Model (2) bears the FAM of Müller and Yao (2008) and Zhu, Yao, and Zhang (2014) as a special case when the functional predictor $X(t)$ is univariate ($d = 1$) and there are no scalar covariates. The partially linear structure is widely used in many popular semiparametric models because it combines the flexibility of nonparametric modeling with easy interpretation of the covariate effects (Carroll et al. 1997; Liu, Wang, and Liang 2011; Wang et al. 2014). In practice, $\boldsymbol{u}$ can include interactions, quadratic terms, and any other low-order nonlinear terms as long as their effects are interpretable and parametric. We show in Section 4 the estimated partially linear coefficient $\widehat{\boldsymbol{\theta}}$ (also referred to as the parametric component of the model) is $\sqrt{n}$-consistent and has an asymptotically normal distribution, despite the existence of nonparametric components, which converge in a slower rate. This is particularly useful if inference on the parametric effects is of primary interest in the study.

Following Lin and Zhang (2006) and Zhu, Yao, and Zhang (2014), we assume that each f_{0k} belongs to a RKHS. We refer interested readers to Wahba (1990) for an introduction of RKHS for penalized regression. The most widely used RKHS is the Sobolev Hilbert space. In such context, an lth order Sobolev Hilbert space $\mathbb{F}^{(l)}[0, 1]$ is the collection of functions on $[0, 1]$ whose first $(l-1)$-th derivatives are absolutely continuous and the lth derivative belongs to $L^2[0, 1]$, and the corresponding norm is chosen as

$$\begin{aligned} \|g\|^2 = &\sum_{\upsilon=0}^{l-1}\left\{\int_0^1 g^{(\upsilon)}(t)dt\right\}^2 \\ &+ \int_0^1 g^{(l)}(t)^2 dt \quad \text{for any } g \in \mathbb{F}^{(l)}[0, 1]. \end{aligned}$$

Let $\mathbb{F}_k$, $k = 1, \ldots, s$, be a sequence of lth order Sobolev spaces on $[0, 1]$ with reproducing kernels R_k, and we assume $f_{0k} \in \mathbb{F}_k$. However, the fact that constant functions belongs to each $\mathbb{F}_k$ leads to an identifiability issue. To provide an identifiable parameterization, we note that each $\mathbb{F}_k$ has an orthogonal decomposition $\mathbb{F}_k = \{1\} \oplus \bar{\mathbb{F}}_k$, where $\{1\}$ is the space of all constant functions. From now on, we assume $m_0 \in \mathbb{M} = \mathbb{I} \oplus \sum_{k=1}^{s} \bar{\mathbb{F}}_k$, where $f_{0k} \in \bar{\mathbb{F}}_k$ for $k = 1, \ldots, s$, and $\mathbb{I} = \{\boldsymbol{u}^\intercal\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^{p+1}\}$. For the rest of the article, we focus on the second-order Sobolev space with $l = 2$.

3. Estimation and Computation

3.1. Estimation in mFPCA

To start with, we assume that the trajectories of $\boldsymbol{x}_i(t)$'s are fully observed. Then, the mean and covariance of $\boldsymbol{X}$ can be estimated by

$$\begin{aligned} \widehat{\boldsymbol{\mu}}(t) &= n^{-1}\sum_{i=1}^{n} \boldsymbol{x}_i(t), \\ \widehat{\mathcal{C}}(s, t) &= n^{-1}\sum_{i=1}^{n}\{\boldsymbol{x}_i(s) - \widehat{\boldsymbol{\mu}}(s)\}\{\boldsymbol{x}_i(t) - \widehat{\boldsymbol{\mu}}(t)\}^\intercal. \end{aligned} \tag{3}$$

Since $\widehat{\mathcal{C}}$ has rank $n-1$, it has a spectral decomposition $\widehat{\mathcal{C}}(s, t) = \sum_{k=1}^{n-1}\widehat{\lambda}_k\widehat{\boldsymbol{\psi}}_k(s)\widehat{\boldsymbol{\psi}}_k^\intercal(t)$, where $\widehat{\lambda}_k$ and $\widehat{\boldsymbol{\psi}}_k(t)$ are the sample eigenvalues and eigenfunctions. The estimated mFPCA scores are

$$\begin{aligned} \widehat{\xi}_{ik} &= \langle \boldsymbol{x}_i, \widehat{\boldsymbol{\psi}}_k\rangle_{\mathbb{X}} = \sum_{j=1}^{d}\int_{\mathcal{T}} x_{ij}(t)\widehat{\psi}_{kj}(t)dt, \\ \widehat{\zeta}_{ik} &= \Phi(\widehat{\lambda}_k^{-1/2}\widehat{\xi}_{ik}), \quad k = 1, \ldots, d. \end{aligned} \tag{4}$$

In practice, we only have discrete noisy observations on $\boldsymbol{x}_i$

$$\begin{aligned} w_{ijk} = x_{ij}(t_{ijk}) + e_{ijk}, \quad i = 1, \ldots, n, \\ j = 1, \ldots, d, \quad k = 1, \ldots, N_{ij}, \end{aligned}$$

where e_{ijk}'s are independent measurement errors with mean 0 and variance $\sigma_{e,j}^2$, $j = 1, \ldots, d$. We will focus on the case where dense measurements are made on each curve such that each functional predictor can be effectively recovered by passing a linear smoother through the discrete observations. Let the recovered functions be $\widetilde{x}_{ij}(t) = \mathfrak{S}(t; \boldsymbol{t}_{ij})\mathfrak{w}_{ij}$, where $\mathfrak{w}_{ij} = (w_{ij1}, \ldots, w_{ij,N_{ij}})^\intercal$ and $\mathfrak{S}(t; \boldsymbol{t}_{ij})$ is a linear smoother depending on the design points $\boldsymbol{t}_{ij} = (t_{ij1}, \ldots, t_{ijN_{ij}})^\intercal$, for example local polynomial or regression splines. The eigenvalues, eigenfunctions and mFPCA scores are estimated by replacing $x_{ij}(t)$ with $\widetilde{x}_{ij}(t)$ in (3) and (4).

For univariate functional data, this presmoothing approach is theoretically justified by Hall, Müller, and Wang (2006), who show that, when $\mathfrak{S}$ is a local linear smoother and $N_{\min} = \min_{i,j} N_{ij} > Cn^{1/4}$, the error incurred by approximating $x_{ij}(t)$ with $\widetilde{x}_{ij}(t)$ is negligible in $\widehat{\lambda}_k$ and $\widehat{\psi}_k$; Li, Wang, and Carroll (2010) further show that this approximation error is negligible to $\widehat{\xi}_{ik}$ if $N_{\min} > Cn^{5/4}$. As commented in Li, Wang, and Carroll (2010), there are two sources of error in $\widehat{\xi}_{ik}$: the error caused by approximating x_{ij} with $\widetilde{x}_{ij}$ and the error in $\widehat{\psi}_k$. If the first type of error prevails, regression analysis using $\widehat{\xi}_{ik}$ will be inconsistent even for linear models. The second type of error, on the other hand, is diminishing to zero as $n \to \infty$. There are mFPCA methodologies for sparse multivariate functional data

(see, e.g., Chiou et al. 2014), but how to consistently estimate FAM or PLFAM when the estimated scores are contaminated with nondiminishing errors is not clear and calls for further research.

In all of our numeric studies, we smooth and register each x_{ij} on B-splines, pool spline coefficients for each component in $\boldsymbol{x}_i$ into a longer vector, then the operator $\widehat{\mathcal{C}}$ is represented as a high-dimensional matrix, and the mFPCA problem reduces to a multivariate PCA problem. For detailed algorithm, we refer the readers to Ramsay and Silverman (2005, Sec. 8.5).

3.2. Estimation of PLFAM with COSSO Penalty

Let $\widehat{\boldsymbol{\zeta}}_i = (\widehat{\zeta}_{i1}, \ldots, \widehat{\zeta}_{is})^\intercal$ be a vector of standardized mFPCA scores for $\boldsymbol{x}_i$ estimated using the procedure in Section 3.1. Since there are potentially infinite number of principal components for $\boldsymbol{X}$, we choose the truncation point s to be a large positive number and use a penalized regression method to select the relevant components.

The proposed estimator $\widehat{m}$ is the minimizer of the following penalized loss $\ell_w(m)$ with respect to $m \in \mathbb{M}$. The loss function is defined as

$$\ell_w(m) = \frac{1}{n}\sum_{i=1}^{n} \pi_i\{y_i - m(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(m), \tag{5}$$

where π_i are the survey weights defined in (1). Here, τ_n^2 is a tuning parameter and $J(m) = \sum_{i=1}^{s} \|\mathcal{P}_k m\|$ with $\mathcal{P}_k$ being the projection operator to $\bar{\mathbb{F}}_k$. The penalty $J(m)$ is first proposed in the COSSO framework (Lin and Zhang 2006) for simultaneous estimation and selection of the nonparametric functions f_{0k}'s.

Following Lin and Zhang (2006), we minimize (5) by iteratively minimizing its equivalent form

$$\frac{1}{n}\sum_{i=1}^{n} \pi_i\{y_i - m(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \kappa_0 \sum_{k=1}^{s} \phi_k^{-1}\|\mathcal{P}_k m\|^2 + \kappa \sum_{k=1}^{s} \phi_k \tag{6}$$

over $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_s)^\intercal \in [0, \infty)^s$ and $m \in \mathbb{M}$, where $\kappa_0 > 0$ is a predetermined constant and κ is a tuning parameter.

The relationship between (5) and (6) is stated in the following lemma, which is an extension of Lemma 2 in Lin and Zhang (2006) to partially linear additive model under a weighted least square loss. Its proof is omitted for brevity.

Lemma 1 (Lemma 2 of Lin and Zhang (2006)). Set $\kappa = \tau_n^4/(4\kappa_0)$. (i) If $\widehat{m}$ minimizes (5), set $\widehat{\phi}_k = \kappa_0^{1/2}\kappa^{-1/2}\|\mathcal{P}_k\widehat{m}\|$; then the pair $(\widehat{\boldsymbol{\phi}}, \widehat{m})$ minimizes (6). (ii) If $(\widehat{\boldsymbol{\phi}}, \widehat{m})$ minimizes (6), then $\widehat{m}$ minimizes (5).

By representer theorem, the minimizer $\widehat{m}(\boldsymbol{u}, \boldsymbol{\zeta})$ takes the form $\boldsymbol{u}^\intercal\boldsymbol{\theta} + \sum_{k=1}^{s}\phi_k\sum_{i=1}^{n} a_i R_k(\widehat{\zeta}_{ik}, \zeta_k)$, for $\boldsymbol{u} = (1, z_1, \ldots, z_p)^\intercal \in \mathbb{R}^{p+1}$, $(\zeta_1, \ldots, \zeta_s)^\intercal \in \mathbb{R}^s$, where $\boldsymbol{a} = (a_1, \ldots, a_n)^\intercal \in \mathbb{R}^n$ is a vector of unknown parameters. Then, minimization of (6) is equivalent to minimizing

$$\frac{1}{n}\left\|\Pi^{1/2}(\boldsymbol{y} - \boldsymbol{U}\boldsymbol{\theta} - \sum_{k=1}^{s}\phi_k\boldsymbol{R}_k\boldsymbol{a})\right\|_E^2 + \kappa_0\sum_{k=1}^{s}\phi_k\boldsymbol{a}^\intercal\boldsymbol{R}_k\boldsymbol{a} + \kappa\sum_{k=1}^{s}\phi_k, \tag{7}$$

where $\|\cdot\|_E$ represents the Euclidean norm, $\Pi = \mathrm{diag}\{\pi_1, \ldots, \pi_n\}$, $\boldsymbol{y} = (y_1, \ldots, y_n)^\intercal$, $\boldsymbol{U} = [u_{ij}]_{i=1,\ldots,n,\, j=1,\ldots,p+1}$ is a $n \times (p+1)$ design matrix, and $\boldsymbol{R}_k = [R_k(\widehat{\zeta}_{ik}, \widehat{\zeta}_{jk})]_{i,j=1,\ldots,n}$ is a $n \times n$ matrix for $k = 1, \ldots, s$. For a fixed $\boldsymbol{\phi}$, minimizing (7) with respect to $(\boldsymbol{\theta}, \boldsymbol{a})$ is similar to solving a weighted ridge regression. For fixed $\boldsymbol{\theta}$ and $\boldsymbol{a}$, let $\boldsymbol{D}$ be the $n \times s$ matrix with the kth column being $\boldsymbol{R}_k\boldsymbol{a}$, then minimization of (7) with respect to $\boldsymbol{\phi} \in [0, \infty)^s$ becomes

$$\min \frac{1}{n}\left[\boldsymbol{\phi}^\intercal\boldsymbol{D}^\intercal\Pi\boldsymbol{D}\boldsymbol{\phi} - 2\left\{\boldsymbol{D}^\intercal\Pi(\boldsymbol{y} - \boldsymbol{U}\boldsymbol{\theta}) - \frac{1}{2}n\kappa_0\boldsymbol{D}^\intercal\boldsymbol{a}\right\}^\intercal\boldsymbol{\phi}\right]$$
$$\text{subject to } \sum_{k=1}^{s}\phi_k < G \text{ and } \boldsymbol{\phi} \in [0, \infty)^s,$$

for some $G > 0$, which is a typical quadratic programming. The practical minimization of (5) is done by iterating over these two minimizations by fixing $(\boldsymbol{\theta}, \boldsymbol{a})$ and $\boldsymbol{\phi}$ in turn. The algorithm starts with solving $(\boldsymbol{\theta}, \boldsymbol{a})$ while fixing $\boldsymbol{\phi} = \mathbf{1}$. Empirically, the objective function decreases quickly in the first iteration, which was also observed in Lin and Zhang (2006) and Storlie et al. (2011). To reduce the computational cost, we limit the number of iterations, and follow a one-step update procedure similar to Lin and Zhang (2006).

As discussed by Lin and Zhang (2006) and Storlie et al. (2011), κ_0 can be fixed at any positive value. We select κ_0 that minimizes the GCV of the partial spline problem when $\boldsymbol{\phi} = \mathbf{1}$. Let $\widehat{\boldsymbol{\phi}}^{(\tau_n)} = (\widehat{\phi}_1^{(\tau_n)}, \ldots, \widehat{\phi}_n^{(\tau_n)})^\intercal$, $\widehat{\boldsymbol{a}}^{(\tau_n)}$ and $\widehat{\boldsymbol{\theta}}^{(\tau_n)}$ be the minimizer of (7) for a fixed τ_n. To select the smoothing parameter τ_n (or equivalently G), we minimize the Bayesian information criterion $n\log(\mathrm{RSS}_w(\tau_n)/n) + \mathrm{df}(\tau_n)\log(n)$, where the effective degrees of freedom $\mathrm{df}(\tau_n)$ is the trace of the smoothing matrix in the partial spline problem (7) when $\boldsymbol{\phi}$ is set to $\widehat{\boldsymbol{\phi}}^{(\tau_n)}$, and the weighted residual sum of squares is $\mathrm{RSS}_w(\tau_n) = \frac{n}{\sum_{i=1}^{n}\pi_i}\|\Pi^{1/2}(\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\theta}}^{(\tau_n)} - \sum_{k=1}^{s}\widehat{\phi}_k^{(\tau_n)}\boldsymbol{R}_k\widehat{\boldsymbol{a}}^{(\tau_n)})\|_E^2$.

4. Theoretical Results

4.1. Basic Results for mFPCA

By the theory of Dauxois, Pousse, and Romain (1982), $\|\widehat{\mathcal{C}} - \mathcal{C}\|_{\mathrm{op}} = O_p(n^{-1/2})$ where the operator norm is defined as $\|\mathcal{A}\|_{\mathrm{op}} = \sup_{\boldsymbol{x}\in\mathbb{X}}\frac{\|\mathcal{A}\boldsymbol{x}\|_{\mathbb{X}}}{\|\boldsymbol{x}\|_{\mathbb{X}}}$ for any bounded linear operator $\mathcal{A}$ on $\mathbb{X}$. To derive the asymptotic expansion for $\widehat{\xi}_{ik}$'s, we use the asymptotic expansion of $\widehat{\lambda}_k$ and $\widehat{\boldsymbol{\psi}}_k$ provided by Hsing and Eubank (2015), which is a generalization of those by Hall and Hosseini-Nasab (2006) for univariate functional data to more general Hilbert space random variables. We adopt the following assumptions:

Assumption 1 (Cai and Hall 2006).

$$C_\lambda^{-1}k^{-\alpha} \le \lambda_k \le C_\lambda k^{-\alpha},$$
$$\lambda_k - \lambda_{k+1} \ge C_\lambda^{-1}k^{-1-\alpha}, \quad k = 1, 2, \ldots. \tag{8}$$

To ensure that $\sum_{k=1}^{\infty}\lambda_k < \infty$, we assume that $\alpha > 1$.

Assumption 2. $\mathbb{E}(\|\boldsymbol{X}\|_{\mathbb{X}}^4) < \infty$ and there exists a constant $C_\xi > 0$ such that $\mathbb{E}(\xi_k^2\xi_{k'}^2) \le C_\xi\lambda_k\lambda_{k'}$ and $\mathbb{E}(\xi_k^2 - \lambda_k)^2 < C_\xi\lambda_k^2$ for all k and $k' \ne k$.

The polynomial decay rate described in Assumption 1 is a slow decay rate assumption on the eigenvalues and allows $\boldsymbol{X}(t)$ to be flexibly modeled as a multivariate L^2 process without strong constraints on the roughness of its sample path. Assumption 2 is a weak moment condition on the functional predictors and is satisfied if $\boldsymbol{X}(t)$ is a multivariate Gaussian process. Both assumptions are widely used in the FLM literature (Cai and Hall 2006; Cai and Yuan 2012; Hsing and Eubank 2015). Define $\delta_k = \frac{1}{2}\min_{k' \neq k}|\lambda_{k'} - \lambda_k|$, which is no less than $\frac{1}{2}C_\lambda^{-1}k^{-1-\alpha}$ under condition (8) and denote $\Delta = n^{1/2}(\widehat{\mathcal{C}} - \mathcal{C})$. By Dauxois, Pousse, and Romain (1982), Δ converges weakly to a Gaussian variable in the space of linear operators, and hence $\|\Delta\|_{\text{op}} = \mathcal{O}_p(1)$.

Proposition 1 (Transformed FPC scores). Suppose the transformation function $\Phi(\cdot)$ has bounded derivative. Under Assumptions 1 and 2, there is a constant $C > 0$ such that $\mathbb{E}(\widehat{\zeta}_{ik} - \zeta_{ik})^2 \le Ck^2/n$ uniformly for $k \le J_n$, where $J_n = \lfloor (2C_\lambda \|\Delta\|_{\text{op}})^{-1/(1+\alpha)} n^{1/(2+2\alpha)} \rfloor$.

The proof of Proposition 1 is given in Appendix A in the supplementary material. It implies that the estimation error of the principal component score increases as the order of the principal component gets higher. Interestingly, the estimation error is of order $\mathcal{O}_p(n^{-1/2}k)$, which does not depend on the decay rate α of the eigenvalues.

4.2. Asymptotic Theory for PLFAM

For simplicity, we assume that $\pi_i = 1$ for $i = 1, \ldots, n$. We begin by introducing several notations. We write P_n as the empirical distribution of $(\boldsymbol{Z}, \boldsymbol{\zeta})$. That is, $P_n = \sum_{i=1}^n \delta_{\boldsymbol{z}_i, \boldsymbol{\zeta}_i}/n$, where $\delta_{\boldsymbol{z},\boldsymbol{\zeta}}$ is the delta function at $(\boldsymbol{z}, \boldsymbol{\zeta})$. Moreover, we denote the distribution of $(\boldsymbol{Z}, \boldsymbol{\zeta})$ by P. We define the corresponding (squared) empirical norm and inner product as

$$\|m_1\|_n^2 = \int m_1^2 dP_n \quad \text{and}$$
$$(m_1, m_2)_n = \int m_1 m_2 dP_n, \qquad \text{for any } m_1, m_2 \in \mathbb{M}.$$

These notations are extended to measurement errors $\{\varepsilon_i\}$. For instance, $(\varepsilon, m_1)_n = \sum_{i=1}^n \varepsilon_i m_1(\boldsymbol{u}_i, \boldsymbol{\zeta}_i)/n$. Moreover, we write the Euclidean norm for vector as $\|\cdot\|_E$. To derive the asymptotic properties, we assume that the parametric component is identifiable. More specifically, $\boldsymbol{\Sigma} = \int \boldsymbol{u}\boldsymbol{u}^\intercal dP$ is nonsingular.

Theorem 1. Suppose, for some $\beta > 0$, $\mathbb{E}(\widehat{\zeta}_{ik} - \zeta_{ik})^2 \le Cn^{-1}k^{2\beta}$ uniformly for all $k \le s$. Assume $0 < J(m_0) < \infty$, $\boldsymbol{\Sigma}$ is nonsingular and $\tau_n^{-1} = \mathcal{O}_p(\min\{n^{2/5}s^{-6/5}, n^{1/2}s^{-(\frac{1}{2}+\beta)}\})$, we have $\|\widehat{m} - m_0\|_n = \mathcal{O}_p(\tau_n)$ and $J(\widehat{m}) = \mathcal{O}_p(1)$. If $J(m_0) = 0$ and $\tau_n \asymp n^{-1/4}s^3$, $\|\widehat{m} - m_0\|_n = \mathcal{O}_p(n^{-1/2})$ and $J(\widehat{m}) = \mathcal{O}_p(n^{-1/2}s^{-6})$.

Remarks:

1. Under the framework laid out in Assumptions 1 and 2, with $s = \mathcal{O}_p(n^{1/\{2(1+\alpha)\}})$, we have $\mathbb{E}(\widehat{\zeta}_{ik} - \zeta_{ik})^2 \le Cn^{-1}k^2$ uniformly for all $k \le s$ followed from Proposition 1. The results in Theorem 1 can be further simplified by identifying $\beta = 1$. In this case, if $0 < J(m_0) < \infty$ and $\tau_n^{-1} = \mathcal{O}_p(n^{2/5}s^{-6/5})$, we have $\|\widehat{m} - m_0\|_n = \mathcal{O}_p(n^{-2/5}s^{6/5})$. If s is fixed, $\|\widehat{m} - m_0\|_n = \mathcal{O}_p(n^{-2/5})$ is the optimal nonparametric convergence rate assuming each f_{0k} belongs to a second-order Sobolev space.
2. Our result can be considered as an extension of Theorem 1 in Zhu, Yao, and Zhang (2014), where we allow $s \to \infty$ in a rate no faster than $\mathcal{O}_p(n^{1/\{2(1+\alpha)\}})$. The reason for setting such a restriction on the rate of s is that, to estimate the principal components consistently, we need the distance between two adjacent eigenvalues to be no smaller than $\|\widehat{\mathcal{C}} - \mathcal{C}\|_{\text{op}}$. This is a fundamental difference with classic high-dimensional additive models (Meier, van de Geer, and Bühlmann 2009; Ravikumar et al. 2009; Liu, Wang, and Liang 2011; Wang et al. 2014).
3. The key issue in achieving consistent estimation of PLFAM is to bound the estimation error in $\widehat{\zeta}_{ik}$. To achieve this goal, Zhu, Yao, and Zhang (2014) assumed (see their Assumption 1)

$$\left|\frac{\partial f(\boldsymbol{\zeta}_i)}{\partial \zeta_{ik}}\right| = |f_k'(\zeta_{ik})| \le B_i \|f\|_2 \qquad \text{with probability 1}$$

for some independent variables $\{B_i\}_{i=1}^n$ with $\mathbb{E}(B_i^2) < \infty$, where $\|\cdot\|_2$ is the $L_2(P)$-norm. This is a strong assumption that eliminates the possibility f_k belonging to the space spanned by high-order Fourier or Demmler–Reinsch basis functions. As an effect, their estimation is restricted in a low-dimensional functional space. We, on the other hand, show in Lemma 2 that $\sup_{\zeta \in [0,1]} |f_k'(\zeta)|$ is bounded by the RKHS norm of f_k for all $k \le s$, and such a result help to control the error caused by the error-contaminated predictor $\widehat{\boldsymbol{\zeta}}_i$.

When s is fixed, better asymptotic results can be derived for the regression coefficients $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^\intercal = (\theta_2, \ldots, \theta_{p+1})^\intercal$. Define

$$\begin{aligned}
\boldsymbol{w}(\boldsymbol{\zeta}) &= (w_1(\boldsymbol{\zeta}), \ldots, w_p(\boldsymbol{\zeta}))^\intercal \\
&= \operatorname{argmin}_{\substack{w_j \in \{1\} \oplus \sum_{k=1}^s \bar{\mathbb{F}}_k \\ j = 1, \ldots, p}} \mathbb{E}\|\boldsymbol{Z} - \boldsymbol{w}(\boldsymbol{\zeta})\|^2, \\
\widetilde{\boldsymbol{w}}(\boldsymbol{z}, \boldsymbol{\zeta}) &= (\widetilde{w}_1(\boldsymbol{z}, \boldsymbol{\zeta}), \ldots, \widetilde{w}_p(\boldsymbol{z}, \boldsymbol{\zeta}))^\intercal = \boldsymbol{z} - \boldsymbol{w}(\boldsymbol{\zeta}), \\
\boldsymbol{M} &= (M_{ij})_{i,j=1}^p, \qquad \text{where } M_{ij} = \int \widetilde{w}_i \widetilde{w}_j dP. \qquad (9)
\end{aligned}$$

It is easy to see that $\boldsymbol{w}(\boldsymbol{\zeta})$ defines an additive regression of $\boldsymbol{Z}$ on $\boldsymbol{\zeta}$, and it can be considered as the projection of $\mathbb{E}(\boldsymbol{Z}|\boldsymbol{\zeta})$ on the additive regression space, and therefore

$$\mathbb{E}\{\widetilde{\boldsymbol{w}}^\intercal(\boldsymbol{z}, \boldsymbol{\zeta})\boldsymbol{g}(\boldsymbol{\zeta})\} = 0 \qquad (10)$$

for any $\boldsymbol{g}(\boldsymbol{\zeta}) = (g_1, \ldots, g_p)^\intercal(\boldsymbol{\zeta})$ such that $g_j(\boldsymbol{\zeta}) \in \{1\} \oplus \sum_{k=1}^s \bar{\mathbb{F}}_k$ for $j = 1, \ldots, p$.

Theorem 2. Assume the conditions of Theorem 1 hold with $s < \infty$ being fixed, $\boldsymbol{\zeta}$ has a nondegenerate joint density on $[0, 1]^s$, which is bounded above and below, $\tau_n = \mathcal{O}_p(n^{-1/4})$, and that $\boldsymbol{M}$ defined in (9) is nonsingular. Then, $n^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \to \text{Normal}(\boldsymbol{0}, \boldsymbol{M}^{-1}(\boldsymbol{V}_1 + \boldsymbol{V}_2)\boldsymbol{M}^{-1})$ in distribution, where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are defined in (S.14) of the supplementary material.

Remark. As shown in our proof, $\boldsymbol{V}_1 = \text{cov}\{n^{1/2}(\varepsilon, \widetilde{\boldsymbol{w}})_n\}$, and $\boldsymbol{M}^{-1}\boldsymbol{V}_1\boldsymbol{M}^{-1}$ is the typical asymptotic covariance matrix of $\widehat{\boldsymbol{\gamma}}$ in classic literature of partially linear additive model (Wang et al.

2014), where $\boldsymbol{\zeta}$ is directly observed. The covariance $\boldsymbol{V}_2$ is the extra variation, caused by the estimation error in the FPCA score $\widehat{\boldsymbol{\zeta}}$. The two sources of variation are asymptotically independent to each other because the model error ε is independent with the error in $\widehat{\boldsymbol{\zeta}}$. A similar effect of FPCA estimation error was discovered by Li, Wang, and Carroll (2010), who investigated a simpler functional linear regression model and found that the FPCA error tends to inflate the asymptotic variance of the parametric component even if the functional predictors are fully observed. Our result in Theorem 2 shows the same phenomenon also exists for nonlinear functional regression models such as the PLFAM.

5. Simulation Study

We extend the simulation setting of Zhu, Yao, and Zhang (2014) to a multivariate functional data setting with an additional vector predictor $\boldsymbol{Z}$. The multivariate functional predictor is $\boldsymbol{x}_i(t) = \{x_{i1}(t), x_{i2}(t)\}^\intercal$ with

$$x_{i1}(t) = t + \sin(t) + \sum_{k=1}^{10} \xi_{ik}^{(1)} \psi_k^{(1)}(t),$$
$$x_{i2}(t) = t + \cos(t) + \sum_{k=1}^{10} \xi_{ik}^{(2)} \psi_k^{(2)}(t),$$

where $\xi_{ik}^{(1)} \sim N(0, \varsigma_{2k-1})$, $\xi_{ik}^{(2)} \sim N(0, \varsigma_{2k})$, $\varsigma_k = 45.25k^{-2}$, $\operatorname{corr}(\xi_{ik}^{(j)}, \xi_{ik'}^{(j)}) = 0$ for $k' \neq k$, and $\psi_k^{(j)}(t) = (1/\sqrt{5})\sin(\pi kt/10)$ for $t \in \mathcal{T} = [0, 10]$, $j = 1, 2$. The equations above define the univariate Karhunen–Loève expansions for the two functional predictors, respectively, scores within the same functional predictor are independent; however, we allow the scores from different functional predictors to be cross-correlated. We let $\operatorname{corr}(\xi_{ik}^{(1)}, \xi_{ik'}^{(2)}) = \varrho$ for $k' = k$ and 0 otherwise, where ϱ is a cross-correlation parameter between 0 and 1.

The mFPCA eigenfunctions are defined through an orthogonalization of the univariate eigenfunctions, as described in Proposition 5 in Happ and Greven (2017). More specifically, suppose the covariance matrix pooling all univariate FPCA scores has the eigenvalue decomposition $\operatorname{var}(\{(\boldsymbol{\xi}_i^{(1)})^\intercal, (\boldsymbol{\xi}_i^{(2)})^\intercal\}^\intercal) = \boldsymbol{PQP}^\intercal$, where $\boldsymbol{\xi}_i^{(j)} = (\xi_{i1}^{(j)}, \ldots, \xi_{i10}^{(j)})^\intercal$, $j = 1, 2$, $\boldsymbol{Q} = \operatorname{diag}\{\lambda_1, \ldots, \lambda_{20}\}$ and $\boldsymbol{P}^\intercal\boldsymbol{P} = \boldsymbol{I}$. The kth mFPCA score $\xi_{ik} \sim N(0, \lambda_k)$ is a linear function of the univariate scores $\{(\boldsymbol{\xi}_i^{(1)})^\intercal, (\boldsymbol{\xi}_i^{(2)})^\intercal\}\boldsymbol{p}_k$, where $\boldsymbol{p}_k = \{(\boldsymbol{p}_k^{(1)})^\intercal, (\boldsymbol{p}_k^{(2)})^\intercal\}^\intercal$ is the kth column of $\boldsymbol{P}$, and the corresponding mFPCA eigenfunction is $\boldsymbol{\psi}_k(t) = (\psi_{k1}, \psi_{k2})^\intercal(t)$, where $\psi_{kj}(t) = \{\boldsymbol{\psi}^{(j)}(t)\}^\intercal \boldsymbol{p}_k^{(j)}$, $\boldsymbol{\psi}^{(j)}(t) = (\psi_1^{(j)}, \ldots, \psi_{10}^{(j)})^\intercal(t)$, $j = 1, 2$.

From model (2), we simulate 1000 iid copies of $\{Y, \boldsymbol{Z}, \boldsymbol{X}(\cdot)\}$, denoted as $\{y_i, \boldsymbol{z}_i, \boldsymbol{x}_i(\cdot)\}_{i=1}^{1000}$, with the first 200 used as training data and the rest as testing data. Observations on $\boldsymbol{x}_i$ are obtained on a regular grid of 100 points in $\mathcal{T} = [0, 10]$ with independent measurement errors following $N(0, 0.2^2)$. For the regression function, we set $f_0(\boldsymbol{\zeta}) = f_{01}(\zeta_{i1}) + f_{02}(\zeta_{i2}) + f_{04}(\zeta_{i4})$, where $f_{01}(\zeta_1) = 3\zeta_1 - 3/2$, $f_{02}(\zeta_2) = \sin\{2\pi(\zeta_2 - 1/2)\}$ and $f_{04}(\zeta_4) = 8(\zeta_4 - 1/3)^2 - 8/9$. There are only three nonzero additive component functions in our simulation: $f_{0k}(\zeta_k) = 0$ for $k \notin \{1, 2, 4\}$. Moreover, we generate the vector predictor $\boldsymbol{z}_i$ independently from the bivariate uniform distribution over $[0, 1]^2$. We consider two settings for the partially linear coefficient $\boldsymbol{\theta}_0$: (I) $(1.4, 0, 0)^\intercal$ and (II) $(1.4, 3, -4)^\intercal$ and two settings of the correlation parameter ϱ: (i) 0.3 (low correlation) and (ii) 0.9 (high correlation). Combining different setups for ϱ and $\boldsymbol{\theta}_0$, we have four settings: {(i), (I)}, {(i), (II)}, {(ii), (I)}, and {(ii), (II)}. The errors ε_i's in the regression model (2) are distributed independently as $N(0, \sigma^2)$ with σ^2 being 1 for setting (I) and 1.9470 for (II) to achieve the signal-to-noise ratio (SNR) of approximately 2.2. The SNR is defined as $\operatorname{var}(m_0(\boldsymbol{\zeta}))/\operatorname{var}(\varepsilon)$. For simplicity, all sampling weights π_i are set to be 1. The simulation is repeated 200 times and we fit the following two models to each simulated dataset: FAM of Zhu, Yao, and Zhang (2014), which is also based on COSSO but ignores the effect of $\boldsymbol{Z}$, and the proposed PLFAM. Throughout this simulation study, s is chosen to recover at least 99.9% of the total variation in $\{\boldsymbol{x}_i\}$ and the COSSO tuning parameters are selected by the Bayesian information criterion.

Tables 1 and 2 summarize the results related to component function selection in FAM and PLFAM under the four settings. Due to space constraint, only percentages of model sizes up to 8 and selection percentages of the first 8 component functions are shown. In Table 2, Column "% correct set" corresponds to the percentages of fittings achieving exact selection of $\widehat{f}_1$, $\widehat{f}_2$, and $\widehat{f}_4$, while Column "% super set" gives the percentages of fittings that include nonzero $\widehat{f}_1$, $\widehat{f}_2$, and $\widehat{f}_4$. Despite a small tendency of over-selection, the COSSO component selection mechanism tends to select parsimonious models and, for each correct component function, the selection percentage is high.

To assess the estimation quality of f_{0k}'s, Table 3 shows the averaged integrated squared errors (AISEs) of the first eight component functions and the overall function $\widehat{f} = \sum_{k=1}^{s} \widehat{f}_k$

Table 1. Percentages of fitted model sizes.

		% for the following model sizes							
Setting	Model	1	2	3	4	5	6	7	8
{(i), (I)}	FAM	0	0	28	49.5	20.5	1.5	0.5	0
	PLFAM	0	0	24	57.5	17	1	0.5	0
{(ii), (I)}	FAM	0	0	20.5	58	16.5	4	1	0
	PLFAM	0	0	19	58.5	18	3.5	0	1
{(i), (II)}	FAM	0	6.5	41	39.5	12	0.5	0.5	0
	PLFAM	0	0	22.5	56	18	3	0.5	0
{(ii), (II)}	FAM	0	3	44.5	38	12.5	2	0	0
	PLFAM	0	0	22.5	61	12.5	3	1	0

Table 2. Percentages of selected components and, correct and super selection.

Setting	Model	% for the following component functions								% correct set	% super set
		$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_3$	$\widehat{f}_4$	$\widehat{f}_5$	$\widehat{f}_6$	$\widehat{f}_7$	$\widehat{f}_8$		
{(i), (I)}	FAM	100	100	14	93	51.5	2	6	1.5	27	93
	PLFAM	100	100	14	93	51.5	3	5	1.5	23	93
{(ii), (I)}	FAM	100	100	20.5	97	51	5.5	1.5	2	18.5	97
	PLFAM	100	100	20	97.5	53	5.5	1.5	2.5	17.5	97.5
{(i), (II)}	FAM	100	90	8.5	93.5	34.5	2.5	4.5	4.5	35	83.5
	PLFAM	100	100	16	97.5	54.5	3.5	3.5	1.5	21.5	97.5
{(ii), (II)}	FAM	100	93.5	12	94	31.5	2.5	3	1	37.5	88
	PLFAM	100	99.5	22.5	98	47.5	2.5	2.5	1.5	22.5	97.5

Table 3. Averaged integrated squared errors.

Setting	Model	AISEs for the following component functions								
		$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_3$	$\widehat{f}_4$	$\widehat{f}_5$	$\widehat{f}_6$	$\widehat{f}_7$	$\widehat{f}_8$	$\widehat{f}$
{(i), (I)}	FAM	0.0172	0.1073	0.0057	0.1689	0.1204	0.0001	0.0015	0.0001	0.4292
	PLFAM	0.0175	0.1070	0.0056	0.1689	0.1205	0.0004	0.0014	0.0002	0.4289
{(ii), (I)}	FAM	0.0198	0.1038	0.0109	0.1290	0.0890	0.0018	0.0004	0.0007	0.3633
	PLFAM	0.0198	0.1046	0.0111	0.1279	0.0896	0.0016	0.0005	0.0011	0.3638
{(i), (II)}	FAM	0.0330	0.2208	0.0064	0.2197	0.0782	0.0008	0.0026	0.0021	0.5780
	PLFAM	0.0177	0.1072	0.0064	0.1320	0.1130	0.0011	0.0009	0.0005	0.3858
{(ii), (II)}	FAM	0.0290	0.2035	0.0087	0.2170	0.0841	0.0017	0.0024	0.0005	0.5642
	PLFAM	0.0179	0.1084	0.0103	0.1398	0.0978	0.0007	0.0008	0.0005	0.3821

(without constant term). The integrated squared errors are defined as

$$\mathrm{ISE}(\widehat{f}_k) = \int_0^1 \{\widehat{f}_k(t) - f_{0k}(t)\}^2 dt \quad \text{and}$$

$$\mathrm{ISE}(\widehat{f}) = \sum_{k=1}^{s} \int_0^1 \{\widehat{f}_k(t) - f_{0k}(t)\}^2 dt.$$

Notice that, under setting (I) where $\boldsymbol{Z}$ has zero effect and FAM is the correct model, the PLFAM estimators perform comparably to those of FAM. However, under setting (II), where $\boldsymbol{Z}$ has nonzero effects, FAM performs significantly worse than PLFAM. This demonstrates the possible risk of ignoring important vector predictors.

We also summarize the prediction errors, and the mean squared errors (MSE) for the estimated partially linear coefficients in Table 4. To show the advantage of mFPCA, we further compare two methods to obtain FPCA scores: the "joint" approach is the mFPCA approach that we advocate; and the "separate" approach is to perform univariate FPCA to each component of $\boldsymbol{X}$, standardize these scores separately, and then pool all standardized scores together as covariates in the additive model. Both FPCA approaches can be used in conjunction with FAM and PLFAM. The prediction error is computed by $n^{-1}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ on the testing dataset. To compute the prediction $\widehat{y}_i$ in the test data, we first compute the transformed FPCA scores of $\boldsymbol{x}_i$ in the test set using the estimates of mean function, eigenvalues and eigenfunctions from the training data, and then plug these scores into the estimated regression $\widehat{m}$. The results in Table 4 suggest that jointly modeling multiple functional predictors leads to smaller MSEs for $\widehat{\boldsymbol{\theta}}$, and lower prediction errors, as opposed to modeling each functional predictor separately using univariate FPCA. In addition, PLFAM has significant lower prediction errors than FAM under setting (II) when there is a nonzero effect from $\boldsymbol{Z}$.

In Section C of the supplementary material, we also report the simulation results when s chosen to recover 90% of the total variation. Under this setting, one important component related to Y is close to the 90% cut-off line and often not included as a candidate for COSSO. As a result, a nonzero component function is often failed to be selected, and the resulted models yield higher prediction errors in the test datasets. Based on these

Table 4. Prediction errors and mean squared errors for FAM and PLFAM, using separate univariate FPCA scores (columns labeled "separate") or mFPCA scores (columns labeled "joint"). For prediction errors, means are presented with corresponding standard deviations in parentheses.

Setting	Model	Prediction error		Mean squared errors					
				Separate			Joint		
		Separate	Joint	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$
{(i), (I)}	FAM	1.55 (0.10)	1.32 (0.13)	—	—	—	—	—	—
	PLFAM	1.57 (0.11)	1.33 (0.13)	0.0746	0.0911	0.1076	0.06	0.0751	0.0831
{(ii), (I)}	FAM	1.65 (0.09)	1.33 (0.12)	—	—	—	—	—	—
	PLFAM	1.66 (0.09)	1.35 (0.13)	0.0678	0.1095	0.0827	0.0585	0.0888	0.0681
{(i), (II)}	FAM	3.84 (0.22)	3.63 (0.21)	—	—	—	—	—	—
	PLFAM	1.59 (0.10)	1.34 (0.13)	0.0639	0.1023	0.0894	0.0545	0.0935	0.0696
{(ii), (II)}	FAM	3.89 (0.24)	3.60 (0.24)	—	—	—	—	—	—
	PLFAM	1.68 (0.12)	1.35 (0.14)	0.0642	0.0879	0.1092	0.0526	0.069	0.0851

results, we recommend to include a large number of components and let the built-in model selection mechanism of COSSO determine the size of the model.

6. Real Data Application

The practical utility of the proposed method is illustrated through an analysis of a crop yield dataset from the National Agricultural Statistics Agency (*https://quickstats.nass.usda.gov/*), which consists of several yield-related variables at the county level (such as annual crop yield in bushels per acre, size of harvested land and the proportion irrigated land to the total harvested land) from 105 counties in Kansas from 1999 to 2011. We have yield-related variables for the two major crops in Kansas, corn and soybean, which are analyzed separately. Variables such as total harvest land and proportion of irrigated land are crop-specific. The weather data (annual averaged precipitation, daily maximum temperature, and daily minimum temperature) are gathered from 1123 weather stations in Kansas provided by the National Climatic Data Center (*https://www.ncdc.noaa.gov/data-access*) and aggregated at the county level.

To apply our model, let Y be the average crop yield per acre (corn or soybean) for a specific year and county; $X_1(t)$ and $X_2(t)$ are the daily maximum and daily minimum temperature trajectories for the same year and county with the time domain $\mathcal{T} = [0, 365]$; $\boldsymbol{Z}$ includes proportion of irrigated land in that county and for that particular type of crop, averaged annul precipitation, and the interaction between the two. In the past several decades, due to sustained improvements in genetics and production technology, there is a consistent increasing trend in the yields of both corn and soybean. To take this effect into consideration, we also include year indicators into $\boldsymbol{Z}$.

Since the response is an average obtained from an agricultural survey, the errors are heteroscedastic with weights π equal to the sizes of harvested land. Some earlier work (Smith 1938; Beran et al. 2013) suggests crop yield may exhibit long range dependency on a scale measured in feet. Our study on the other hand is based on county level aggregated data. The crop yields are usually averaged over tens of thousands of acres within a county and not from a continuous piece of land. At this scale, the spatial correlation is already quite weak, and therefore it is reasonable to assume the variance of the average crop yield is proportional to the inverse of the total harvest land. Furthermore, land use rotates between the major crops across years: land used to grow corn this year is usually used to grow soybean the next year. Variables such as the proportion of irrigated land and size of harvest land are different in different years even for the same crop and same county. Even though our theory and methods are developed under the independence assumption, they can still be applied as long as the crop yields are conditionally independent across counties and years, given the local meteorology information, which seems reasonable because of the rotation in land use and because crops of different genotypes are planted in different years.

To illustrate the functional predictors, we show in Figure S.1 of the supplementary material 50 randomly selected trajectories for $X_1(t)$ and $X_2(t)$, with the mean functions $\mu_1(t)$ and $\mu_2(t)$ marked as solid curves in the two panels. As one can see, there are a lot of local fluctuations in the temperature trajectories, which is normal since heat and chill alternate throughout the year. In Figure 1, we show the heat plots for the (cross-)covariance functions. The kernel function for $\mathcal{C}_{12}$ shows great resemblance to $\mathcal{C}_{11}$ and $\mathcal{C}_{22}$, which implies that the two functional predictors are strongly correlated. This also suggests mFPCA would achieve more efficient dimension reduction than univariate FPCA done separately to the two processes, and the latter would include too much redundant information into the regression model.

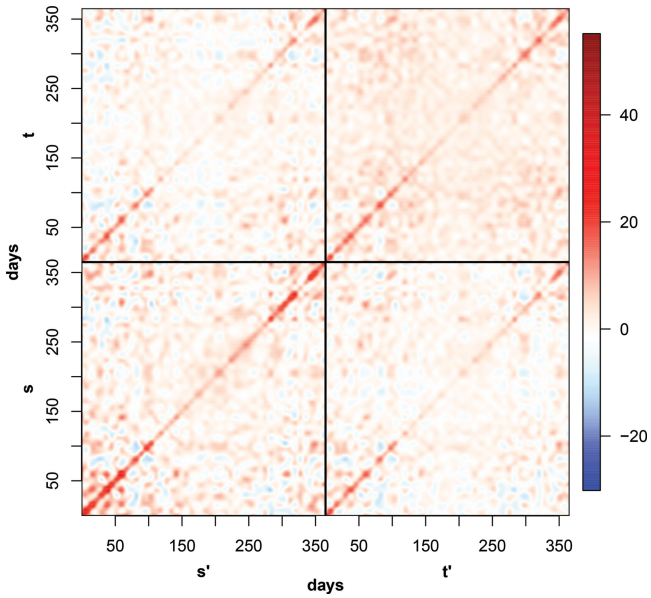


Figure 1. Heat plot for the covariance and cross-covariance functions. From bottom to top and from left to right are the kernel functions of the (cross-) covariance operators $\mathcal{C}_{jj'}$.

6.1. Crop Yield Prediction Experiment

Since our goal of this study is to find the best model for yield prediction, we divide the data into smaller training and validation datasets and compare the prediction of the following 10 competing models.

1. PLFAM(joint): the proposed PLFAM based on mFPCA scores;
2. PLFAM(separate): PLFAM based on univariate FPCA scores from X_1 and X_2 separately;
3. FAM(joint): FAM based on mFPC scores (without $\boldsymbol{Z}$);
4. FAM(separate): FAM based on univariate FPC scores (without $\boldsymbol{Z}$);
5. FLM-Cov(joint): FLM based on mFPCA scores, with covariates;
6. FLM-Cov(separate): FLM based on separate univariate FPCA scores, with covariates;
7. FLM(joint): FLM based on joint mFPCA scores (without $\boldsymbol{Z}$);
8. FLM(separate): FLM based on separate FPCA scores (without $\boldsymbol{Z}$);
9. LM: linear model on $\boldsymbol{Z}$ only;
10. LM-GDD: linear model on $\boldsymbol{Z}$ and Growing Degree Days (GDD), to be explained below.

The models we consider can be divided into three categories: (a) FAM (Models 1–4), (b) FLM (Models 5–8), and (c) nonfunctional model (Models 9–10). For all functional regression models, including those in categories (a) and (b), FPCA scores that account up to 99.9% of the total variation are admitted into the model. For the methods based on separate FPCA on X_1 and X_2, we include FPCA scores that explain 99.9% of total variation in each functional predictor, and thus use twice as many FPCA scores in the regression analysis as the joint modeling methods. For all models in category (a), we rely on the model selection mechanism of COSSO to prevent overfitting and select the tuning parameters by 5-fold cross-validation; for the FLM in category (b), we avoid overfitting by introducing ridge penalties, the tuning parameters of which are chosen by generalized cross-validation. It is worth noting that Model 10 serves as the benchmark model for yield prediction with temperature information enters into the model as the GDD variable. GDD is a measure of heat accumulation commonly used to predict plant development (Gilmore and Rogers 1958; Yang, Logan, and Coffey 1995; McMaster and Wilhelm 1997). Here we adopt the definition used in the EPIC (Erosion Productivity Impact Calculator) plant growth model (Williams et al. 1989), in which GDD is defined as the sum of $[\{X_1(t) + X_2(t)\}/2 - T_{\text{base}}]_+$ over growing season, where T_{base} is the crop-specific base temperature in $^\circ$C. For corn $T_{\text{base}} = 8$, and for soybean $T_{\text{base}} = 10$.

Table 5. Average of five-year overall prediction errors.

		Corn	Soybean
(a) Functional additive models	PLFAM(joint)	298.43	35.64
	PLFAM(separate)	306.50	38.85
	FAM(joint)	830.17	48.54
	FAM(separate)	839.00	51.06
(b) Functional linear models	FLM-Cov(joint)	303.81	35.29
	FLM-Cov(separate)	308.57	35.69
	FLM(joint)	704.19	47.31
	FLM(separate)	767.42	50.42
(c) Nonfunctional model	LM	391.18	61.74
	LM-GDD	389.76	49.58

To account for heteroscedasticity, the sizes of harvested land are used as weights in fitting all models.

For each five-year window (i.e., 1999–2003, 2000–2004, . . . , 2007–2011), we pull the data from those five years into a smaller dataset. For each five-year dataset, we randomly divide it into five subsets, hold out one subset at a time as a validation set, fit the ten models described above to the remaining four subsets, and then use the trained models to predict the responses in the validation data. The mean squared prediction errors are weighted by the sizes of harvested land, averaged over the five validation sets and over all five-year periods. The averaged overall prediction errors are reported in Table 5. From the table,

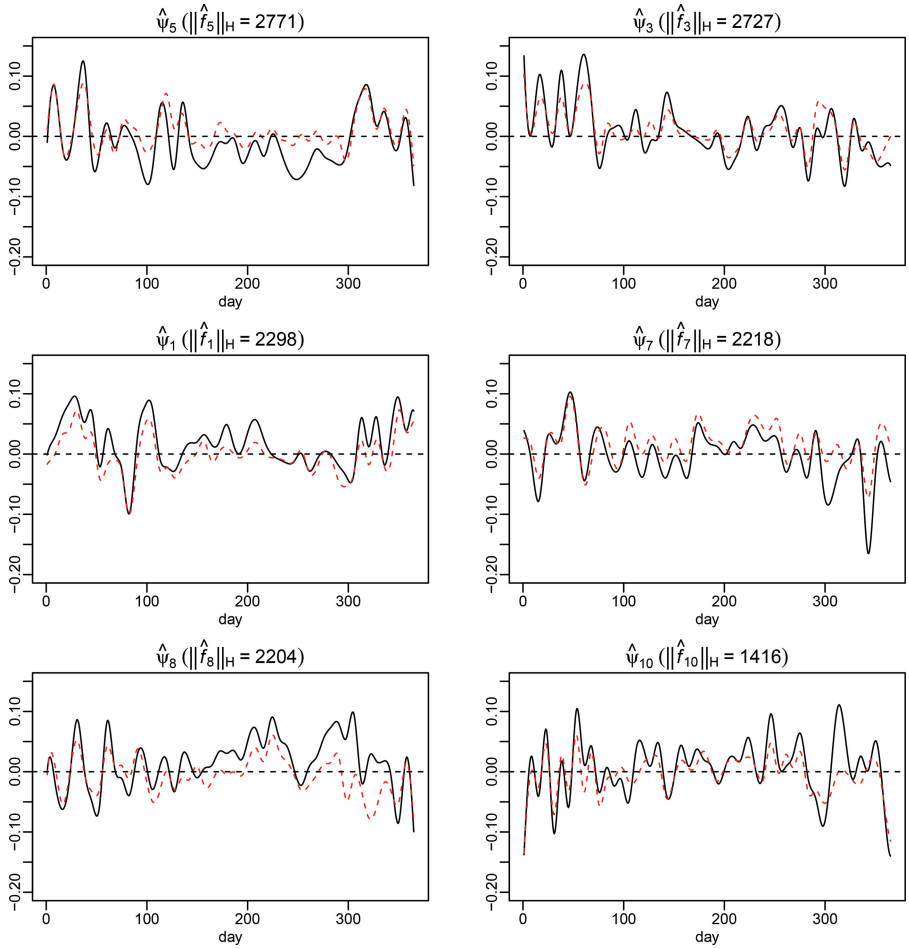


Figure 2. Corn yield prediction: top six principal components selected by COSSO for corn yield prediction, sorted by the decreasing order of the RKHS norm of $\widehat{f}_k(\zeta)$ ($k = 5, 3, 1, 7, 8, 10$). Each principal component is a vector $\boldsymbol{\psi}_k(t) = \{\psi_{k1}(t), \psi_{k2}(t)\}^{\mathsf{T}}$. The solid curve in each panel is $\psi_{k1}(t)$ and the dashed curve is $\psi_{k2}(t)$.

models without the covariate effects, including FAM(separate), FAM(joint), FLM(separate), and FLM(joint), perform significantly worse than the rest. These results agree with the general belief that irrigation and precipitation are informative in yield prediction, which also stress the importance of extending the FAM of Müller and Yao (2008) to our PLFAM. We can also see that including the functional predictors can reduce the prediction error, and functional regression model such as PLFAM(separate), PLFAM(joint), FLM-Cov(separate), and FLM-Cov(joint) perform better than the nonfunctional models (LM and LM-GDD). Joint modeling the two functional predictors using mFPCA also leads to lower prediction error for both PLFAM and FLM-Cov. Overall, PLFAM(joint) performs the best in corn yield prediction and achieves comparable result to FLM-Cov(joint) for soybean.

Part of the reason that PLFAM performs slightly worse than FLM in soybean yield prediction is that the nonlinear effect is less significant for soybean and PLFAM requires a larger sample size. In another experiment where we include more years of data in the training set, PLFAM predicts soybean yield better than FLM.

In addition to the 10 models described above, we also consider another 12 models that use X_1, X_2 or $(X_1 + X_2)/2$ alone. These models yield higher prediction errors than the proposed PLFAM(joint) model, which utilizes both functional predictors. Even though the two functional covariates in the real data are strongly correlated as suggested by Figure 1, these results show that each covariate does provide additional information that complements the other and it is beneficial to jointly model them. Due to space limitation, these results are presented in Section D of the supplementary material.

6.2. Regression Analysis of the Whole Data

We now apply PLFAM(joint) to the whole dataset pooling all available years. For corn yield prediction, we include 52 principal components in the regression model which account for ∼ 99% of variation in the temperature trajectories, and 10 principal components are selected by COSSO. In Figure 2, we show the top six most significant principal components; and in Figure 3, we show the corresponding additive component functions $\widehat{f}_k(\zeta)$. These components are ranked by the importance of their contribution to Y. More specifically, we sort the principal components by the RKHS norm of the component function $\widehat{f}_k$. The dashed curves in Figure 3 are the pointwise confidence bands $\widehat{f}_k(\zeta) \pm 2 \times se\{\widehat{f}_k(\zeta)\}$, and the dotted curves are the three times standard error bands. The standard errors are estimated using a bootstrap procedure detailed in the supplementary material.

Since each principal component in mFPCA is a vector of functions $\boldsymbol{\psi}_k(t) = \{\psi_{k1}(t), \psi_{k2}(t)\}^\intercal$, we show $\psi_{k1}(t)$ as the solid curve and $\psi_{k2}(t)$ as the dashed curve in each panel of Figure 2. It is not surprising that $\psi_{k2}(t)$ largely coincides with $\psi_{k1}(t)$, given the observation from the covariance functions that the two processes are strongly correlated. However, the plots do

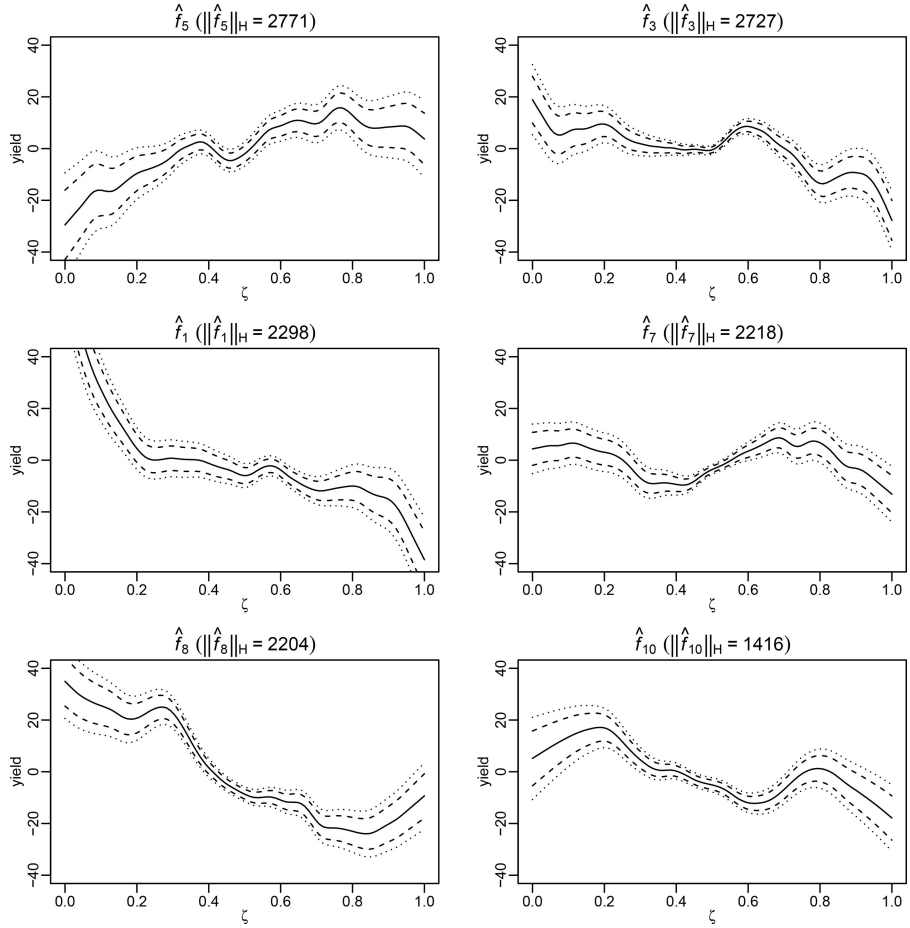


Figure 3. Corn yield prediction: top six additive component functions $\widehat{f}_k(\zeta)$, sorted by the decreasing order of the RKHS norm of $\widehat{f}_k(\zeta)$ ($k = 5, 3, 1, 7, 8, 10$).

Table 6. Estimated regression coefficients (bootstrap standard error) in the PLFAM for crop yield prediction.

	Irrigate	Prec	Irrigat*Prec
Corn	168.38 (6.42)	20.98 (2.72)	− 33.87 (3.32)
Soybean	33.30 (3.35)	3.91 (0.70)	− 4.88 (1.65)

Note: Irrigate: proportion of irrigated land in a county for the specific crop and growing year; Prec: averaged precipitation for county and year; Irrigat*Prec: the interaction.

reveal subtle differences between the two temperature trajectories. The component most related to corn yield ψ_5 features a temperature pattern with near average daily minimum temperature and lower than average daily maximum temperature during the summer months from May to September. A higher loading on ψ_5 means a milder summer, less heat stress and less chance of draught, and corn yield is an increasing function of ζ_5 in Figure 3. In contrast, ψ_1 and ψ_8 represent hot summers, and crop yield is a decreasing function of their loadings ζ_1 and ζ_8. These are consistent with the findings in Westcott and Jewison (2013), which conclude that hot July–August weather lowered the corn yield. A prominent feature in ψ_3 is warm spring months from January to March, which may lead to less snow coverage on the ground, early insect activities, and hence the negative association with corn yield as described by $f_3(\zeta)$ in Figure 3. For soybean yield prediction, graphs of the selected eigenfunctions and the corresponding additive component functions are similar to those in Figures 2 and 3, and are hence omitted.

The estimated partially linear coefficients and their bootstrap standard errors for both corn and soybean yield models are summarized in Table 6. As we can see, both the proportion of irrigated land (Irrigate) and precipitation (Prec) have significant positive effects on crop yield. The significant negative interaction means the effect of Prec is mitigated when a big portion of the lands in the county are equipped with irrigation systems. For corn yield prediction, the first and third quartiles for Irrigate are 0.027 and 0.485, respectively. Changing Irrigate from its first quartile to the third, the partial slope on Prec reduces from 167.47 to 151.95. The bootstrap procedure, provided in the supplementary material, is based on the assumption that the errors in model (2) are independent. To validate this assumption, we also estimate the spatial variogram for each year and temporal autocorrelation for each county based on the residuals of the fitted model, see Figures S.2 and S.3 in the supplementary material. The variograms and ACFs are contained in their confidence bands based on the assumption of no dependency, which means there is no significant evidence for spatial or temporal correlation.

7. Concluding Remarks

7.1. Our Contributions

We have extended the FAM of Müller and Yao (2008) to a class of PLFAM which takes into account of the effects of a multivariate covariate $\boldsymbol{Z}$. As demonstrated in our crop yield application, including the covariate effects significantly improves the prediction accuracy. The effect of functional predictors are modeled through an additive model on the principal component scores. Since the FPC scores are estimated with error, our theory and methods also shine a new light on the area of additive models with covariate measurement errors.

We have also made a number of important theoretical contributions. First, we develop a more general model framework, which includes multivariate functional predictors and multivariate covariates. Second, we allow the number of principal components admitted in the additive model to diverge to infinity, which is fundamentally different from Zhu, Yao, and Zhang (2014). Third, we are able to quantify and bound the nuisance from the estimation errors in mFPCA scores without the artificial assumption in Zhu, Yao, and Zhang (2014). Finally, when the number of principal components does not diverge to infinity, we establish root-n consistency and asymptotic normal distribution for the partially linear regression coefficients.

7.2. Interpretability of the Model

Functional regression models based on principal components are in general hard to interpret because FPCs are the maximum modes of variation in the functional predictors that are not necessarily the features most related to the response variable. This is part of the reason that many authors focused on prediction using FLM (Cai and Hall 2006; Cai and Yuan 2012). Our proposed PLFAM adopts the philosophy of semiparametric statistics: we model the effects of functional covariates nonparametrically to increase the model flexibility and prediction performance, and model the effects of the multivariate covariates parametrically for better interpretations and statistical inference. Our Theorem 2 provides a basis for statistical inference on the parametric component $\boldsymbol{\gamma}$. There is also another class of functional additive regression models proposed by Müller, Wu, and Yao (2013), McLean et al. (2014), and Kim et al. (2018), which offer an alternative view on modeling nonlinear effects of functional covariates.

7.3. mFPCA Versus Separate FPCA

For multivariate functional data, mFPCA usually provides more efficient dimension reduction than separate FPCA to each functional covariate. However, mFPCA estimates are subject to higher variability due to the need of estimating all cross-covariance functions and performing eigenvalue decomposition on a much larger covariance matrix. When the sample size is small, the extra variation in mFPCA can offset its benefit. There are also other situations where separate FPCA is more preferable, such as when different functional covariates are of different scales or even defined on different domains (Happ and Greven 2017). Under these situations, our theory and methods can also be easily extended to the model based on separate FPCA scores. A separate FPCA version of model (2) is

$$y_i = \boldsymbol{u}_i^{\mathsf{T}}\boldsymbol{\theta}_0 + \sum_{k=1}^{s}\sum_{j=1}^{d} f_{0jk}(\zeta_{ijk}) + \varepsilon_i, \tag{11}$$

where ζ_{ijk} is the kth standardized principal component score for x_{ij}. The model can be fitted using the same COSSO algorithm described in Section 3.2 except that the mFPCA scores are replaced by the separate FPCA scores. As long as the separate FPC scores can be estimated with a similar accuracy as

assumed in Theorem 1, that is $\mathbb{E}(\widehat{\zeta}_{ijk} - \zeta_{ijk})^2 \leq Cn^{-1}k^{2\beta}$ uniformly for all $j = 1, \ldots, d$ and $k \leq s$, the same asymptotic results in Theorems 1 and 2 hold for the model in (11).

Supplemental Material

The online supplementary material contains technical proofs for the theorems, additional simulation and data analysis results, a bootstrap procedure for statistical inference, as well as R codes for the proposed methodology.

Acknowledgments

Wong's research was partially supported by the National Science Foundation under Grants DMS-1612985 and DMS-1711952 (subcontract). Li's research was partially supported by the National Science Foundation grant DMS-1317118. Zhu's research was funded in part by cooperative agreement 68-7482-17-009 between the USDA Natural Resources Conservation Service and Iowa State University.

ORCID

Yehua Li http://orcid.org/0000-0002-5945-378X

References

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013), *Long-Memory Processes*, New York: Springer. [413]

Cadson, R., Todey, D. P., and Taylor, S. E. (1996), "Midwestern Corn Yield and Weather in Relation to Extremes of the Southern Oscillation," *Journal of Production Agriculture*, 9, 347–352. [406]

Cai, T. T., and Hall, P. (2006), "Prediction in Functional Linear Regression," *The Annals of Statistics*, 34, 2159–2179. [406,409,416]

Cai, T. T., and Yuan, M. (2012), "Minimax and Adaptive Prediction for Functional Linear Regression," *Journal of the American Statistical Association*, 107, 1201–1216. [406,410,416]

Cardot, H., Ferraty, F., and Sarda, P. (2003), "Spline Estimators for the Functional Linear Model," *Statistica Sinica*, 13, 571–591. [406]

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 92, 477–489. [408]

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, Boca Raton, FL: Chapman and Hall/CRC. [407]

Chiou, J.-M., Chen,Y.-T., and Yang, Y.-F. (2014), "Multivariate Functional Principal Component Analysis: A Normalization Approach," *Statistica Sinica*, 24, 1571–1596. [407,409]

Crainiceanu, C. M., Staicu, A.-M., and Di, C.-Z. (2009), "Generalized Multilevel Functional Regression," *Journal of the American Statistical Association*, 104, 155–1561. [406]

Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference," *Journal of Multivariate Analysis*, 12, 136–154. [409]

Gilmore, E., and Rogers, J. (1958), "Heat Units as a Method of Measuring Maturity in Corn," *Agronomy Journal*, 50, 611–615. [414]

Hall, P., and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society*, Series B, 68, 109–126. [409]

Hall, P., Müller, H. G., and Wang, J. L. (2006), "Properties of Principal Component Methods for Functional and Longitudinal Data Analysis," *Annals of Statistics*, 34, 1493–1517. [406,408]

Hansen, J. W. (2002), "Realizing the Potential Benefits of Climate Prediction to Agriculture: Issues, Approaches, Challenges," *Agricultural Systems*, 74, 309–330. [406]

Happ, C., and Greven, S. (2017), "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains," *Journal of American Statistical Association*, to appear. [411,416]

Hsing, T., and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators*, Chichester, UK: Wiley. [406,407,409]

James, G. (2002), "Generalized Linear Models With Functional Predictor Variables," *Journal of the Royal Statistical Society*, Series B, 64, 411–432. [406]

James, G., and Silverman, B. W. (2005), "Functional Adaptive Model Estimation," *Journal of the American Statistical Association*, 100, 565–576. [406]

Kim, J., Staicu, A.-M., Maity, A., Carroll, R. J., and Ruppert, D. (2018), "Additive Function-on-Function Regression," *Journal of Computational and Graphical Statistics*, 27, 234–244. [416]

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017), "A Bayesian Multivariate Functional Dynamic Linear Model," *Journal of the American Statistical Association*, 112, 733–744. [406]

Li, Y., and Hsing, T. (2010a), "Deciding the Dimension of Effective Dimension Reduction Space for Functional and High-Dimensional Data," *Annals of Statistics*, 38, 3028–3062. [406]

Li, Y., and Hsing, T. (2010b), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *Annals of Statistics*, 38, 3321–3351. [406]

Li, Y., Wang, N., and Carroll, R. J. (2010), "Generalized Functional Linear Models With Semiparametric Single-Index Interactions," *Journal of the American Statistical Association*, 105, 621–633. [406,408,411]

Lin, Y., and Zhang, H. H. (2006), "Component Selection and Smoothing in Multivariate Nonparametric Regression," *The Annals of Statistics*, 34, 2272–2297. [407,408,409]

Liu, X., Wang, L., and Liang, H. (2011), "Estimation and Variable Selection for Semiparametric Additive Partial Linear Models," *Statistica Sinica*, 21, 1225–1248. [408]

Lobell, D. B., and Burke, M. B. (2010), "On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change," *Agricultural and Forest Meteorology*, 150, 1443–1452. [406]

McLean, M. W., Hooker, G., Staicu, A. M., Scheipl, F., and Ruppert, D. (2014), "Functional Generalized Additive Models," *Journal of Computational and Graphical Statistics*, 23, 249–269. [416]

McMaster, G. S., and Wilhelm, W. (1997), "Growing Degree-Days: One Equation, Two Interpretations," *Agricultural and Forest Meteorology*, 87, 291–300. [414]

Meier, L., van de Geer, S., and Bühlmann, P. (2009), "High-Dimensional Additive Modeling," *The Annals of Statistics*, 37, 3779–3821. [410]

Müller, H. G., and Stadtmüller, U. (2005), "Generalized Functional Linear Models," *Annals of Statistics*, 33, 774–805. [406]

Müller, H.-G., Wu, Y., and Yao, F. (2013), "Continuously Additive Models for Nonlinear Functional Regression," *Biometrika*, 103, 607–622. [416]

Müller, H.-G., and Yao, F. (2008), "Functional Additive Models," *Journal of the American Statistical Association*, 103, 1534–1544. [406,408,415,416]

Prasad, A. K., Chai, L., Singh, R. P., and Kafatos, M. (2006), "Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters," *International Journal of Applied Earth Observation and Geoinformation*, 8, 26–33. [406]

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer. [406,409]

Ravikumar, P., Lafferty, J., Liu, H., and Wasserman, L. (2009), "Sparse Additive Models," *Journal of the Royal Statistical Society*, Series B, 71, 1009–1030. [410]

Smith, H. F. (1938), "An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops," *Journal of Agricultural Science*, 28, 1–23. [413]

Storlie, C. B., Bondell, H. D., Reich, B. J., and Zhang, H. H. (2011), "Surface Estimation, Variable Selection, and the Nonparametric Oracle Property," *Statistica Sinica*, 21, 679–705. [409]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia: SIAM. [408]

Wang, L., Xue, L., Qu, A., and Liang, H. (2014), "Estimation and Model Selection in Generalized Additive Partial Linear Models for Correlated Data With Diverging Number of Covariates," *Annals of Statistics*, 42, 592–624. [408]

Westcott, P. C., and Jewison, M. (2013), *Weather Effects on Expected Corn and Soybean Yields*, Washington DC: USDA Economic Research Service FDS-13g-01. [416]

Williams, J., Jones, C., Kiniry, J., and Spanel, D. (1989), "The Epic Crop Growth Model," *Transactions of the ASAE*, 32, 0497–0511. [414]

Yang, S., Logan, J., and Coffey, D. L. (1995), "Mathematical Formulae for Calculating the Base Temperature for Growing Degree Days," *Agricultural and Forest Meteorology*, 74, 61–74. [414]

Yao, F., Lei, E., and Wu, Y. (2015), "Effective Dimension Reduction for Sparse Functional Data," *Biometrika*, 102, 421–437. [406]

Yao, F., Müller, H. G., and Wang, J. L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [406]

Zhou, L., Huang, J. Z., and Carroll, R. J. (2008), "Joint Modelling of Paired Sparse Functional Data Using Principal Components," *Biometrika*, 95, 601–619. [406]

Zhu, H., Yao, F., and Zhang, H. H. (2014), "Structured Functional Additive Regression in Reproducing Kernel Hilbert Spaces," *Journal of the Royal Statistical Society*, Series B, 76, 581–603. [406,407,408,411,416]

$$= \omega_k^{-1/2}\left\{n^{-1/2}\sum_{j\neq k}(\omega_k-\omega_j)^{-1}\xi_{ij}\langle\Delta\psi_j,\psi_k\rangle\right\}\times\{1+O_p(n^{-1/2}\varrho_k^{-1})\}+O_p(n^{-1/2}),$$

which leads to (15). A closer calculation shows $\widehat{\zeta}_{ik}^d=\omega_k^{-1/2}\sum_{j\neq k}(\omega_k-\omega_j)^{-1}\xi_{ij}(\frac{1}{n}\sum_{i_1=1}^n\xi_{i_1k}\xi_{i_1j})$, and by Lemma 3,

$$\mathrm{E}(\widehat{\zeta}_{ik}^d)^2\lesssim n^{-1}\sum_{j\neq k}(\omega_k-\omega_j)^{-2}\omega_j^2\leq Cn^{-1}k^2.$$

LEMMA **1** *(Error from truncated KL expansion) The canonical parameter in the functional generalized linear model (1) is* $\eta_i=g\{\mu_Y(X_i,\mathbf{Z}_i)\}=\sum_{j=1}^\infty\zeta_{ij}\vartheta_{1j}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\sum_{j=1}^\infty\zeta_{ij}\vartheta_{2j}+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}$ *and its approximation using the first* m *principal components is*

$$\widetilde{\eta}_i=\boldsymbol{\vartheta}_1^{\mathrm{T}}\boldsymbol{\zeta}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}.$$

Define $\epsilon_i=q_1(\eta_i,Y_i)$ *and* $\widetilde{\epsilon}_i=q_1(\widetilde{\eta}_i,Y_i)$*, then under the assumptions in (C1),*

$$\widetilde{\eta}_i-\eta_i=O_p(m^{\frac{1-a-2b}{2}}),\quad \widetilde{\epsilon}_i-\epsilon_i=O_p(m^{\frac{1-a-2b}{2}}). \tag{S.1}$$

Proof of Lemma 1: Under assumption (C1), for $\ell=1,2$,

$$\mathrm{E}(\textstyle\sum_{j=m+1}^\infty\zeta_{ij}\vartheta_{\ell j})^2\leq C\sum_{j=m+1}^\infty j^{-a-2b}\leq Cm^{1-a-2b},$$

which implies $\sum_{j=m+1}^\infty\zeta_{ij}\vartheta_{\ell j}=O_p(m^{\frac{1-a-2b}{2}})$ and hence

$$\widetilde{\eta}_i-\eta_i=\sum_{j=m+1}^\infty\zeta_{ij}\vartheta_{1j}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\sum_{j=m+1}^\infty\zeta_{ij}\vartheta_{2j}=O_p(m^{\frac{1-a-2b}{2}}).$$

By Taylor expansion, $\widetilde{\epsilon}_i-\epsilon_i=q_2(\eta_i,Y_i)(\widetilde{\eta}_i-\eta_i)\times\{1+o_p(1)\}=O_p(m^{\frac{1-a-2b}{2}})$.

LEMMA **2** *If* $\boldsymbol{A}$ *is nonsingular, then* $(\boldsymbol{A}+\boldsymbol{B})^-=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})^-\boldsymbol{A}^{-1}$.

Proof: This lemma is a direct result of Section 4 of Henderson and Searle (1981).

LEMMA **3** *If* $|p_k|\leq Ck^{-\pi}$ *for each* k*, then, for* $d=2,4,6,\ldots,$

$$\sum_{k:k\neq j}(\omega_j-\omega_k)^{-d}p_k^2\leq C\{\lambda_{ad-2\pi}(j)+j^{ad+d-2\pi}\}\leq C(1+j^{ad+d-2\pi}),$$

where, for each real number r*,*

$$\lambda_r(m)=\begin{cases}m^{r+1}, & \text{if}\quad r>-1,\\ \log\,m, & \text{if}\quad r=-1,\\ 1, & \text{if}\quad r<-1.\end{cases}$$

Proof: see page 2174 of Cai and Hall (2006).

LEMMA **4** *Under a nonparametric regression model $Y_i = m(X_i) + e_i$, $i = 1, \ldots, n$, where X_i are iid random variables on $[0,1]$ with a continuous density $f_X(x) > 0$ for all $x \in [0,1]$, both $f_X(\cdot)$ and $m(\cdot)$ are twice continuously differentiable functions on $[0,1]$, and e_i's are iid zero-mean error with $\sigma^2 = var(e_i)$. Suppose the kernel function $K(\cdot)$ and the bandwidth satisfy assumptions (C2), denote $w_i(x) = K_h(X_i - x)/\{\sum_{j=1}^n K_h(X_j - x)\}$ and let $q(y,x)$ be a function such that $E\{q(Y,X)|X = x\}$ is a twice continuously differentiable function of x, then*

(a) $\sup_{x\in[0,1]} \left| \sum_{i=1}^n w_i(x) Y_i - m(x) \right| = O(h^2 + \delta_n) \quad a.s.;$

(b) $\left\{ \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n w_i(X_j) q(Y_i, X_i) \right\} = \mathrm{E}\{q(Y,X)\} \times \{1 + O_p(h^2 + \delta_n)\} + O_p(n^{-1/2}).$

Proof of Lemma 4: Part (a) is a standard uniform convergence rate result for kernel estimators (Härdle et al., 1988). Part (b) follows from

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n w_i(X_j) q(Y_i, X_i) \quad = \quad \left[\frac{1}{n} \sum_{j=1}^n \mathrm{E}\{q(Y,X)|X_j\} \right] \times \{1 + O(h^2 + \delta_n)\}.$$

LEMMA **5** *Denote*

$$\boldsymbol{U}_1(\mathbf{z};\boldsymbol{\theta}) = \sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) q_2(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)[\{\boldsymbol{\vartheta}_1 + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\vartheta}_2\}^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)] \begin{pmatrix} 1 \\ \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}^*_{i0,1} \end{pmatrix},$$

$$\boldsymbol{U}_2(\mathbf{z};\boldsymbol{\theta}) = \sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) q_1(\widetilde{\eta}_i, Y_i)\{\boldsymbol{\vartheta}_2^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)\} \begin{pmatrix} 1 \\ \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}^*_{i0,1} \end{pmatrix},$$

$$\boldsymbol{U}_3(\mathbf{z};\boldsymbol{\theta}) = \sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) q_2(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \begin{pmatrix} 1 \\ \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}^*_{i0,1} \end{pmatrix} (\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)^{\mathrm{T}},$$

$$\boldsymbol{U}_4(\mathbf{z};\boldsymbol{\theta}) = \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z})\{q_2(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}) + q_1(\widetilde{\eta}_i, Y_i)\} \begin{pmatrix} 1 \\ \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}^*_{i0,1} \end{pmatrix} (\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)^{\mathrm{T}}.$$

Let $U_{11}(\mathbf{z},\boldsymbol{\theta})$ be the first element of $\boldsymbol{U}_1(\mathbf{z},\boldsymbol{\theta})$. Under conditions C1 and C2, when $m = o(n^{\frac{1}{2a+2}})$,

$$U_{11}(\mathbf{z},\boldsymbol{\theta}) = n^{-1/2} \sum_{k=1}^m \sum_{l\neq k} \left[\frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k - \omega_l)} \mathrm{E}\{q_2(\eta_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\zeta_{il}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1} = \boldsymbol{\theta}\mathbf{z}\} \right.$$
$$\left. \times\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\vartheta_{2k}\}\langle \Delta\psi_k, \psi_l\rangle \right] + O_p\{n^{-1/2}(h^2 + \delta_n + m^{\frac{1-a-2b}{2}})\}.$$

In addition, $\|\boldsymbol{U}_1(\mathbf{z};\boldsymbol{\theta})\| = O_p(n^{-1/2})$, $\|\boldsymbol{U}_2(\mathbf{z};\boldsymbol{\theta})\| = O_p\{n^{-1/2}(\delta_n + m^{\frac{1-a-2b}{2}})$, $\|\boldsymbol{U}_3(\mathbf{z},\boldsymbol{\theta})\| = O_p(m^{3/2}n^{-1/2})$, $\|\boldsymbol{U}_4(\mathbf{z},\boldsymbol{\theta})\| = O_p(m^{3/2}n^{-1/2})$.

Proof of Lemma 5: The asymptotic expansion of $U_{11}(\mathbf{z},\widehat{\boldsymbol{\theta}})$ follows from Proposition 1. Since $\vartheta_{\ell k} \lesssim k^{-b-a/2}$ for $\ell = 1, 2$ and all $k \leq m$,

$$\begin{aligned} \mathrm{E}\{U_{11}^2(\mathbf{z},\boldsymbol{\theta})\} &\lesssim n^{-1}\sum_{k=1}^{m}\sum_{l\neq k}\frac{\omega_l}{\omega_k(\omega_k-\omega_l)^2}k^{-a-2b}\omega_k\omega_l \\ &\lesssim n^{-1}\sum_{k=1}^{m}k^{2-a-2b} \quad \text{(by Lemma 3)} \\ &\lesssim n^{-1}. \end{aligned}$$

Similar calculation applies to $U_{12}(\mathbf{z};\boldsymbol{\theta})$, hence $\|\boldsymbol{U}_1(\mathbf{z};\boldsymbol{\theta})\| = O_p(n^{-1/2})$. Further, by (15)

$$\mathrm{E}|\boldsymbol{\vartheta}_\ell^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)| \lesssim \sum_{k\leq m} k^{-b-a/2}kn^{-1/2} \lesssim n^{-1/2}.$$

On the other hand, since $q_1(\eta_i, Y_i)$ has mean 0 and is uncorrelated with $\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i$, by Lemmas 1 and 4,

$$\begin{aligned} |\{U_{21}(\mathbf{z};\boldsymbol{\theta})| &\lesssim \left|\sum_{i=1}^{n} w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})q_1(\eta_i,Y_i)\{\boldsymbol{\vartheta}_2^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{\zeta}_i)\}\right| + \left|\sum_{i=1}^{n} w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})q_2(\eta_i,Y_i)(\widetilde{\eta}_i-\eta_i)\{\boldsymbol{\vartheta}_2^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{\zeta}_i)\}\right| \\ &\lesssim n^{-1/2}\delta_n + n^{-1/2}m^{\frac{1-a-2b}{2}}. \end{aligned}$$

Hence, we have $\|\boldsymbol{U}_2(\mathbf{z};\boldsymbol{\theta})\| = O_p\{n^{-1/2}(\delta_n + m^{\frac{1-a-2b}{2}})\}$. Furthermore, by Proposition 1 and Lemma 4,

$$\mathrm{E}\|\boldsymbol{U}_3^{\mathrm{T}}(\mathbf{z},\boldsymbol{\theta})\|^2 \lesssim \sum_{k\leq m} k^2/n \lesssim m^3/n,$$

which implies $\|\boldsymbol{U}_3^{\mathrm{T}}(\mathbf{z},\boldsymbol{\theta})\| = O_p(m^{3/2}n^{-1/2})$. A similar calculation shows that $\|\boldsymbol{U}_4^{\mathrm{T}}(\mathbf{z},\boldsymbol{\theta})\|$ has the same rate.

LEMMA **6** *Denote*

$$\widetilde{\mathcal{A}} = -n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}q_2(\widetilde{\eta}_i, Y_i)\begin{Bmatrix} \widehat{\boldsymbol{\zeta}}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\widehat{\boldsymbol{\zeta}}_i \\ \mathbf{Z}_i \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})\mathbf{Z}_{ij,1} \end{Bmatrix}^{\otimes 2} \tag{S.2}$$

Under conditions C1 and C2, when $m = o(n^{\frac{1}{2a+2}})$, *we have* $\|\widetilde{\mathcal{A}} - \mathcal{A}\| = O_p\{mh^2 + m^2 n^{-1/2} + m^{\frac{3-a-2b}{2}}\} = o_p(1)$, *where* $\mathcal{A}$ *is defined in (A.2). As a result*

$$|\lambda_{max}(\widetilde{\mathcal{A}}) - \lambda_{max}(\mathcal{A})| = o_p(1), \quad |\lambda_{min}(\widetilde{\mathcal{A}}) - \lambda_{min}(\mathcal{A})| = o_p(1).$$

Denote

$$\mathbb{U}_n = n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_2(\eta_i, Y_i) \begin{Bmatrix} \boldsymbol{\zeta}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i)\boldsymbol{\zeta}_i \\ \mathbf{Z}_i \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})\mathbf{Z}_{ij,1} \end{Bmatrix} \{\boldsymbol{\vartheta}_1 + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\vartheta}_2\}^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i). \tag{S.3}$$

then

$$\mathbb{U}_n = \sum_{k=1}^{m}\sum_{l\neq k} \frac{\omega_j^{1/2}}{\omega_k^{1/2}(\omega_k - \omega_l)} \left[n^{-1}\sum_{i=1}^{n} \{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\vartheta_{2k}\}\zeta_{il}\boldsymbol{\varpi}_i \right] \langle \Delta\psi_k, \psi_l\rangle + \widetilde{O}_p\{n^{-1/2}(h^2 + \delta_n)\}.$$

where $\boldsymbol{\varpi}_i$ *is defined in (A.1).*

Proof of Lemma 6: Denote

$$\mathcal{A}^* = -n^{-1}\textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_2(\widetilde{\eta}_i, Y_i) \begin{Bmatrix} \boldsymbol{\zeta}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i \\ \mathbf{Z}_i \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})\mathbf{Z}_{ij,1} \end{Bmatrix}^{\otimes 2},$$

Let $\mathcal{A}^*_{jj'}$ and $\mathcal{A}_{jj'}$ be the (j,j')th entries of $\mathcal{A}^*$ and $\mathcal{A}$ respectively, and by Lemmas 1 and 4 $\mathcal{A}^*_{jj'} - \mathcal{A}_{jj'} = O_p(h^2 + n^{-1/2} + m^{\frac{1-a-2b}{2}})$ and hence

$$\|\mathcal{A}^* - \mathcal{A}\|^2 = O_p\{m^2(h^4 + n^{-1} + m^{1-a-2b})\}.$$

By the asymptotic expansion of FPCA in Proposition 1, $\mathrm{E}\{(\widehat{\zeta}_{ik_1} - \zeta_{ik_1})\zeta_{ik_2}\} = O(n^{-1}\varrho_{k_1}^{-1})$ and $\mathrm{var}\{(\widehat{\zeta}_{ik_1} - \zeta_{ik_1})\zeta_{ik_2}\} = O(n^{-1}k_1^2)$, for any $k_1, k_2 \leq m$. By Lemma 4,

$$\mathrm{E}\|\widetilde{\mathcal{A}} - \mathcal{A}^*\|^2 \lesssim \sum_{k_1\leq m}\sum_{k_2\leq m} (n^{-2}\varrho_{k_1}^{-2} + n^{-1}k_1^2) \lesssim n^{-1}m^4.$$

Combining the results, $\|\widetilde{\mathcal{A}} - \mathcal{A}\| = O_p(mh^2 + m^2n^{-1/2} + m^{-\frac{a+2b-3}{2}}) = o_p(1)$. By Weyl's inequality, $|\lambda_{max}(\widetilde{\mathcal{A}}) - \lambda_{max}(\mathcal{A})| \leq \|\widetilde{\mathcal{A}} - \mathcal{A}\|$, $\|\lambda_{min}(\widetilde{\mathcal{A}}) - \lambda_{min}(\mathcal{A})\| \leq \|\widetilde{\mathcal{A}} - \mathcal{A}\|$. The calculation of $\mathbb{U}_n$ is similar with that of $\boldsymbol{U}_{11}(\mathbf{z}, \boldsymbol{\theta})$ in Lemma 5. The proof is completed.

S.2 Proof of Theorem 1:

S.2.1 Asymptotic expansion of the nonparametric estimator

For any $\mathbf{z} \in \mathcal{D}$, denote $u = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}$, $(a_0, a_1) = (\mathcal{S}, \mathcal{S}^{(1)})(u)$, $a_1^* = ha_1$, $\mathbf{Z}_{i0,1} = \mathbf{Z}_{i1} - \mathbf{z}$ and $\mathbf{Z}_{i0,1}^* = (\mathbf{Z}_{i1} - \mathbf{z})/h$. Denote $w_i(u) = K\{(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{Z}_{i1} - u)/h\}/\sum_{l=1}^n K\{(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{Z}_{l1} - u)/h\}$, $u \in \mathbb{R}$. The updated estimate of (a_0, a_1^*) is the minimizer of (9) denoted as $(\widehat{a}_{0,curr}, \widehat{a}_{1,curr}^*)$. Given the parametric estimate $\widehat{\boldsymbol{\Theta}}_{prev}$ from the previous iteration, $(\widehat{a}_{0,curr}, \widehat{a}_{1,curr}^*)$ satisfies the following estimating equation.

$$
\begin{aligned}
\mathbf{0} \;=\; & \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_1 \left\{ \widehat{\boldsymbol{\vartheta}}_{1,prev}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i + (\widehat{a}_{0,curr} + \widehat{a}_{1,curr}^* \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^*)\widehat{\boldsymbol{\vartheta}}_{2,prev}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i + \widehat{\boldsymbol{\beta}}_{prev}^{\mathrm{T}}\mathbf{Z}_i, Y_i \right\} \\
& \qquad \times \left(\widehat{\boldsymbol{\vartheta}}_{2,prev}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i\right) \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \\
= \; & \Bigg[\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_1(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \\
& \quad + \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_2(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)^2 \{\widehat{a}_{0,curr} + \widehat{a}_{1,curr}\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1} - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\} \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \\
& \quad + \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_2(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \begin{pmatrix} 1 \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}) \\ \mathbf{Z}_i \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \widehat{\boldsymbol{\vartheta}}_{1,prev}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\vartheta}_1^{\mathrm{T}}\boldsymbol{\zeta}_i \\ \widehat{\boldsymbol{\vartheta}}_{2,prev}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i \\ \widehat{\boldsymbol{\beta}}_{prev} - \boldsymbol{\beta} \end{pmatrix} \\
& \quad + \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_1(\widetilde{\eta}_i, Y_i) \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} (\widehat{\boldsymbol{\vartheta}}_{2,prev}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \Bigg] \times \{1 + o_p(1)\} \\
= \; & \Bigg[\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_1(\widetilde{\eta}_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \\
& \quad + \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_2(\eta_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)^2 \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \{\widehat{a}_{0,curr} + \widehat{a}_{1,curr}\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1} - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\} \\
& \quad + \sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z}) q_2(\eta_i, Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i \\ \mathbf{Z}_i \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \widehat{\boldsymbol{\vartheta}}_{1,prev} - \boldsymbol{\vartheta}_1 \\ \widehat{\boldsymbol{\vartheta}}_{2,prev} - \boldsymbol{\vartheta}_2 \\ \widehat{\boldsymbol{\beta}}_{prev} - \boldsymbol{\beta} \end{pmatrix}
\end{aligned}
$$

$$+\{\boldsymbol{U}_1(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})+\boldsymbol{U}_2(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})+\boldsymbol{U}_3(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})(\widehat{\boldsymbol{\vartheta}}_{1,prev}-\boldsymbol{\vartheta}_1)+\boldsymbol{U}_4(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})(\widehat{\boldsymbol{\vartheta}}_{2,prev}-\boldsymbol{\vartheta}_2)\}\Big]$$
$$\times\{1+o_p(1)\},$$

where $\boldsymbol{U}_\ell(\mathbf{z},\boldsymbol{\theta})$, $\ell=1,\ldots,4$, are defined in Lemma 5. For observations with $|\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i1}-u|\le h$, we have

$$\widehat{a}_{0,curr}+\widehat{a}^*_{1,curr}\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{i0,1}-\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})=\begin{pmatrix}1\\ \widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}^*_{i0,1}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\widehat{a}_{0,curr}-a_0\\ \widehat{a}^*_{1,curr}-a^*_1\end{pmatrix}+\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\mathbf{Z}_{i0,1}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_{prev}-\boldsymbol{\theta}_0)$$
$$-\frac{1}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i0,1})^2-\frac{1}{6}\mathcal{S}^{(3)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i0,1})^3+O(h^4).$$

By Lemma 5, solving the local estimation equation leads to

$$\begin{aligned}\widehat{a}_{0,curr}-a_0 &= -\{\widehat{\mathcal{V}}_1^{-1}(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})\widehat{\mathcal{V}}_2^{\mathrm{T}}(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})\}\times(\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta}_0)+O_p(n^{-1/2}m^{3/2}\|\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta}_0\|)\\ &\quad +R_{0,n}(\mathbf{z})+\frac{\sigma_K^2}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})h^2+O_p\{h^4+n^{-1/2}(\delta_n+m^{\frac{1-a-2b}{2}})\},\end{aligned}$$

where

$$\begin{aligned}\widehat{\mathcal{V}}_1(\mathbf{z};\boldsymbol{\theta}) &= -\sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})q_2(\eta_i,Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)^2,\\ \widehat{\mathcal{V}}_2(\mathbf{z};\boldsymbol{\theta}) &= \begin{bmatrix}\sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\boldsymbol{\pi}_{i1}\\ \mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\{\sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\boldsymbol{\pi}_{i2}-\mathbf{z}\widehat{\mathcal{V}}_1(\mathbf{z};\boldsymbol{\theta})\}\end{bmatrix},\\ R_{0,n}(\mathbf{z}) &= \widehat{\mathcal{V}}_1^{-1}(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\widetilde{\epsilon}_i+\widehat{\mathcal{V}}_1^{-1}(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev})U_{11}(\mathbf{z};\widehat{\boldsymbol{\theta}}_{prev}).\end{aligned}$$

By Lemma 4, $\widehat{\mathcal{V}}_1(\mathbf{z};\boldsymbol{\theta})-\mathcal{V}_1(\mathbf{z};\boldsymbol{\theta})=O_p(h^2+\delta_n)$, $\|\widehat{\mathcal{V}}_2(\mathbf{z};\boldsymbol{\theta})-\mathcal{V}_2(\mathbf{z};\boldsymbol{\theta})\|=O_p\{m^{1/2}(\delta_n+h^2)\}=o_p(1)$. As a result,

$$\begin{aligned}\widehat{a}_{0,curr}-a_0 &= -\{\mathcal{V}_1^{-1}(\mathbf{z};\boldsymbol{\theta})\mathcal{V}_2^{\mathrm{T}}(\mathbf{z};\boldsymbol{\theta})\}(\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta}_0)\times\{1+o_p(1)\}+R_{0,n}(\mathbf{z})\\ &\quad +\frac{\sigma_K^2}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})h^2+O_p\{h^4+n^{-1/2}(\delta_n+m^{\frac{1-a-2b}{2}})\}. \qquad \text{(S.4)}\end{aligned}$$

Similarly, we have

$$\widehat{a}^*_{1,curr}-a^*_1=O_p\{(m^{1/2}h+n^{-1/2}m^{3/2})\|\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta}\|+h^3+\delta_n+h(m^{\frac{1-a-2b}{2}}+n^{-1/2})\}. \qquad \text{(S.5)}$$

S.2.2 Asymptotic expansion of the parametric estimator

Given $\widehat{a}_{0j} = \widehat{\mathcal{S}}(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{j1})$ and $\widehat{a}_{1j} = \widehat{\mathcal{S}}^{(1)}(\widehat{\boldsymbol{\theta}}_{prev}^{\mathrm{T}}\mathbf{Z}_{j1})$, $j = 1, \ldots, n$, the updated parametric component $\widehat{\boldsymbol{\Theta}}_{curr}$ satisfies the following estimating equation

$$
\begin{aligned}
\mathbf{0} &= n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_1 \left\{ \widehat{\boldsymbol{\vartheta}}_{1,curr}\widehat{\boldsymbol{\zeta}}_i + (\widehat{a}_{0j} + \widehat{a}_{1j}\widehat{\boldsymbol{\theta}}_{curr}^{\mathrm{T}}\mathbf{Z}_{ij,1})\widehat{\boldsymbol{\vartheta}}_{2,curr}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i + \widehat{\boldsymbol{\beta}}_{curr}\mathbf{Z}_i, Y_i \right\} \\
&\qquad \times \begin{Bmatrix} \widehat{\boldsymbol{\zeta}}_i \\ (\widehat{a}_{0j} + \widehat{a}_{1j}\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{Z}_{ij,1})\widehat{\boldsymbol{\zeta}}_i \\ \mathbf{Z}_i \\ (\widehat{\boldsymbol{\vartheta}}_{2,curr}\widehat{\boldsymbol{\zeta}}_i)\widehat{a}_{1j}\mathbf{Z}_{ij,1} \end{Bmatrix} \\
&= \Bigg[n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_1(\widetilde{\eta}_i, Y_i) \begin{Bmatrix} \widehat{\boldsymbol{\zeta}}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i)\widehat{\boldsymbol{\zeta}}_i \\ \mathbf{Z}_i \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\widehat{a}_{1j}\mathbf{Z}_{ij,1} \end{Bmatrix} \\
&\quad + n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} q_2(\widetilde{\eta}_i, Y_i) \begin{Bmatrix} \widehat{\boldsymbol{\zeta}}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\widehat{\boldsymbol{\zeta}}_i \\ \mathbf{Z}_i \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\widehat{a}_{1j}\mathbf{Z}_{ij,1} \end{Bmatrix} \begin{Bmatrix} 1 \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}) \\ \mathbf{Z}_i \\ \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i \end{Bmatrix}^{\mathrm{T}} \begin{Bmatrix} \widehat{\boldsymbol{\vartheta}}_{1,curr}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\vartheta}_1^{\mathrm{T}}\boldsymbol{\zeta}_i \\ \widehat{\boldsymbol{\vartheta}}_{2,curr}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i \\ \widehat{\boldsymbol{\beta}}_{curr} - \boldsymbol{\beta} \\ (\widehat{a}_{0j} + \widehat{a}_{1j}\widehat{\boldsymbol{\theta}}_{curr}^{\mathrm{T}}\mathbf{Z}_{ij,1}) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}) \end{Bmatrix} \\
&\quad + n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_1(\widetilde{\eta}_i, Y_i) \begin{Bmatrix} \mathbf{0}_m \\ \{(\widehat{a}_{0j} + \widehat{a}_{1j}\widehat{\boldsymbol{\theta}}_{curr}^{\mathrm{T}}\mathbf{Z}_{ij,1}) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\}\widehat{\boldsymbol{\zeta}}_i \\ \mathbf{0}_m \\ \widehat{a}_{0j}\mathbf{Z}_{ij,1}(\widehat{\boldsymbol{\vartheta}}_{2,curr}^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i) \end{Bmatrix} \Bigg] \times \{1 + o_p(1)\} \\
&= \Bigg(n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} \begin{Bmatrix} \boldsymbol{\zeta}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i \\ \mathbf{Z}_i \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})\mathbf{Z}_{ij,1} \end{Bmatrix} \Big[q_1(\widetilde{\eta}_i, Y_i) + q_2(\widetilde{\eta}_i, Y_i)\{\boldsymbol{\vartheta}_1 + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\vartheta}_2\}^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i) \Big]
\end{aligned}
$$

$$+n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}q_2(\widetilde{\eta}_i,Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\begin{Bmatrix}\boldsymbol{\zeta}_i\\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i\\ \mathbf{Z}_i\\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})\mathbf{Z}_{ij,1}\end{Bmatrix}\left\{\begin{pmatrix}1\\ \widehat{\boldsymbol{\theta}}_{curr}^{\mathrm{T}}\mathbf{Z}_{ij,1}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\widehat{a}_{0j}-a_{0j}\\ \widehat{a}_{1j}-a_{1j}\end{pmatrix}\right.$$

$$\left.-\frac{1}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{ij,1})^2-\frac{1}{6}\mathcal{S}^{(3)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{ij,1})^3+O_p(h^4)\right\}-\widetilde{\mathcal{A}}(\widehat{\boldsymbol{\Theta}}_{curr}-\boldsymbol{\Theta})\Bigg)\times\{1+o_p(1)\}$$

where $\widetilde{\mathcal{A}}$ is defined in (S.2).

By (S.5),

$$n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}q_2(\widetilde{\eta}_i,Y_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\begin{Bmatrix}\boldsymbol{\zeta}_i\\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i\\ \mathbf{Z}_i\\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{j1})\mathbf{Z}_{ij,1}\end{Bmatrix}(\widehat{\boldsymbol{\theta}}_{curr}^{\mathrm{T}}\mathbf{Z}_{ij,1})(\widehat{a}_{1j,curr}-a_{1j})$$

$$=\quad \mathcal{D}(\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta})+\widetilde{O}_p\{h^4+h^2(m^{\frac{1-a-2b}{2}}+n^{-1/2})\},$$

where $\mathcal{D}$ is a square matrix with order $2m+d+d_1$, and $\|\mathcal{D}\|=O_p(mh^2+m^2n^{-1/2}h)=o_p(1)$.

Substituting (S.4) into the estimating equation and combining results in lemma 6,

$$\begin{aligned}&\widetilde{\mathcal{A}}(\widehat{\boldsymbol{\Theta}}_{curr}-\boldsymbol{\Theta})\\ =\quad&\left[\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}\boldsymbol{\zeta}_i\\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i\\ \mathbf{Z}_i\\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\{\mathbf{Z}_{i1}-\nu_{Z_1}(\mathbf{Z}_{i1},\boldsymbol{\theta})\}\end{bmatrix}\left[\widetilde{\epsilon}_i+q_2(\widetilde{\eta}_i,Y_i)\{\boldsymbol{\vartheta}_1+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\vartheta}_2\}^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{\zeta}_i)\right]\right.\\ &\qquad-\frac{1}{n}\sum_{j=1}^{n}\mathcal{V}_2(\mathbf{Z}_{j1},\boldsymbol{\theta}_0)\mathcal{V}_1^{-1}(\mathbf{Z}_{j1},\boldsymbol{\theta}_0)\Bigg(\mathcal{V}_2^{\mathrm{T}}(\mathbf{Z}_{j1},\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta}_0)\\ &\qquad\qquad+\sum_{i=1}^{n}w_{ij}(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\left[\widetilde{\epsilon}_i+q_2(\widetilde{\eta}_i,Y_i)\{\boldsymbol{\vartheta}_1+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\vartheta}_2\}^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{\zeta}_i)\right]\Bigg)\Bigg]\times\{1+o_p(1)\}\\ =\quad&\{\mathcal{N}_n+\mathcal{N}_{1,n}+\widetilde{\mathcal{C}}(\widehat{\boldsymbol{\Theta}}_{prev}-\boldsymbol{\Theta})\}\times\{1+o_p(1)\},\end{aligned}$$

where

$$\mathcal{N}_n = n^{-1}\sum_{i=1}^{n}\widetilde{\epsilon}_i\mathcal{E}_i, \tag{S.6}$$

$$\mathcal{N}_{1,n} = \frac{1}{n}\sum_{i=1}^{n} q_2(\eta_i, Y_i)\mathcal{E}_i\{\boldsymbol{\vartheta}_1 + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\vartheta}_2\}^{\mathrm{T}}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i),$$

$$\mathcal{E}_i = \begin{bmatrix} \boldsymbol{\zeta}_i - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta})\boldsymbol{\zeta}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_{i1}\} \\ \mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta})\mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_1)\boldsymbol{\zeta}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_{i1}\} \\ \mathbf{Z}_i - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{Z}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_{i1}\} \\ (\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_{i1})\{\mathbf{Z}_{i1} - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)\widehat{\boldsymbol{\pi}}_2(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)\} \end{bmatrix},$$

$$\widetilde{\mathcal{C}} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_1^{-1}(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)\mathcal{V}_2(\mathbf{Z}_{i1};\boldsymbol{\theta}_0)\mathcal{V}_2^{\mathrm{T}}(\mathbf{Z}_{i1};\boldsymbol{\theta}_0).$$

Define

$$\mathcal{C} = \mathrm{E}\left[\mathcal{V}_1^{-1}\begin{Bmatrix} \widehat{\boldsymbol{\pi}}_1\widehat{\boldsymbol{\pi}}_1^{\mathrm{T}} & \mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\widehat{\boldsymbol{\pi}}_1(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\mathbf{Z}_1)^{\mathrm{T}} \\ \mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\mathbf{Z}_1)\widehat{\pi}_1^{\mathrm{T}} & \{\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\}^2(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\mathbf{Z}_1)^{\otimes 2} \end{Bmatrix}\right], \tag{S.7}$$

it is easily to see $\|\widetilde{\mathcal{C}} - \mathcal{C}\| = O_p(n^{-\frac{1}{2}}m) = o_p(1)$, hence $|\lambda_{max}(\widetilde{\mathcal{C}}) - \lambda_{max}(\mathcal{C})| = o_p(1)$ and $|\lambda_{min}(\widetilde{\mathcal{C}}) - \lambda_{min}(\mathcal{C})| = o_p(1)$.

The estimation equation leads to

$$\widetilde{\mathcal{A}}(\widehat{\boldsymbol{\Theta}}_{curr} - \boldsymbol{\Theta}) = \mathcal{N}_n + \mathcal{N}_{1,n} + \widetilde{\mathcal{C}}(\widehat{\boldsymbol{\Theta}}_{prev} - \boldsymbol{\Theta}) + \widetilde{O}_p(n^{-1/2}\delta_n + h^4 + m^{\frac{1}{2}-\frac{a}{2}-b}h^2),$$

hence

$$\begin{aligned}\widehat{\boldsymbol{\Theta}}_{curr} - \boldsymbol{\Theta} &= \mathcal{A}^{-}(\mathcal{N}_n + \mathcal{N}_{1,n}) + \mathcal{A}^{-}\mathcal{C}(\widehat{\boldsymbol{\Theta}}_{prev} - \boldsymbol{\Theta}) \times \{1 + o_p(1)\} \\ &\quad + \widetilde{O}_p(n^{-1/2}\delta_n m + mh^4 + m^{\frac{3}{2}-\frac{a}{2}-b}h^2).\end{aligned}$$

By Generalized Cauchy-Schwarz inequality, the eigenvalues of $\mathcal{A}^{-}\mathcal{C}$ are strictly less than 1. At convergence,

$$\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} = \mathcal{A}^{-}(\mathcal{N}_n + \mathcal{N}_{1,n}) + \widetilde{O}_p(n^{-1/2}\delta_n m + mh^4 + m^{\frac{3}{2}-\frac{a}{2}-b}h^2). \tag{S.8}$$

By the asymptotic expansion of $\widehat{\zeta}_{ik} - \zeta_{ik}$ given Proposition 1 and by calculations similar to those in Lemma 5,

$$\begin{aligned}\mathcal{N}_{1,n} &= -n^{-1/2}\sum_{k=1}^{m}\sum_{l\neq k}\frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k - \omega_l)}\mathrm{E}\left[\mathcal{E}_i\rho_2(\eta_i)\zeta_{il}\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\vartheta_{2k}\}\right]\langle\Delta\psi_l, \psi_k\rangle \\ &\quad + \widetilde{O}_p\{n^{-1/2}(h^2 + \delta_n)\}.\end{aligned} \tag{S.9}$$

S.2.3 Asymptotic distribution of the parametric estimator

The matrix $\mathcal{A}$ defined in (A.2) in Appendix A can be decomposed as $\mathcal{A} = \mathcal{G} + \mathcal{H}$, where

$$\mathcal{G} = \mathrm{E}\left[\rho_2(\eta_i)\left\{\begin{array}{c}\boldsymbol{\zeta}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i \\ \mathbf{Z}_i \\ \mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})(\boldsymbol{\vartheta}_{2,0}^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathbf{Z}_{i1}\end{array}\right\}^{\otimes 2}\right], \tag{S.10}$$

and

$$\mathcal{H} = \begin{pmatrix} \mathbf{0}_{(2m+d)\times(2m+d)} & -\mathrm{E}\left\{\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\widehat{\boldsymbol{\pi}}_1(\mathbf{Z}_{i1},\boldsymbol{\theta})\mathbf{Z}_{i1}^{\mathrm{T}}\right\} \\ -\mathrm{E}\left\{\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\mathbf{Z}_{i1}\widehat{\boldsymbol{\pi}}_1^{\mathrm{T}}(\mathbf{Z}_{i1},\boldsymbol{\theta})\right\} & \mathrm{E}\left[\left\{\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\right\}^2\left\{-\widehat{\boldsymbol{\pi}}_2\mathbf{Z}_{i1}^{\mathrm{T}} - \mathbf{Z}_{i1}\widehat{\boldsymbol{\pi}}_2^{\mathrm{T}} + \mathcal{V}_1(\mathbf{Z}_{i1},\boldsymbol{\theta})\mathbf{Z}_{i1}\mathbf{Z}_{i1}^{\mathrm{T}}\right\}\right] \end{pmatrix}. \tag{S.11}$$

We partition $\mathcal{G}$ as $\mathcal{G} = \begin{pmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} \\ \mathcal{G}_{21} & \mathcal{G}_{22} \end{pmatrix}$, with

$$\mathcal{G}_{11} = \mathrm{E}\left[\rho_2(\eta_i)\left\{\begin{array}{c}\boldsymbol{\zeta}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i\end{array}\right\}^{\otimes 2}\right]. \tag{S.12}$$

$\mathcal{A}$ and $\mathcal{H}$ are accordingly partitioned as

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix}, \qquad \mathcal{H} = \begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{21} & \mathcal{H}_{22} \end{pmatrix}, \tag{S.13}$$

with $\mathcal{A}_{ij} = \mathcal{G}_{ij} + \mathcal{H}_{ij}, i = 1, 2, j = 1, 2$.

Under assumption (C2.7), $\mathcal{A}_{11}$ is invertible. By matrix algebra,

$$\mathcal{A}^{-} = \begin{pmatrix} \mathcal{A}^{11} & \mathcal{A}^{12} \\ \mathcal{A}^{21} & \mathcal{A}^{22} \end{pmatrix},$$

where

$$\begin{aligned} \mathcal{A}^{11} &= \mathcal{A}_{11}^{-1} + \mathcal{A}_{11}^{-1}\mathcal{A}_{12}(\mathcal{A}_{22} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{A}_{12})^{-}\mathcal{A}_{21}\mathcal{A}_{11}^{-1}, \\ \mathcal{A}^{12} &= -\mathcal{A}_{11}^{-1}\mathcal{A}_{12}(\mathcal{A}_{22} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{A}_{12})^{-}, \\ \mathcal{A}^{21} &= -(\mathcal{A}_{22} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{A}_{12})^{-}\mathcal{A}_{21}\mathcal{A}_{11}^{-1}, \\ \mathcal{A}^{22} &= (\mathcal{A}_{22} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{A}_{12})^{-}. \end{aligned} \tag{S.14}$$

We now derive the asymptotic distribution of $\widehat{\mathbf{\Theta}}_Z - \mathbf{\Theta}_Z$. Using the same partition, we write $\mathcal{E}_i = (\mathcal{E}_i^{[1]}, \mathcal{E}_i^{[2]})^{\mathrm{T}}$, $\mathcal{N}_n = (\mathcal{N}_n^{[1]}, \mathcal{N}_n^{[2]})^{\mathrm{T}}$, $\mathcal{N}_{1,n} = (\mathcal{N}_{1,n}^{[1]}, \mathcal{N}_{1,n}^{[2]})^{\mathrm{T}}$ where

$$\mathcal{E}_i^{[1]} = \begin{bmatrix} \boldsymbol{\zeta}_i - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1}, \boldsymbol{\theta})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta})\boldsymbol{\zeta}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1 = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}\} \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\boldsymbol{\zeta}_i - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1}, \boldsymbol{\theta})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta})\mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}_1)\boldsymbol{\zeta}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1 = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}\} \end{bmatrix},$$

$$\mathcal{E}_i^{[2]} = \begin{bmatrix} \mathbf{Z}_i - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1}, \boldsymbol{\theta})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{Z}|\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1 = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1}\} \\ (\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\{\mathbf{Z}_{i1} - \mathcal{V}_1^{-1}(\mathbf{Z}_{i1}, \boldsymbol{\theta})\widehat{\pi}_2(\mathbf{Z}_{i1}, \boldsymbol{\theta})\} \end{bmatrix},$$

$$\mathcal{N}_n^{[1]} = n^{-1}\sum_{i=1}^n \widetilde{\epsilon}_i\mathcal{E}_i^{[1]}, \qquad \mathcal{N}_n^{[2]} = n^{-1}\sum_{i=1}^n \widetilde{\epsilon}_i\mathcal{E}_i^{[2]},$$

$$\mathcal{N}_{1,n}^{[1]} = -n^{-1/2}\sum_{k=1}^m\sum_{l\neq k}\frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k-\omega_l)}\mathrm{E}\left[\mathcal{E}_i^{[1]}\rho_2(\eta_i)\zeta_{il}\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\vartheta_{2k}\}\right]\langle\Delta\psi_k, \psi_l\rangle + \widetilde{O}_p\{n^{-1/2}(h^2+\delta_n)\},$$

$$\mathcal{N}_{1,n}^{[2]} = -n^{-1/2}\sum_{k=1}^m\sum_{l\neq k}\frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k-\omega_l)}\mathrm{E}\left[\mathcal{E}_i^{[2]}\rho_2(\eta_i)\zeta_{il}\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\vartheta_{2k}\}\right]\langle\Delta\psi_k, \psi_l\rangle + \widetilde{O}_p\{n^{-1/2}(h^2+\delta_n)\}. \tag{S.15}$$

As defined in (A.4),

$$\mathcal{F}_i = \mathcal{E}_i^{[2]} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{E}_i^{[1]}.$$

Following (S.8), when $m = o(n^{\frac{1}{2a+2}})$, $mh^4 = o(n^{-1/2})$, and $m^{\frac{3}{2}-\frac{a}{2}-b}h^2 = o(n^{-1/2})$,

$$\begin{aligned} & \sqrt{n}(\widehat{\mathbf{\Theta}}_Z - \mathbf{\Theta}_Z) \\ = \;& \mathcal{A}^{22}\sqrt{n}(\mathcal{N}_n^{[2]} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{N}_n^{[1]}) + \mathcal{A}^{22}\sqrt{n}(\mathcal{N}_{1,n}^{[2]} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{N}_{1,n}^{[1]}) + o_p(1) \\ = \;& \mathcal{A}^{22}n^{-\frac{1}{2}}\sum_{i=1}^n \epsilon_i\mathcal{F}_i + \mathcal{A}^{22}n^{-\frac{1}{2}}\sum_{i=1}^n(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{F}_i \\ & -\mathcal{A}^{22}\sum_{k=1}^m\sum_{l\neq k}\frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k-\omega_l)}\mathrm{E}\left[\mathcal{F}_i\rho_2(\eta_i)\zeta_{il}\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_{i1})\vartheta_{2k}\}\right]\langle\Delta\psi_k, \psi_l\rangle + o_p(1) \\ \doteq \;& \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + o_p(1). \end{aligned} \tag{S.16}$$

For the first term $\mathcal{I}_1$, since $\{\epsilon_i\mathcal{F}_i, i = 1, \ldots, n\}$ are zero mean i.i.d random vectors, by the central limit theorem, $n^{-1/2}\sum_{i=1}^n \epsilon_i\mathcal{F}_i \to \mathrm{N}\{\mathbf{0}, \sigma^2\mathrm{E}(\mathcal{F}_{\mathrm{i}}\mathcal{F}_{\mathrm{i}}^{\mathrm{T}})\}$ in distribution, where $\sigma^2 = \mathrm{E}\epsilon_i^2$. As defined in (A.5)

$$\mathcal{K}_n = \sigma^2\mathcal{A}^{22}\mathrm{E}\mathcal{F}_i\mathcal{F}_i^{\mathrm{T}}\mathcal{A}^{22}, \quad \mathcal{K}^* = \lim_{n\to\infty}\mathcal{K}_n,$$

we have $\mathcal{I}_1 \to \mathrm{N}(0, \mathcal{K}^*)$ in distribution, where $\mathcal{K}_n$ and $\mathcal{K}^*$ have bounded eigenvalues (Assumption (C2.8)). For $\mathcal{I}_2$, since $\mathrm{E}\mathcal{F}_i\mathcal{F}_i^{\mathrm{T}}$ has bounded eigenvalues, we have $\mathrm{E}\|\mathcal{F}_i\|_2^2 = \sum_{r=1}^{d+d_1} \mathrm{E}\mathcal{F}_{ij}^2 = O(1)$, where $\mathcal{F}_{ij}$ is the j-th element of $\mathcal{F}_i$. By Lemma 1, $\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)^2 = O(m^{1-a-2b})$, hence

$$|\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{F}_{ij}| \leq \left\{\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)^2\right\}^{1/2} \left(\mathrm{E}\mathcal{F}_{ij}^2\right)^{1/2} = O(m^{\frac{1}{2}-\frac{a}{2}-b}),$$

thus $\mathcal{I}_2 = O_p(n^{\frac{1}{2}} m^{\frac{1}{2}-\frac{a}{2}-b}) = O_p(1)$ when $m \gtrsim n^{\frac{1}{a+2b-1}}$.

It is easy to see that $\mathcal{I}_3 = \mathcal{A}^{22} n^{-1/2} \sum_{i=1}^n W_i$, where $W_i = \sum_{k=1}^m \sum_{l\neq k} \frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k - \omega_l)} \mathrm{E}\big[\mathcal{F}\rho_2(\eta)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\big]\xi_{ik}\xi_{il}$. $W_i, i = 1, \ldots, n$ are zero mean i.i.d random vectors. Denote $\boldsymbol{\Omega} = \mathrm{cov}(W_i)$. For any unit norm vector $\boldsymbol{a} \in \mathbb{R}^{d+d_1}$,

$$\begin{aligned}
\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{a} &= \mathrm{E}\left(\sum_{k=1}^m \sum_{l\neq k} \frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k - \omega_l)} \mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{F}\rho_2(\eta)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]\xi_{ik}\xi_{il}\right)^2 \\
&= \sum_{k=1}^m \sum_{l\neq k} \frac{\omega_l}{\omega_k(\omega_k - \omega_l)^2}(\omega_k - \omega_l)^{-2}\left(\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{F}\rho_2(\eta)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]\right)^2 \mathrm{E}\xi_{ik}^2\xi_{il}^2 \\
&\quad - \sum_{k=1}^m \sum_{\substack{l=1 \\ l\neq k}}^m (\omega_k - \omega_l)^{-2}\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{F}\rho_2(\eta)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right] \\
&\quad \times \mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{F}\rho_2(\eta)\{\vartheta_{1l} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2l}\}\zeta_k\right]\mathrm{E}\xi_{ik}^2\xi_{il}^2 \\
&\lesssim \sum_{k=1}^m \sum_{l\neq k} (\omega_k - \omega_l)^{-2}\omega_l^2(\boldsymbol{a}^{\mathrm{T}}\mathrm{E}\mathcal{F}\mathcal{F}^{\mathrm{T}}\boldsymbol{a})\mathrm{E}\left[\rho_2^2(\eta)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}^2\zeta_l^2\right] \\
&\quad + (\boldsymbol{a}^{\mathrm{T}}\mathrm{E}\mathcal{F}\mathcal{F}^{\mathrm{T}}\boldsymbol{a}) \sum_{k=1}^m \sum_{\substack{l=1 \\ l\neq k}}^m (\omega_k - \omega_l)^{-2}\omega_k\omega_l \left(\mathrm{E}\left[\rho_2^2(\eta)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}^2\zeta_l^2\right]\right)^{\frac{1}{2}} \\
&\quad \times \left(\mathrm{E}\left[\rho_2^2(\eta)\{\vartheta_{1l} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2l}\}^2\zeta_k^2\right]\right)^{\frac{1}{2}} \\
&= O\left\{\sum_{k=1}^m k^{-a-2b} \sum_{l\neq k} (\omega_k - \omega_l)^{-2}\omega_l^2 + \sum_{k=1}^m \omega_k k^{-\frac{a}{2}-b} \sum_{l\neq k} (\omega_k - \omega_l)^{-2}\omega_l l^{-\frac{a}{2}-b}\right\} \\
&= O(1) \qquad \text{(By Lemma 3).}
\end{aligned}$$

Therefore, $\boldsymbol{\Omega}$ has bounded eigenvalues. By CLT, we have $\mathcal{I}_3 \to \mathrm{N}(0, \mathcal{K}_1^*)$ in distribution, where

$$\mathcal{K}_{1,n} = \mathrm{cov}(\mathcal{I}_3) = \mathcal{A}^{22}\boldsymbol{\Omega}\mathcal{A}^{22}, \quad \mathcal{K}_1^* = \lim_{n\to\infty} \mathcal{K}_{1,n},$$

as defined in (A.6). Since $\mathcal{I}_3$ is a function of $\{X_i(\cdot), \mathbf{Z}_i\}_{i=1}^n$ and $\mathcal{I}_1$ is a linear function of ϵ_i's which satisfy $\mathrm{E}(\epsilon_i|X_i, \mathbf{Z}_i) = 0$, one can easily see $\mathcal{I}_1$ and $\mathcal{I}_3$ are uncorrelated and and therefore asymptotically independent.

Combining above results, when m satisfies $m \gtrsim n^{\frac{1}{a+2b-1}}$, $m = o(n^{\frac{1}{2a+2}})$, and $mh^4 = o(n^{-1/2})$, $\widehat{\mathbf{\Theta}}_Z - \mathbf{\Theta}_Z = O_p(n^{-\frac{1}{2}})$. If we further have $m^{-1}n^{\frac{1}{a+2b-1}} = o(1)$,

$$\sqrt{n}(\widehat{\mathbf{\Theta}}_Z - \mathbf{\Theta}_Z) = \mathcal{I}_1 + \mathcal{I}_3 + o_p(1) \to \mathrm{N}(0, \mathcal{K}^* + \mathcal{K}_1^*) \quad \text{in distribution.}$$

S.3 Proof of Theorem 2

From equation (S.4), we can deduce that

$$\begin{aligned}
&\widehat{\mathcal{S}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) \\
&= -\left[\mathcal{V}_1^{-1}(\mathbf{z};\boldsymbol{\theta})\{\mathcal{V}_2(\mathbf{z};\boldsymbol{\theta})\}\right]^{\mathrm{T}}\left(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\right) \times \{1 + o_p(1)\} + \mathcal{V}_1^{-1}(\mathbf{z},\boldsymbol{\theta})\sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\widetilde{\epsilon}_i \\
&\quad -n^{-1/2}\mathcal{V}_1^{-1}(\mathbf{z};\boldsymbol{\theta})\sum_{k=1}^m\sum_{l\neq k}\frac{\omega_l^{1/2}}{\omega_k^{-1/2}(\omega_k - \omega_l)}\mathrm{E}\left[\rho_2(\eta_i)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\{\vartheta_{1k} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\vartheta_{2k}\}\zeta_{il}\right]\langle\Delta\psi_k, \psi_l\rangle \\
&\quad +\frac{\sigma_K^2}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})h^2 + O_p\{h^4 + n^{-1/2}(\delta_n + m^{\frac{1-a-2b}{2}})\},
\end{aligned} \tag{S.17}$$

for all $\mathbf{z} \in \mathcal{D}$.

By Theorem 1 and Lemma 7, we can show that

$$\left(\mathcal{V}_1^{-1}\mathcal{V}_2^{\mathrm{T}}\right)(\mathbf{z},\boldsymbol{\theta})\left(\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\right) = O_p(n^{-1/2}m^{1/2}).$$

By Lemmas 1 and 4,

$$\mathcal{V}_1^{-1}(\mathbf{z},\boldsymbol{\theta})\sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)(\widetilde{\epsilon}_i - \epsilon_i) = O_p(m^{\frac{1}{2}-\frac{a}{2}-b}).$$

By Lemma 5, the third term on the right hand of (S.17) is of order $O_p(n^{-1/2})$. Hence,

$$\widehat{\mathcal{S}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) = \mathcal{V}_1^{-1}(\mathbf{z},\boldsymbol{\theta})\sum_{i=1}^n w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\epsilon_i + \frac{\sigma_K^2}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})h^2 + O_p(h^4 + m^{\frac{1}{2}-\frac{a}{2}-b} + n^{-\frac{1}{2}}m^{\frac{1}{2}}).$$

When m and h satisfies $m \gtrsim n^{\frac{1}{a+2b-1}}$, $m = o(n^{\frac{1}{2a+2}})$, and $mh = o(1)$, we can obtain the asymptotic normality for $\widehat{\mathcal{S}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})$ using CLT.

S.4 Proof of Theorem 3 and its Corollaries

Assume that $\{x(\cdot), \mathbf{z}\}$ is a new observation. Denote $x_j = \int x(t)\psi_j(t)dt$, $\widehat{x}_j = \int x(t)\widehat{\psi}_j(t)dt$, $j = 1, 2, \ldots$. We assume the new observation is fixed, and $|x_j| \lesssim j^{-a/2}$.

LEMMA **7** *Denote* $\boldsymbol{\Theta}_X = (\boldsymbol{\vartheta}_1^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}})^{\mathrm{T}}$, *and* $\mathcal{Q} = \mathrm{E}\{(\widehat{\boldsymbol{\Theta}}_X - \boldsymbol{\Theta}_X)(\widehat{\boldsymbol{\Theta}}_X - \boldsymbol{\Theta}_X)^{\mathrm{T}}\}$. *Under assumptions C1, C2 and* $m = o(n^{\frac{1}{2a+2}})$, *there exists a constant* $C > 0$ *such that*

$$\boldsymbol{a}^{\mathrm{T}}\mathcal{Q}\boldsymbol{a} \leq C\{(n^{-1} + m^{1-a-2b} + m^2h^8 + m^{3-a-2b}h^4)\|\boldsymbol{a}\|^2\},$$

for any constant vector $\boldsymbol{a} \in \mathbb{R}^{2m}$. *Specifically, when* $m \gtrsim n^{\frac{1}{a+2b-1}}$ *and* $mh^4 = o(n^{-1/2})$, $n\mathcal{Q}$ *has bounded eigenvalues.*

Proof: From (S.8), we have

$$\begin{aligned}
\widehat{\boldsymbol{\Theta}}_X - \boldsymbol{\Theta}_X &= \mathcal{A}^{11}(\mathcal{N}_n^{[1]} + \mathcal{N}_{1,n}^{[1]}) + \mathcal{A}^{12}(\mathcal{N}_n^{[2]} + \mathcal{N}_{1,n}^{[2]}) + \widetilde{O}_p(n^{-1/2}\delta_n m + mh^4 + m^{\frac{3}{2}-\frac{a}{2}-b}h^2) \\
&= \mathcal{G}_{11}^{-1}(\mathcal{N}_n^{[1]} + \mathcal{N}_{1,n}^{[1]}) - \mathcal{G}_{11}^{-1}\mathcal{A}_{12}(\widehat{\boldsymbol{\Theta}}_Z - \boldsymbol{\Theta}_Z) + \widetilde{O}_p(n^{-1/2}\delta_n m + mh^4 + m^{\frac{3}{2}-\frac{a}{2}-b}h^2).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathcal{Q} \leq\ & 4\mathcal{G}_{11}^{-1}\mathrm{E}\mathcal{N}_n^{[1]}(\mathcal{N}_n^{[1]})^{\mathrm{T}}\mathcal{G}_{11}^{-1} + 4\mathcal{G}_{11}^{-1}\mathrm{E}\mathcal{N}_{1,n}^{[1]}(\mathcal{N}_{1,n}^{[1]})^{\mathrm{T}}\mathcal{G}_{11}^{-1} + 4\mathcal{G}_{11}^{-1}\mathcal{A}_{12}\mathrm{E}(\widehat{\boldsymbol{\Theta}}_Z - \boldsymbol{\Theta}_Z)(\widehat{\boldsymbol{\Theta}}_Z - \boldsymbol{\Theta}_Z)^{\mathrm{T}}\mathcal{A}_{21}\mathcal{G}_{11}^{-1} \\
& + \widetilde{O}(n^{-1}\delta_n^2 m^2 + m^2h^8 + m^{3-a-2b}h^4).
\end{aligned}$$

For the first term on right hand of the above equation,

$$\begin{aligned}
\mathrm{E}\mathcal{N}_n^{[1]}\mathcal{N}_n^{[1]\mathrm{T}} &= \frac{1}{n^2}\mathrm{E}\left(\sum_{i=1}^n \widetilde{\epsilon}_i\mathcal{E}_i^{[1]}\right)\left(\sum_{i=1}^n \widetilde{\epsilon}_i\mathcal{E}_i^{[1]}\right)^{\mathrm{T}} \\
&= \frac{1}{n^2}\mathrm{E}\left\{\sum_{i=1}^n \epsilon_i\mathcal{E}_i^{[1]} + \sum_{i=1}^n(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{E}_i^{[1]}\right\}\left\{\sum_{i=1}^n \epsilon_i\mathcal{E}_i^{[1]} + \sum_{i=1}^n(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{E}_i^{[1]}\right\}^{\mathrm{T}} \\
&\leq \frac{2}{n^2}\mathrm{E}\left(\sum_{i=1}^n \epsilon_i\mathcal{E}_i^{[1]}\right)\left(\sum_{i=1}^n \epsilon_i\mathcal{E}_i^{[1]}\right)^{\mathrm{T}} + \frac{2}{n^2}\mathrm{E}\left\{\sum_{i=1}^n(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{E}_i^{[1]}\right\}\left\{\sum_{i=1}^n(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{E}_i^{[1]}\right\}^{\mathrm{T}} \\
&\leq \frac{2}{n}\sigma^2\mathrm{E}\mathcal{E}_i^{[1]}(\mathcal{E}_i^{[1]})^{\mathrm{T}} + \frac{2}{n}\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)^2\mathcal{E}_i^{[1]}(\mathcal{E}_i^{[1]})^{\mathrm{T}} + 2\left\{\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{E}_i^{[1]}\right\}\left\{\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)\mathcal{E}_i^{[1]}\right\}^{\mathrm{T}}.
\end{aligned}$$

Since $\mathcal{G}_{11}$, $\mathrm{E}\mathcal{E}_i^{[1]}\mathcal{E}_i^{[1]\mathrm{T}}$ has bounded eigenvalues, for any constant vector $\boldsymbol{a} \in \mathbb{R}^{2m}$

$$\begin{aligned}
& \boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathrm{E}\mathcal{N}_n^{[1]}(\mathcal{N}_n^{[1]})^{\mathrm{T}}\mathcal{G}_{11}^{-1}\boldsymbol{a} \\
\leq\ & 2n^{-1}\sigma^2\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathrm{E}\mathcal{E}_i^{[1]}(\mathcal{E}_i^{[1]})^{\mathrm{T}}\mathcal{G}_{11}^{-1}\boldsymbol{a} + 2n^{-1}\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)^2(\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}_i^{[1]})^2 + 2\left\{\mathrm{E}(\widetilde{\epsilon}_i - \epsilon_i)(\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}_i^{[1]})\right\}^2 \\
=\ & O\{(n^{-1} + m^{1-a-2b})\|\boldsymbol{a}\|^2\}.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
& \boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathrm{E}\mathcal{N}_{1,n}^{[1]}(\mathcal{N}_{1,n}^{[1]})^{\mathrm{T}}\mathcal{G}_{11}^{-1}\boldsymbol{a} \\
= \; & n^{-1}\mathrm{E}\left(\sum_{k=1}^{m}\sum_{l\neq k}\frac{\omega_l^{1/2}}{\omega_k^{1/2}(\omega_k-\omega_l)}\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}^{[1]}\rho_2(\eta)\{\vartheta_{1k}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]\xi_{ik}\xi_{il}\right)^2 \\
= \; & n^{-1}\sum_{k=1}^{m}\sum_{l\neq k}\frac{\omega_l}{\omega_k(\omega_k-\omega_l)^2}\left(\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}^{[1]}\rho_2(\eta)\{\vartheta_{1k}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]\right)^2\mathrm{E}\xi_{ik}^2\xi_{il}^2 \\
& -n^{-1}\sum_{k=1}^{m}\sum_{l\neq k}(\omega_k-\omega_l)^{-2}\mathrm{E}\xi_{ik}^2\xi_{il}^2\left(\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}^{[1]}\rho_2(\eta)\{\vartheta_{1k}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]\right) \\
& \qquad\qquad \times\left(\mathrm{E}\left[\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}^{[1]}\rho_2(\eta)\{\vartheta_{1l}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2l}\}\zeta_k\right]\right) \\
\leq \; & n^{-1}\mathrm{E}\left(\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}^{[1]}\right)^2\sum_{k=1}^{m}\sum_{l\neq k}\frac{\omega_l}{\omega_k(\omega_k-\omega_l)^2}\mathrm{E}\left[\rho_2(\eta)\{\vartheta_{1k}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]^2\mathrm{E}\xi_{ik}^2\xi_{il}^2 \\
& +n^{-1}\mathrm{E}\left(\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{E}^{[1]}\right)^2\sum_{k=1}^{m}\sum_{l\neq k}(\omega_k-\omega_l)^{-2}\mathrm{E}\xi_{ik}^2\xi_{il}^2\left(\mathrm{E}\left[\rho_2(\eta)\{\vartheta_{1k}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2k}\}\zeta_l\right]^2\right)^{1/2} \\
& \qquad\qquad \times\left(\mathrm{E}\left[\rho_2(\eta)\{\vartheta_{1l}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_1)\vartheta_{2l}\}\zeta_k\right]^2\right)^{1/2} \\
= \; & O\left\{n^{-1}\sum_{k=1}^{m}k^{-a-2b}\sum_{l\neq k}(\omega_k-\omega_l)^{-2}\omega_l^2+n^{-1}\sum_{k=1}^{m}\omega_k k^{-\frac{a}{2}-b}\sum_{l\neq k}(\omega_k-\omega_l)^{-2}\omega_l l^{-\frac{a}{2}-b}\right\}\|\boldsymbol{a}\|^2 \\
= \; & O(n^{-1}\|\boldsymbol{a}\|^2) \quad \text{(By Lemma 3).}
\end{aligned}
$$

By the asymptotic expansion (S.16),

$$\widehat{\boldsymbol{\Theta}}_Z-\boldsymbol{\Theta}_Z=n^{-1/2}\mathcal{I}_1+n^{-1/2}\mathcal{I}_2+n^{-1/2}\mathcal{I}_3+\widetilde{O}_p(n^{-1/2}\delta_n m+mh^4+m^{\frac{3}{2}-\frac{a}{2}-b}h^2).$$

Under Assumption (C2.8), $\mathcal{G}_{11}^{-1}$ and $\mathcal{A}_{12}\mathcal{A}_{21}$ have bounded eigenvalues. It is then easy to see that $\mathcal{G}_{11}^{-1}\mathcal{A}_{12}\mathrm{E}\mathcal{I}_l\mathcal{I}_l^{\mathrm{T}}\mathcal{A}_{21}\mathcal{G}_{11}^{-1}, l=1,3$ has bounded eigenvalues. Using similar arguments, $\boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{A}_{12}\mathrm{E}\mathcal{I}_2\mathcal{I}_2^{\mathrm{T}}\mathcal{A}_{21}\mathcal{G}_{11}^{-1}\boldsymbol{a}=O(m^{1-a-2b}\|\boldsymbol{a}\|^2)$. Therefore,

$$
\begin{aligned}
& \boldsymbol{a}^{\mathrm{T}}\mathcal{G}_{11}^{-1}\mathcal{A}_{12}\mathrm{E}\left(\widehat{\boldsymbol{\Theta}}_Z-\boldsymbol{\Theta}_Z\right)\left(\widehat{\boldsymbol{\Theta}}_Z-\boldsymbol{\Theta}_Z\right)^{\mathrm{T}}\mathcal{A}_{21}\mathcal{G}_{11}^{-1}\boldsymbol{a} \\
& =O\{(n^{-1}+m^{1-a-2b}+n^{-1}\delta_n^2m^2+m^2h^8+m^{3-a-2b}h^4)\|\boldsymbol{a}\|^2\}.
\end{aligned}
$$

Combining the above results, when $m=o(n^{\frac{1}{2a+2}})$,

$$\boldsymbol{a}^{\mathrm{T}}\mathcal{Q}\boldsymbol{a}=O\{(n^{-1}+m^{1-a-2b}+m^2h^8+m^{3-a-2b}h^4)\|\boldsymbol{a}\|^2\}.$$

LEMMA **8** *When $m = o(n^{\frac{1}{2a+2}})$, for $j = 1, \ldots, m$, we have*

$$\mathrm{E}(\widehat{x}_j - x_j)^2 = O\left(n^{-1} j^{2-a}\right), \quad \mathrm{E}\Big\{ \sum_{j=1}^{m} \alpha_{1j}(\widehat{x}_j - x_j) \Big\}^2 = O(n^{-1}).$$

Proof: This result comes from Cai and Hall (2006), equations (5.14) and (5.15).

Proof of Theorem 3:

Now we investigate the performance of $\widehat{T}$, as noted in (16).

$$\begin{aligned} \widehat{T} - T &= \sum_{j=1}^{m} \{\widehat{\alpha}_{1j} + \widehat{\alpha}_{2j}\widehat{\mathcal{S}}(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}_1)\}\widehat{x}_j + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{z} - \sum_{j=1}^{\infty} \{\alpha_{1j} + \alpha_{2j}\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\}x_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z} \\ &= \sum_{\ell=1}^{9} \mathcal{T}_\ell, \end{aligned} \tag{S.18}$$

where

$$\mathcal{T}_1 = \sum_{j=1}^{m} x_j \left\{ \widehat{\alpha}_{1j} - \alpha_{1j} + (\widehat{\alpha}_{2j} - \alpha_{2j})\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \right\}, \quad \mathcal{T}_2 = \left\{ \widehat{\mathcal{S}}(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}_1) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \right\} \sum_{j=1}^{m} x_j \left(\widehat{\alpha}_{2j} - \alpha_{2j}\right),$$

$$\mathcal{T}_3 = \left\{ \widehat{\mathcal{S}}(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}_1) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \right\} \sum_{j=1}^{m} x_j \alpha_{2j}, \quad \mathcal{T}_4 = \sum_{j=1}^{m} (\widehat{x}_j - x_j)\{\widehat{\alpha}_{1j} - \alpha_{1j} + (\widehat{\alpha}_{2j} - \alpha_{2j})\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\},$$

$$\mathcal{T}_5 = \left\{ \widehat{\mathcal{S}}(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}_1) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \right\} \sum_{j=1}^{m} (\widehat{x}_j - x_j)\left(\widehat{\alpha}_{2j} - \alpha_{2j}\right),$$

$$\mathcal{T}_6 = \left\{ \widehat{\mathcal{S}}(\widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}_1) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \right\} \sum_{j=1}^{m} \alpha_{2j}(\widehat{x}_j - x_j), \quad \mathcal{T}_7 = \sum_{j=1}^{m} (\widehat{x}_j - x_j)\{\alpha_{1j} + \alpha_{2j}\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\},$$

$$\mathcal{T}_8 = -\sum_{j=m+1}^{\infty} x_j \{\alpha_{1j} + \alpha_{2j}\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\}, \text{ and } \quad \mathcal{T}_9 = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{z}.$$

$E(\widehat{T} - T)^2$ involves the expectations of the squared and cross terms of $\mathcal{T}_\ell$. For $m = o(n^{\frac{1}{2a+2}})$, we have $\mathrm{E}(\mathcal{T}_8^2) = O(m^{2-a-2b})$, and $\mathrm{E}(\mathcal{T}_9^2) = O(n^{-1} + m^{1-a-2b} + m^2 h^8 + m^{3-a-2b}h^4)$. In what follows we deduce orders for the rest of the squared terms.

Denote

$$\mathcal{X} = \left\{ x_1, \ldots, x_m, x_1\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1), \ldots, x_m\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \right\}^{\mathrm{T}}, \mathcal{W} = \mathrm{diag}\left(\omega_1^{-1/2}, \ldots, \omega_m^{-1/2}, \omega_1^{-1/2}, \ldots, \omega_m^{-1/2} \right).$$

By Lemma 7,

$$\mathrm{E}(\mathcal{T}_1^2) = \mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{Q}\mathcal{W}\mathcal{X} = O\left(n^{-1}m + m^{2-a-2b} + m^3 h^8 + m^{4-a-2b}h^4 \right).$$

Using similar derivations as those for $\mathrm{E}(\mathcal{T}_1^2)$, we obtain

$$\mathrm{E}\Big\{\sum_{j=1}^{m} x_j(\widehat{\alpha}_{2j}-\alpha_{2j})\Big\}^2 = O\left(n^{-1}m + m^{2-a-2b} + m^3h^8 + m^{4-a-2b}h^4\right).$$

From proof of Theorem 2 and Lemma 7, we have

$$\{\widehat{\mathcal{S}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\}^2 \;=\; O_p(n^{-1}m + m^{2-a-2b} + m^3h^8 + n^{-1}h^{-1} + h^4),$$

which leads to,

$$\begin{aligned}\mathrm{E}(\mathcal{T}_2^2) &= o\left(n^{-1}m + m^{2-a-2b} + m^3h^8 + m^{4-a-2b}h^4\right), \text{ and}\\ \mathrm{E}(\mathcal{T}_3^2) &= O(n^{-1}m + m^{2-a-2b} + m^3h^8 + n^{-1}h^{-1} + h^4).\end{aligned}$$

By Lemma 7, we obtain

$$\begin{aligned}& E\left[\sum_{j=1}^{m}\{\widehat{\alpha}_{1j}-\alpha_{1j}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)(\widehat{\alpha}_{2j}-\alpha_{2j})\}^2\right]\left\{\mathbf{1}_m^{\mathrm{T}} \quad \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\mathbf{1}_m^{\mathrm{T}}\right\}\mathcal{W}\mathcal{Q}\mathcal{W}\begin{Bmatrix}\mathbf{1}_m\\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\mathbf{1}_m\end{Bmatrix}\\ =\;& O\{m^{a+1}(n^{-1}+m^{1-a-2b}+m^2h^8+m^{3-a-2b}h^4)\}\end{aligned}$$

Based on Lemma 8,

$$\begin{aligned}\mathrm{E}(\mathcal{T}_4^2) &\le \mathrm{E}\Big\{\sum_{j=1}^{m}(\widehat{x}_j-x_j)^2\Big\}\mathrm{E}\Big[\sum_{j=1}^{m}\{\widehat{\alpha}_{1j}-\alpha_{1j}+\mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)(\widehat{\alpha}_{2j}-\alpha_{2j})\}^2\Big]\\ &= o(n^{-1}m + m^{2-a-2b} + m^3h^8 + m^{4-a-2b}h^4).\end{aligned}$$

For $\mathcal{T}_5$, using Lemmas 7 and 8,

$$\mathrm{E}\Big\{\sum_{j=1}^{m}(\widehat{x}_j-x_j)(\widehat{\alpha}_{2j}-\alpha_{2j})\Big\}^2 \le \mathrm{E}\left\{\sum_{j=1}^{m}(\widehat{x}_j-x_j)^2\right\}\mathrm{E}\left\{\sum_{j=1}^{m}(\widehat{\alpha}_{2j}-\alpha_{2j})^2\right\},$$

and it leads to that, $\mathrm{E}(\mathcal{T}_5^2) = o(n^{-1}m + m^{2-a-2b} + m^3h^8 + m^{4-a-2b}h^4)$. Similarly, based on result of Lemma 8, $\mathrm{E}(\mathcal{T}_6^2) = o(n^{-1}m + m^{2-a-2b} + m^3h^8 + n^{-1}h^{-1} + h^4)$, and $\mathrm{E}(\mathcal{T}_7^2) = O(n^{-1})$.

Combining the results above, we obtain that when $m = o(n^{\frac{1}{2a+2}})$,

$$\mathrm{E}(\widehat{T}-T)^2 = O(n^{-1}m + m^{2-a-2b} + n^{-1}h^{-1} + h^4 + m^3h^8).$$

Note that under Assumption (C2.5), the order of $n^{-1}h^{-1}+h^4$ is between $O(n^{-\frac{4}{5}})$ and $O(n^{-\frac{2}{3}})$. This leads to the outcomes given in (17) and Corollary 1 when $h \asymp n^{-1/5}$.

Proof of Corollary 2:

Under optimal tuning parameters $h \asymp n^{-1/5}$ and $m \asymp n^{\frac{1}{a+2b-1}}$, $\mathcal{T}_3$ is the dominating term in the decomposition of $\widehat{T} - T$ in (S.18). By Theorems 1 and 2

$$\begin{aligned}\widehat{T} - T &= \left\{\widehat{\mathcal{S}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) - \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)\right\} \sum_{j=1}^{m} x_j \alpha_{2j} + o_p\{h^2 + (nh)^{-1/2}\} \\ &= \mathcal{V}_1^{-1}(\mathbf{z}, \boldsymbol{\theta}) \int x(t) \mathscr{A}_2(t) dt \sum_{i=1}^{n} w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\epsilon_i + \frac{\sigma_K^2}{2} h^2 \mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \int x(t) \mathscr{A}_2(t) dt \\ &\qquad + o_p\{h^2 + (nh)^{-1/2}\}.\end{aligned}$$

By (16),

$$\begin{aligned}\widehat{\mu}_Y(x, \mathbf{z}) - \mu_Y(x, \mathbf{z}) &= (g^{-1})'(T) \mathcal{V}_1^{-1}(\mathbf{z}, \boldsymbol{\theta}) \int x(t) \mathscr{A}_2(t) dt \sum_{i=1}^{n} w_i(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1)(\boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\zeta}_i)\epsilon_i \qquad \text{(S.19)} \\ &\quad + \frac{\sigma_K^2}{2} h^2 (g^{-1})'(T) \mathcal{S}^{(2)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}_1) \int x(t) \mathscr{A}_2(t) dt + o_p\{h^2 + (nh)^{-1/2}\}.\end{aligned}$$

The asymptotic normal distribution in Corollary 2 follows directly from the asymptotic expansion (S.19) and the Central Limit Theorem.

S.5 Additional numerical results

In this subsection, we report additional numerical results as described in Section 5.1. They include two figures of confidence intervals and an outcome table of the estimated $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ for the corn-yield data set, constructed using the analysis outcomes with $m = 20, 21, \ldots, 25$.

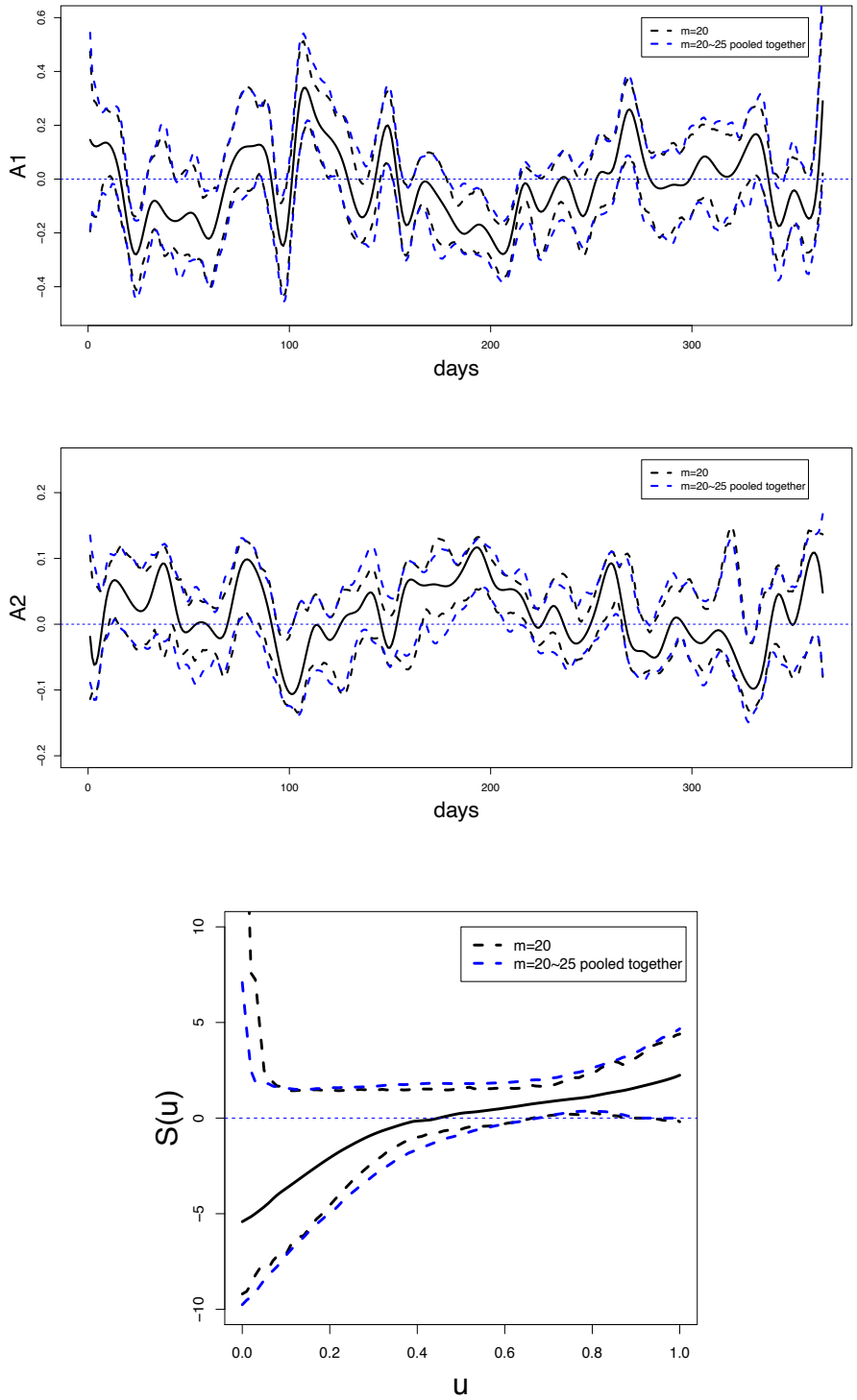
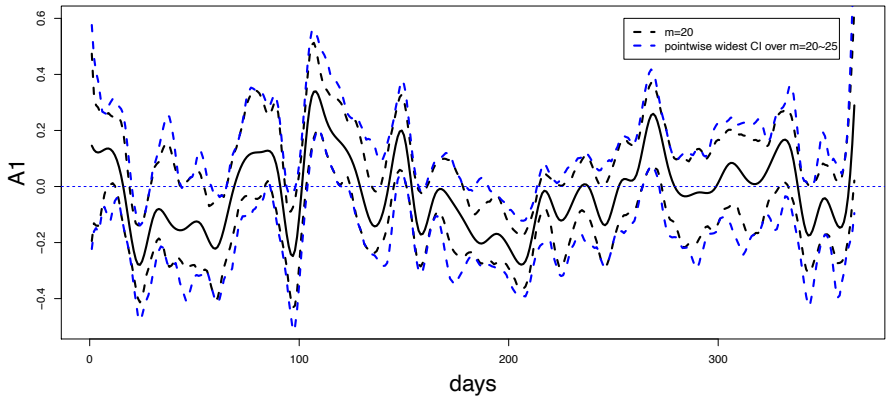
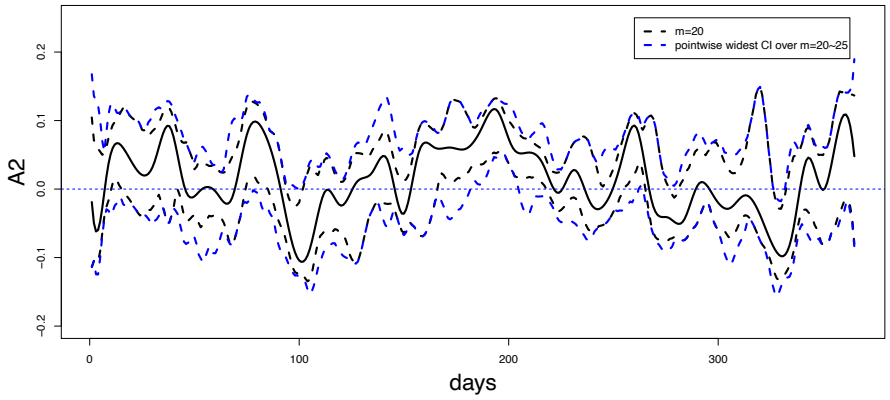


Figure S.1: Estimated nonparametric functions and two types of confidence intervals for the corn-yield data set. From top to bottom, the three plots correspond to $\mathscr{A}_1(\cdot)$, $\mathscr{A}_2(\cdot)$ and $\mathcal{S}(\cdot)$, respectively. In each plot, we provide the estimates (solid lines) and the corresponding 95% pointwise confidence intervals obtained with $m = 20$, and with all bootstrap outcomes from different m pooled together (dashed lines).





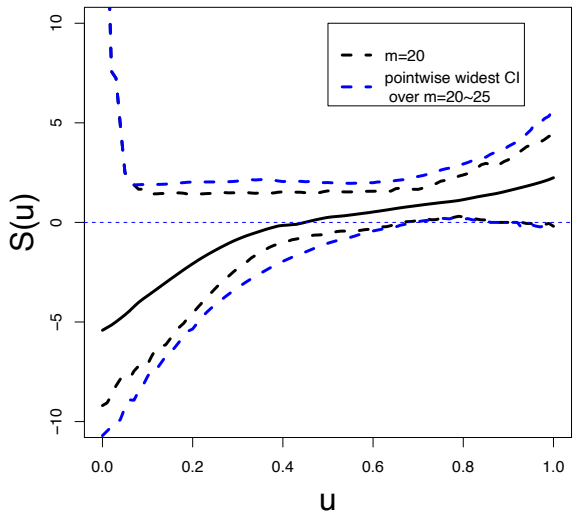


Figure S.2: Estimated nonparametric functions and two types of confidence intervals for the corn-yield data set. As in Figure S.1 but with the wider intervals constructed by taking the minimum/maximum of the lower/upper endponts of the 95% pointwise confidence intervals with $m = 20, 21, \ldots, 25$.

Table S.1: Mean (standard deviation) for the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ in bootstrap results with $m = 20, 21, \ldots, 25$ and pooled bootstrap results.

	m=20	m=21	m=22	m=23	m=24	m=25	Pooled
	Mean (Sd)	Mean (Sd)	Mean (Sd)	Mean (Sd)	Mean (Sd)	Mean (Sd)	Mean (Sd)
Estimated $\boldsymbol{\theta}$							
irrigated.eprop	0.99 (0.04)	0.99 (0.03)	0.99 (0.04)	0.99 (0.04)	0.99(0.03)	0.99 (0.02)	0.99 (0.03)
avgprcp	0.02 (0.11)	0.02 (0.10)	0.02 (0.09)	0.00 (0.13)	0.00 (0.11)	0.00 (0.11)	0.01 (0.11)
Estimated $\boldsymbol{\beta}$							
intercept	42.85 (20.63)	49.30 (20.81)	51.12 (20.17)	52.93 (20.59)	53.80 (20.81)	53.16 (20.24)	50.53 (20.87)
irrigated.eprop	172.07 (13.63)	171.90 (13.83)	171.90 (13.67)	169.82 (13.40)	170.98 (14.26)	169.54 (14.36)	171.04 (13.89)
avgprcp	19.38 (4.23)	18.97 (4.22)	18.76 (4.29)	17.86 (4.29)	18.39 (4.46)	17.96 (4.55)	18.55 (4.37)
irrigated.eprop*avgprcp	-33.25 (6.36)	-32.86 (6.36)	-32.79 (6.33)	-31.43 (6.29)	-31.96 (6.77)	-31.13 (6.92)	-32.24 (6.55)