Varying-Coefficient Functional Linear Regression

Yichao Wu* **, Jianqing Fan and Hans-Georg Müller**

NCSU, Princeton University, and UC-Davis

Abstract: Functional linear regression analysis aims to model regression relations which include a functional predictor. The analogue to the regression parameter vector or matrix in conventional multivariate or multiple-response linear regression models is a regression parameter function in one or two arguments. If in addition one has scalar predictors, as is often the case in applications to longitudinal studies, the question arises how to incorporate these into a functional regression model. We study a varying-coefficient approach where the scalar covariates are modeled as additional arguments of the regression parameter function. This extension of the functional linear regression model is analogous to the extension of conventional linear regression models to varying-coefficient models, and shares the advantages such as increased flexibility, however the details of this extension are more challenging in the functional case. Our methodology combines smoothing methods with regularization by truncation at a finite number of functional principal components. A practical version is developed and demonstrated to perform better than functional linear regression for longitudinal data. We investigate the asymptotic properties of varying-coefficient functional linear regression and establish consistency properties.

AMS 2000 subject classifications: Primary 62M20; secondary 60G15, 62G05.

Keywords and phrases: Asymptotics, eigenfunctions, functional data analysis, local polynomial smoothing, longitudinal data, varying-coefficient models.

Running title: Varying-coefficient Functional Linear Regression.

*Author for correspondence. Yichao Wu is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695. Jianqing Fan is Frederick L. Moore '18 Professor of Finance and Professor of Statistics, Princeton University, Princeton 08544. Hans-Georg Müller is Professor, Department of Statistics, University of California, Davis, CA 95616.

1. Introduction

Functional linear regression analysis is an extension of ordinary regression to the case where predictors include random functions and responses are scalars or functions. This methodology recently has met with increasing interest due to its inherent applicability in longitudinal data analysis and other areas of modern data analysis. For an excellent introduction, see Ramsay and Silverman (2005). Assuming that predictor process X possesses a square integrable trajectory (*i.e.*, $X \in L^2(\mathcal{S})$, where $\mathcal{S} \subset I\!R$), commonly considered functional linear regression models include

$$E(Y|X) = \mu_Y + \int_{\mathcal{S}} \beta(s)(X(s) - \mu_X(s))ds \tag{1.1}$$

with a scalar response $Y \in R$, and

$$E(Y(t)|X) = \mu_Y(t) + \int_{\mathcal{S}} \beta(s,t)(X(s) - \mu_X(s))ds \tag{1.2}$$

with a functional response $Y \in L^2(\mathcal{T})$ and $\mathcal{T}$ being a subset of the real line $I\!R$, where $\mu_X(s) = E(X(s))$, $s \in \mathcal{S}$ and $\mu_Y(t) = E(Y(t))$, $t \in \mathcal{T}$ (Ramsay and Dalzell, 1991). In analogy to the classical regression case, estimating equations for the regression function are based on minimizing the deviation

$$\beta^*(s,t) = \underset{\beta \in L_2(\mathcal{S}\times\mathcal{T})}{\operatorname{argmin}} E\left\{\int_{\mathcal{T}} (Y(t) - \mu_Y(t) - \int_{\mathcal{S}} \beta(s,t)[X(s) - \mu_X(s)]ds)^2 dt\right\},$$

and analogously for (1.1). To provide a regularized estimator, one approach is to expand $\beta(\cdot,\cdot)$ in terms of the eigenfunctions of the covariance functions of X and Y and to use an appropriately chosen finite number of the resulting functional principal component (FPC) scores of X as predictors. See, for example, Silverman (1996), Ramsay and Silverman (2002), Ramsay and Silverman (2005), Besse and Ramsay (1986), Ramsay and Dalzell (1991), Rice and Silverman (1991), James et al. (2000), Cuevas et al. (2002), Cardot et al. (2003), Hall and Horowitz (2007), Cai and Hall (2006), Cardot (2007), and many others.

Advances of modern technology enable us to collect massive amounts of data at fairly low cost. In such settings, one may observe scalar covariates in addition to functional predictor and response trajectories. For example, when predicting a response such as blood pressure from functional data, one may wish to utilize functional covariates such as

body mass index and also additional non-functional covariates Z such as age of a subject. It is often realistic to expect the regression relation to change as an additional covariate such as age varies. To broaden the applicability of functional linear regression models, we propose to generalize models (1.1) and (1.2) by allowing the slope function to depend on some additional scalar covariates Z. Previous work on varying-coefficient functional regression models, assuming the case of a scalar response and of continuously observed predictor processes, is due to Cardot and Sarda (2005), and recent investigations of the varying-coefficient approach include Fan et al. (2007) and Zhang et al. (2007).

For ease of presentation, we consider the case of a one-dimensional covariate $Z \in \mathcal{Z} \subset I\!R$, extending (1.1) and (1.2) to the varying-coefficient functional linear regression models

$$E(Y|X,Z) = \mu_{Y|Z} + \int_{\mathcal{S}} \beta(Z,s)(X(s) - \mu_{X|Z}(s))ds \tag{1.3}$$

and

$$E(Y(t)|X,Z) = \mu_{Y|Z}(t) + \int_{\mathcal{S}} \beta(Z,s,t)(X(s) - \mu_{X|Z}(s))ds \tag{1.4}$$

for scalar and functional responses, respectively, with corresponding characterizations for the regression parameter functions

$$\beta^*(z,s) = \underset{\beta(z,\cdot)\in L_2(\mathcal{S})}{\operatorname{argmin}} E\left\{(Y - \mu_{Y|Z} - \int_{\mathcal{S}} \beta(Z,s)[X(s) - \mu_{X|Z}(s)]ds)^2 \mid Z = z\right\},$$

$$\beta^*(z,s,t) = \underset{\beta(z,\cdot,\cdot)\in L_2(\mathcal{S}\times\mathcal{T})}{\operatorname{argmin}} E\left\{\int_{\mathcal{T}}(Y(t) - \mu_{Y|Z}(t) - \int_{\mathcal{S}} \beta(Z,s,t)[X(s) - \mu_{X|Z}(s)]ds)^2 dt \mid Z = z\right\}.$$

Here $\mu_{X|Z}(s)$ and $\mu_{Y|Z}(t)$ denote the conditional mean function of X and Y, given Z.

Intuitively, after observing a sample of n observations: $\{X_i, Y_i, Z_i\}_{i=1}^n$, the estimation of the varying slope functions can be achieved using kernel methods, as follows:

$$\tilde{\beta}^*(z,s) = \operatorname{argmin} \sum_{i=1}^n K_b(Z_i - z)\left[Y_i - \mu_{Y|Z_i} - \int_{\mathcal{S}} \beta(Z_i,s)[X_i(s) - \mu_{X|Z_i}(s)]ds\right]^2 \text{ and}$$

$$\tilde{\beta}^*(z,s,t) = \operatorname{argmin} \sum_{i=1}^n K_b(Z_i - z)\int_{\mathcal{T}}\left[Y_i(t) - \mu_{Y|Z_i}(t) - \int_{\mathcal{S}} \beta(Z_i,s,t)[X_i(s) - \mu_{X|Z_i}(s)]ds\right]^2 dt$$

for (1.3) and (1.4) respectively, where $K_b(z) = K(z/b)/b$ for a kernel function $K(\cdot)$ and a bandwidth $b > 0$. The necessary regularization of the slope function is conveniently achieved by truncating the Karhunen-Loève expansion of the covariance function for the predictor process (and the response process, if applicable). To avoid difficult technical

issues and enable straightforward and fast implementation, it is expedient to adopt the two-step estimation scheme proposed and extensively studied by Fan and Zhang (2000).

To this end, we first bin our observations according to the values taken by the additional covariate Z into a partition of $\mathcal{Z}$. On each bin, we obtain the sample covariance functions based on the observations within this bin. Assuming that the covariance functions of the predictor and response processes are continuous in z guarantees that these sample covariance functions converge to the corresponding true covariance functions evaluated at the bin centers as bin width goes to zero and sample size increases. This allows us to estimate the slope function at each bin center consistently using the technique studied in Yao et al. (2005b), providing initial raw estimates. Next, local linear smoothing (Fan and Gijbels, 1996) is applied to improve estimation efficiency, providing our final estimator of the slope function for any $z \in \mathcal{Z}$.

The remainder of the paper is organized as follows. In Section 2, we introduce basic notations and present our estimation scheme. Asymptotic consistency properties are reported in Section 3. Finite-sample implementation issues are discussed in Section 4, results of simulation studies in Section 5 and real data applications in Section 6, with conclusions in Section 7. Technical proofs and auxiliary results are given in an Appendix.

2. Varying coefficient functional linear regression for sparse and irregular data

To facilitate the presentation, we focus on the case of a functional response, which remains largely unexplored. The case with a scalar response can be handled similarly. We also emphasize the case of sparse and irregularly observed data with errors, due to its relevance in longitudinal studies. The motivation of the varying-coefficient functional regression models (1.3) and (1.4) is to borrow strength across subjects, while adequately reflecting the effects of the additional covariate. We impose the following smoothness conditions:

[A0] The conditional mean and covariance functions of the predictor and response processes depend on Z and are continuous in Z, *i.e.*, $\mu_{X,z}(s) = E(X(s)|Z = z)$, $\mu_{Y,z}(t) = E(Y(t)|Z = z)$, $G_{X,z}(s_1, s_2) = \mathrm{cov}(X(s_1), X(s_2)|Z = z)$,$G_{Y,z}(t_1, t_2) = \mathrm{cov}(Y(t_1), Y(t_2)|Z = z)$, and $C_{XY,z}(s, t) = \mathrm{cov}(X(s), Y(t)|Z = z)$ are continuous in z and their respective arguments and have continuous second-order partial deriva-

tives with respect to z.

Note that [A0] implies that the conditional mean and covariance functions of predictor and response processes do not change radically in a small neighborhood of $Z = z$. This facilitates the estimation of $\beta(z, s, t)$, using the two-step estimation scheme proposed by Fan and Zhang (2000). While there the additional covariate Z is assumed to take values on a grid, in our case Z is more generally assumed to be continuously distributed. In this case, we assume that the additional variable Z has a compact domain $\mathcal{Z}$ and its density $f_Z(z)$ is continuous and bounded away from both zero and infinity.

[A1] $\mathcal{Z}$ is compact, $f_Z(z) \in C^0$, $\underline{f_Z} = \inf_{z\in\mathcal{Z}} f_Z(z) > 0$ and $\bar{f_Z} = \sup_{z\in\mathcal{Z}} f_Z(z) < \infty$.

2.1. Representing predictor and response functions via functional principal components for sparse and irregular data

Suppose we have observations on n subjects. For each subject i, conditional on $Z_i = z_i$, the square integrable predictor trajectory X_i and response trajectory Y_i are unobservable realizations of the smooth random processes $(X, Y|Z = z_i)$, with unknown mean and covariance functions (Condition [A0]). The arguments of $X(\cdot)$ and $Y(\cdot)$ are usually referred to as time. Without loss of generality, their domains $\mathcal{S}$ and $\mathcal{T}$ are assumed to be finite and closed intervals. Adopting the general framework of functional data analysis, we assume for each z that there exist orthogonal expansions of the covariance functions $G_{X,z}(\cdot,\cdot)$ (resp. $G_{Y,z}(\cdot,\cdot)$) in the L_2 sense via the eigenfunctions $\psi_{z,m}$ (resp. $\phi_{z,k}$), with nonincreasing eigenvalues $\rho_{z,m}$ (resp. $\lambda_{z,k}$), *i.e.*, $G_{X,z}(s_1, s_2) = \sum_{m=1}^{\infty} \rho_{z,m}\psi_{z,m}(s_1)\psi_{z,m}(s_2)$, $G_{Y,z}(t_1, t_2) = \sum_{k=1}^{\infty} \lambda_{z,k}\phi_{z,k}(t_1)\phi_{z,k}(t_2)$.

Instead of observing the full predictor trajectory X_i and response trajectory Y_i, typical longitudinal data consist of noisy observations that are made at sparse and irregularly spaced locations or time points, providing sparse measurements of predictor and response trajectories that are contaminated with additional measurement errors (Staniswalis and Lee, 1998; Rice and Wu, 2001; Yao et al., 2005a,b). To adequately reflect the situation of sparse, irregular, and possibly subject-specific time points underlying these measurements, we assume that a random number L_i (resp. N_i) of measurements for X_i (resp. Y_i) is made, at times denoted by $S_{i1}, S_{i2}, \cdots, S_{iL_i}$ (resp. $T_{i1}, T_{i2}, \cdots, T_{iN_i}$).

Independent of any other random variables, the numbers of points sampled from each trajectory correspond to random variables L_i and N_i that are assumed to be *i.i.d.* as L and N (which may be correlated), respectively. For $1 \leq i \leq n$, $1 \leq l \leq L_i$, $1 \leq j \leq N_i$, denote U_{il} (resp. V_{ij}) to be the observation of the random trajectory X_i (resp. Y_i) made at a random time S_{il} (resp. T_{ij}), contaminated with measurement errors ε_{il} (resp. ϵ_{ij}). Here the random measurement errors ε_{il} and ϵ_{ij} are assumed to be *i.i.d.*, with mean zero and variances σ_X^2 and σ_Y^2, respectively. They are independent of all other random variables. The following two assumptions are made.

[A2] For each subject i, $L_i \overset{i.i.d.}{\sim} L$ (resp. $N_i \overset{i.i.d.}{\sim} N$) for a positive discrete valued random variable with $EL < \infty$ (resp. $EN < \infty$) and $P(L > 1) > 0$ (resp. $P(N > 1) > 0$).

[A3] For each subject i, observations on X_i (resp. Y_i) are independent of L_i (resp. N_i), *i.e.*, $\{(S_{il}, U_{il} : l \in \mathcal{L}_i)\}$ is independent of L_i for any $\mathcal{L}_i \subset \{1, 2, \cdots, L_i\}$ (resp. $\{(T_{ij}, V_{ij}) : j \in \mathcal{N}_i\}$ is independent of N_i for any $\mathcal{N}_i \subset \{1, 2, \cdots, N_i\}$).

It it surprising that under these "longitudinal assumptions" where the number of observations per subject is fixed and does not increase with sample size, one can nevertheless obtain asymptotic consistency results for the regression relation. This phenomenon was observed in Yao et al. (2005b) and is due to the fact that according to (2.3) the target regression function depends only on localized eigenfunctions and localized eigenvalues and cross-covariances of localized functional principal components. However, even though localized, these eigenfunctions and moments can be estimated from pooled data and do not require the fitting of individual trajectories. Even for the case of fitted trajectories, conditional approaches have been implemented successfully, allowing even to obtain reasonable derivative estimates from very sparse data (Liu and Müller, 2009).

Conditional on $Z_i = z$, the FPC scores of X_i and Y_i are $\zeta_{z,im} = \int [X_i(s) - \mu_{X,z}(s)]\psi_{z,m}(s)\, ds$ and $\xi_{z,ik} = \int [Y_i(s) - \mu_{Y,z}(s)]\phi_{z,k}(s)\, ds$, respectively. For all z, these FPC scores $\zeta_{z,im}$ satisfy $E\zeta_{z,im} = 0$, $\mathrm{corr}(\zeta_{z,im_1}, \zeta_{z,im_2}) = 0$ for any $m_1 \neq m_2$, and $\mathrm{var}(\zeta_{z,im}) = \rho_{z,m}$; analogously for $\xi_{z,ik}$. With these notations, using the Karhunen-Loève expansion as in Yao et al. (2005b), conditional on Z_i, the available measurements of the i-th predictor and response trajectories can be represented as

$$U_{il} \;=\; X_i(S_{il}) + \varepsilon_{il} = \mu_{X,Z_i}(S_{il}) + \sum_{m=1}^{\infty} \zeta_{Z_i,im}\psi_{Z_i,m}(S_{il}) + \varepsilon_{il}, \qquad 1 \leq l \leq L_i, \tag{2.1}$$

$$V_{ij} \;=\; Y_i(T_{ij}) + \epsilon_{ij} = \mu_{Y,Z_i}(T_{ij}) + \sum_{k=1}^{\infty} \xi_{Z_i,ik}\phi_{Z_i,k}(T_{ij}) + \epsilon_{ij}, \qquad 1 \le j \le N_i. \quad (2.2)$$

2.2. Estimation of the slope function

For estimation of the slope function, one standard approach is to expand it in terms of an orthonormal functional basis and to estimate the coefficients of this expansion to estimate the slope function in the non-varying model (1.2) (Yao et al., 2005b). As a result of the nonincreasing property of the eigenvalues of the covariance functions, such expansions of the slope function are often efficient and require only a few components for a good approximation. Truncation at a finite number of terms provides the necessary regularization. Departing from Yao et al. (2005b), we assume here that an additional covariate Z plays an important role and must be incorporated into the model, motivating (1.4). To make this model as flexible as possible, the conditional mean and covariance functions of the predictor and response processes are allowed to change smoothly with the value of the covariate Z (Assumption [A0]) which facilitates implementation and analysis of the two-step estimation scheme as in Fan and Zhang (2000).

Efficient implementation of the two-step estimation scheme begins by binning subjects according to the levels of the additional covariate Z_i, $i = 1, 2, \cdots, n$. For ease of presentation, we use equal-width bins, although in practice non-equidistant bins can occasionally be advantageous. Denoting the bin centers by $z^{(p)}, p = 1, 2, \cdots, P$ and the bin width by h, the p-th bin is $[z^{(p)} - \frac{h}{2}, z^{(p)} + \frac{h}{2})$ with $h = \frac{|\mathcal{Z}|}{P}$ where $|\mathcal{Z}|$ denotes the size of the domain of Z, $z^{(1)} - h/2 \equiv \inf\{z : z \in \mathcal{Z}\}$ and $z^{(P)} + h/2 \equiv \sup\{z : z \in \mathcal{Z}\}$ (note that the last bin is $[z^{(P)} - h/2, z^{(P)} + h/2]$). Let $\mathcal{N}_{z,h} = \left\{i : Z_i \in [z - \frac{h}{2}, z + \frac{h}{2})\right\}$ be the index set of those subjects falling into bin $[z - \frac{h}{2}, z + \frac{h}{2})$ and $n_{z,h} = \#\mathcal{N}_{z,h}$ the number of those subjects.

2.2.1. Raw estimates

For each bin $[z^{(p)} - \frac{h}{2}, z^{(p)} + \frac{h}{2})$, we use the Yao et al. (2005a) method to obtain our raw estimates $\tilde{\mu}_{X,z^{(p)}}(\cdot)$ and $\tilde{\mu}_{Y,z^{(p)}}(\cdot)$ of the conditional mean trajectories and the raw slope function estimate $\tilde{\beta}(z^{(p)}, s, t)$. The corresponding raw estimates of σ_X^2 and σ_Y^2 are denoted by $\tilde{\sigma}^2_{X,z^{(p)}}$ and $\tilde{\sigma}^2_{Y,z^{(p)}}$ for $p = 1, 2, \cdots, P$. For each $1 \le p \le P$, the local linear scatterplot

smoother of $\tilde{\mu}_{X,z^{(p)}}(s)$ is defined through minimizing

$$\sum_{i\in\mathcal{N}_{z^{(p)},h}}\sum_{j=1}^{N_i}\kappa_1(\frac{S_{ij}-s}{b_{X,z^{(p)}}})(U_{ij}-d_0-d_1(S_{ij}-s))^2,$$

with respect to d_0 and d_1 and setting $\tilde{\mu}_{X,z^{(p)}}(s)$ to be the minimizer $\hat{d}_0$, where $\kappa_1(\cdot)$ is a kernel function and $b_{X,z^{(p)}}$ is the smoothing bandwidth, the choice of which will be discussed in Section 4. We define a similar local linear scatterplot smoother of $\tilde{\mu}_{Y,z^{(p)}}(t)$. According to Lemma 2 in Appendix, raw estimates $\tilde{\mu}_{X,z^{(p)}}(s)$ and $\tilde{\mu}_{Y,z^{(p)}}(t)$ are consistent uniformly for $z^{(p)}$, $p=1,2,\cdots,P$, for appropriate bandwidths $b_{X,z^{(p)}}$ and $b_{Y,z^{(p)}}$.

Extending Yao et al. (2005b), the conditional slope function can be represented as

$$\beta(z,s,t)=\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\frac{E\zeta_{z,m}\xi_{z,k}}{E\zeta_{z,m}^2}\psi_{z,m}(s)\phi_{z,k}(t) \tag{2.3}$$

for each z, where $\psi_{z,m}(\cdot)$ and $\phi_{z,k}(\cdot)$ are the eigenfunctions of covariance functions $G_{X,z}(\cdot,\cdot)$ and $G_{Y,z}(\cdot,\cdot)$ respectively, and $\zeta_{z,m}$ and $\xi_{z,k}$ are the functional principal component scores of X and Y respectively, conditional on $Z=z$.

To obtain raw slope estimates $\tilde{\beta}(z^{(p)},s,t)$, for $p=1,2,\cdots,P$, we first estimate the conditional covariance functions $G_{X,z^{(p)}}(s_1,s_2)$, $G_{Y,z^{(p)}}(t_1,t_2)$, and $C_{XY,z^{(p)}}(s,t)$ at each bin center, based on the observations falling into the bin, using the approach of Yao et al. (2005b). From "raw" covariances: $G_{X,i,z^{(p)}}(S_{ij},S_{ik})=(U_{ij}-\tilde{\mu}_{X,z^{(p)}}(S_{ij}))(U_{ik}-\tilde{\mu}_{X,z^{(p)}}(S_{ik}))$ for $1\le j,k\le L_i$, $i\in\mathcal{N}_{z^{(p)},h}$ and $p=1,2,\cdots,P$, the locally smoothed conditional covariance $\tilde{G}_{X,z^{(p)}}(s_1,s_2)$ is defined as the minimizer $\hat{b}_0$ of the following local linear problem:

$$\min_{b_0,b_{11},b_{12}}\sum_{i\in\mathcal{N}_{z^{(p)},h}}\sum_{1\le j\ne l\le L_i}\kappa_2(\frac{S_{ij}-s_1}{h_{X,z^{(p)}}},\frac{S_{il}-s_2}{h_{X,z^{(p)}}})\left[G_{X,i,z^{(p)}}(S_{ij},S_{il})-b_0-b_{11}(S_{ij}-s_1)-b_{12}(S_{il}-s_2)\right]^2,$$

where $\kappa_2(\cdot,\cdot)$ is a bivariate kernel function and $h_{X,z^{(p)}}$ a smoothing bandwidth. The diagonal "raw" covariances $G_{X,i,z^{(p)}}(S_{ij},S_{ij})$ are removed from the objective function of the above minimization problem, because $EG_{X,i,z^{(p)}}(S_{ij},S_{il})\approx\mathrm{cov}(X(S_{ij}),X(S_{il}))+\delta_{jl}\sigma_X^2$ where $\delta_{jl}=1$ if $j=l$ and 0 otherwise. Analogous considerations apply for $\tilde{G}_{Y,z^{(p)}}(T_{ij},T_{il})$. The diagonal "raw" covariances $G_{X,i,z^{(p)}}(S_{ij},S_{ij})$ and $G_{Y,i,z^{(p)}}(T_{ij},T_{ij})$ can be smoothed with bandwidths $b_{X,z^{(p)},V}$ and $b_{Y,z^{(p)},V}$ to estimate $V_{X,z^{(p)}}(s)=G_{X,z^{(p)}}(s,s)+\sigma_X^2$ and $V_{Y,z^{(p)}}(t)=G_{Y,z^{(p)}}(t,t)+\sigma_Y^2$, and the resulting estimators are denoted by $\tilde{V}_{X,z^{(p)}}(s)$ and $\tilde{V}_{Y,z^{(p)}}(t)$, respectively, and the differences $(\tilde{V}_{X,z^{(p)}}(s)-\tilde{G}_{X,z^{(p)}}(s,s))$ and analogously for

Y can be used to obtain estimates $\tilde{\sigma}^2_{X,z^{(p)}}$ for σ^2_X and $\tilde{\sigma}^2_{Y,z^{(p)}}$ for σ^2_Y by integration. Furthermore, "raw" conditional cross-covariances $C_{i,z^{(p)}}(S_{il}, T_{ij}) = (U_{il} - \tilde{\mu}_{X,z^{(p)}}(S_{il}))(V_{ij} - \tilde{\mu}_{Y,z^{(p)}}(T_{ij}))$ are used to estimate $C_{XY,z^{(p)}}(s,t)$, by minimizing

$$\sum_{i\in\mathcal{N}_{z^{(p)},h}} \sum_{1\le l\le L_i} \sum_{1\le j\le N_i} \kappa_2(\frac{S_{ij}-s}{h_{1,z^{(p)}}}, \frac{T_{ij}-t}{h_{2,z^{(p)}}}) \left[C_{i,z^{(p)}}(S_{il},T_{ij}) - b_0 - b_{11}(S_{il}-s) - b_{12}(T_{ij}-t)\right]^2,$$

with respect to b_0, b_{11}, and b_{12}, and setting $\tilde{C}_{XY,z^{(p)}}(s,t)$ to be the minimizer $\hat{b}_0$, with smoothing bandwidths $h_{1,z^{(p)}}$ and $h_{2,z^{(p)}}$.

In (2.3), the slope function may be represented via the eigenvalues and eigenfunctions of the covariance operators. To obtain the estimates $\tilde{\rho}_{z^{(p)},m}$ and $\tilde{\psi}_{z^{(p)},m}(\cdot)$ (resp. $\tilde{\lambda}_{z^{(p)},k}$ and $\tilde{\phi}_{z^{(p)},k}(\cdot)$) of eigenvalue-eigenfunction pairs $\rho_{z^{(p)},m}$ and $\psi_{z^{(p)},m}(\cdot)$ (resp. $\lambda_{z^{(p)},k}$ and $\phi_{z^{(p)},k}(\cdot)$), we use Conditional Functional Principal Component Analysis (CFPCA) for $\tilde{G}_{X,z^{(p)}}(\cdot,\cdot)$ (resp. $\tilde{G}_{Y,z^{(p)}}(\cdot,\cdot)$), by numerically solving the conditional eigen-equations

$$\int_{\mathcal{S}} \tilde{G}_{X,z^{(p)}}(s_1,s_2)\tilde{\psi}_{z^{(p)},m}(s_1)ds_1 = \tilde{\rho}_{z^{(p)},m}\tilde{\psi}_{z^{(p)},m}(s_2), \qquad m=1,2,\cdots, \quad (2.4)$$

$$\int_{\mathcal{T}} \tilde{G}_{Y,z^{(p)}}(t_1,t_2)\tilde{\phi}_{z^{(p)},k}(t_1)dt_1 = \tilde{\lambda}_{z^{(p)},k}\tilde{\phi}_{z^{(p)},k}(t_2), \qquad k=1,2,\cdots. \quad (2.5)$$

Note that we estimate the conditional mean functions and conditional covariance functions over dense grids of $\mathcal{S}$ and $\mathcal{T}$. Numerical integrations like the one on the left hand side of (2.4) are done over these dense grids using the Trapezoid Rule. Note further that integrals over individual trajectories are not needed for the regression focus in that we use conditional expectation to estimate principal component scores as in (4.1).

Due to the fact that

$$C_{XY,z}(s,t) = \mathrm{cov}(X(s),Y(t)|Z=z) = \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} E(\zeta_{z,m}\xi_{z,k})\psi_{z,m}(s)\phi_{z,k}(t),$$

we then obtain preliminary estimates of $\sigma_{z,mk} = E(\zeta_{z,m}\xi_{z,k})$ at the bin centers $z^{(p)}$, $p=1,2,\cdots,P$, by numerical integration,

$$\tilde{\sigma}_{z^{(p)},mk} = \int_{\mathcal{T}}\int_{\mathcal{S}} \tilde{\psi}_{z^{(p)},m}(s)\tilde{C}_{XY,z^{(p)}}(s,t)\tilde{\phi}_{z^{(p)},k}(t)dsdt. \quad (2.6)$$

With (2.3), (2.4), (2.5), and (2.6), the raw estimates of $\beta(z^{(p)},s,t)$ are

$$\tilde{\beta}(z^{(p)},s,t) = \sum_{k=1}^{K}\sum_{m=1}^{M} \frac{\tilde{\sigma}_{z^{(p)},mk}}{\tilde{\rho}_{z^{(p)},m}}\tilde{\psi}_{z^{(p)},m}(s)\tilde{\phi}_{z^{(p)},k}(t). \quad (2.7)$$

Further details on the "global" case can be found in Yao et al. (2005b).

2.2.2. Refining the raw estimates

We establish in the Appendix that the raw estimates $\tilde{\mu}_{X,z^{(p)}}(s)$, $\tilde{\mu}_{Y,z^{(p)}}(t)$, and $\tilde{\beta}(z^{(p)},s,t)$ are consistent. As has been demonstrated in Fan and Zhang (2000), there are several reasons to refine such raw estimates. For example, the raw estimates are generally not smooth and are based on local observations, hence inefficient. Most importantly, applications require that the function $\beta(z,s,t)$ is available for any $z \in \mathcal{Z}$.

To refine the raw estimates, the classical way is smoothing, for which we adopt the local polynomial smoother. Defining $\mathbf{c}_p = (1, z^{(p)} - z, \cdots, (z^{(p)} - z)^r)^T$, $p = 1, 2, \cdots, P$, the local polynomial smoothing weights for estimating the q-th derivative of an underlying function are

$$\omega_{q,r+1}(z^{(p)}, z, b) = q!\mathbf{e}_{q+1,r+1}^T(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}\mathbf{c}_p K_b(z^{(p)} - z), \quad p = 1, 2, \cdots, P,$$

where $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_P)^T$, $\mathbf{W} = diag(K_b(z^{(1)} - z), K_b(z^{(2)} - z), \cdots, K_b(z^{(P)} - z))$, and $\mathbf{e}_{q+1,r+1} = (0, \cdots, 0, 1, 0, \cdots, 0)^T$ is a unit vector of length $r+1$ with the $q+1$-th element being 1 (see Fan and Gijbels, 1996). Our final estimators are given by

$$\begin{aligned}
\hat{\mu}_{X,z}(s) &= \sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)\tilde{\mu}_{X,z^{(p)}}(s), \\
\hat{\mu}_{Y,z}(t) &= \sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)\tilde{\mu}_{Y,z^{(p)}}(t), \\
\hat{\beta}(z,s,t) &= \sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)\tilde{\beta}(z^{(p)}, s, t).
\end{aligned}$$

Due to the assumption that the variance of the measurement error does not depend on the additional covariate, the final estimators of σ_X^2 and σ_Y^2 can be taken as simple averages

$$\hat{\sigma}_X^2 = \sum_{p=1}^{P} \tilde{\sigma}_{X,z^{(p)}}^2/P \text{ and } \hat{\sigma}_Y^2 = \sum_{p=1}^{P} \tilde{\sigma}_{Y,z^{(p)}}^2/P. \tag{2.8}$$

Remark 1. *The localization to $Z = z$ as needed for the proposed varying coefficient model, coupled with the extreme sparseness assumption [A2], which adequately reflects longitudinal designs, is not conducive to obtain explicit results in terms of convergence rates for the general case. However, one may find by suitably moifiying our arguments and coupling them with the rates of convergence provided on p.2891 of Yao et al. (2005b) that*

we can obtain rates if desired. These are the rates given there, which depend on complex intrinsic properties of the underlying processes, provided everywhere the sample size n is replaced by nh, the sample size for each bin.

Remark 2. *In this work, we focus on sparse and irregularly observed longitudinal data. For the case where entire processes are observed without noise and error-free, one can estimate the localized eigenfunctions at rates of L^2 convergence of $(n\tilde{h})^{-1/2}$ (see Hall et al., 2006), where $\tilde{h}$ is the smoothing bandwidth. For the moments of the functional principal components a smoothing step is not needed. Known results will be adjusted by replacing n with nh, when conditioning on a fixed covariate level $Z = z$; see Cai and Hall (2006) and Hall and Horowitz (2007).*

3. Asymptotic properties

We establish some key asymptotic consistency properties for the proposed estimators. Detailed technical conditions and proofs can be found in the Appendix.

The observed dataset is denoted by $\{Z_i, (S_{il}, U_{il})_{l=1}^{L_i}, (T_{ij}, V_{ij})_{j=1}^{N_i} : i = 1, 2, \cdots, n\}$. We assume that it comes from (1.2) and satisfies [A0], [A1], [A2], and [A3].

For $\tilde{n} \propto \sqrt{n}$, define the event

$$E_n = \{\min n_{z^{(p)},h} > \tilde{n}\}, \tag{3.1}$$

where $n_{z^{(p)},h}$ is the number of observations in the pth bin and $\tilde{n} \propto \sqrt{n}$ means that there exist c_0 and C_0 such that $0 < c_0 \leq \tilde{n}/\sqrt{n} \leq C_0 < \infty$. It is shown in Proposition 1 in Appendix that $P(E_n) \to 1$ as $n \to \infty$ for $P \propto n^{1/8}$ as specified by Condition (xi).

The global consistency of the final mean and slope function estimates follows from

Theorem 1 (Consistency of time-varying functional regression)**.** *Under Conditions [A0], [A1], [A2], and [A3] in Section 2 and Conditions [A4], [A5], (i-xi) in Appendix, on event E_n with $P(E_n) \to 1$ as $n \to \infty$ we have*

$$\int_{\mathcal{Z}} \int_{\mathcal{R}} (\hat{\mu}_{W,z}(r) - \mu_{W,z}(r))^2 dr dz \xrightarrow{P} 0 \text{ for } W = X, \mathcal{R} = \mathcal{S} \text{ and } W = Y, \mathcal{R} = \mathcal{T}$$

and

$$\int_{\mathcal{Z}} \int_{\mathcal{T}} \int_{\mathcal{S}} (\hat{\beta}(z, s, t) - \beta(z, s, t))^2 ds dt dz \xrightarrow{P} 0.$$

To study prediction through time-varying functional regression, consider a new predictor process X^* with associated covariate Z^*. The corresponding conditional expected response process Y^* and its prediction $\hat{Y}^*$ are given by

$$Y^*(t) = E(Y(t)|X^*, Z^*) = \mu_{Y,Z^*}(t) + \int_{\mathcal{S}} \beta(Z^*, s, t)(X^*(s) - \mu_{X,Z^*}(s))ds, \quad (3.2)$$

$$\hat{Y}^*(t) = \hat{\mu}_{Y,Z^*}(t) + \int_{\mathcal{S}} \hat{\beta}(Z^*, s, t)(X^*(s) - \hat{\mu}_{X,Z^*}(s))ds. \quad (3.3)$$

Theorem 2 (Consistency of prediction)**.** *For a new predictor process X^* with associated covariate Z^*, it holds under conditions of Theorem 1 that $\int_{\mathcal{T}}(Y^*(t) - \hat{Y}^*(t))^2 dt \xrightarrow{P} 0$, where $Y^*(t)$ and $\hat{Y}^*(t)$ are given by (3.2) and (3.3).*

4. Finite-sample implementation

For the finite-sample case, several smoothing parameters need to be chosen. Following Yao et al. (2005a), the one-curve-leave-out cross-validation method can be used to select smoothing parameters $b_{X,z^{(p)}}$, $b_{Y,z^{(p)}}$, $b_{X,z^{(p)},V}$, $b_{Y,z^{(p)},V}$, $h_{X,z^{(p)}}$, $h_{Y,z^{(p)}}$, $h_{1,z^{(p)}}$, and $h_{2,z^{(p)}}$, individually for each bin. Further required choices concern the bin width h, the smoothing bandwidth b, and the numbers M and K of included expansion terms in (2.7). The method of cross-validation could also be used for these additional choices, incurring however a heavy computational load. A fast alternative is a pseudo-Akaike Information Criterion (AIC) (or pseudo-Bayesian Information Criterion (BIC)).

[1] Choose the number of terms in the truncated double summation representation $\tilde{\beta}(z^{(p)}, s, t)$, for $M(n)$ and $K(n)$ using AIC or BIC as in Yao et al. (2005b).

[2] For each bin width h, choose the best smoothing bandwidth $b^*(h)$, by minimizing AIC or BIC.

[3] Choose the bin width h^* which minimizes AIC or BIC, while for each h investigated we use $b^*(h)$ for b.

For [1], we will choose M and K simultaneously for all bins, minimizing the conditional penalized pseudo-deviance given by

$$C(K) = \sum_{p=1}^{P} \sum_{i \in \mathcal{N}_p} \left\{ \frac{1}{\tilde{\sigma}^2_{Y,z^{(p)}}} \tilde{\boldsymbol{\epsilon}}_i^T \tilde{\boldsymbol{\epsilon}}_i + N_i \log(2\pi) + N_i \log \tilde{\sigma}^2_{Y,z^{(p)}} \right\} + \mathcal{P}$$

where $\mathcal{P} = 2PK$ for AIC and $\mathcal{P} = (\log n)PK$ for BIC, with respect to K. Here, for $i \in \mathcal{N}_p$, $\tilde{\boldsymbol{\epsilon}}_i = \mathbf{V}_i - \tilde{\boldsymbol{\mu}}_{Y,z^{(p)},i} - \sum_{k=1}^{K} \tilde{\xi}^*_{z^{(p)},k,i} \tilde{\boldsymbol{\phi}}_{z^{(p)},k,i}$ with $\tilde{\boldsymbol{\mu}}_{Y,z^{(p)},i} = (\tilde{\mu}_{Y,z^{(p)}}(T_{i1}), \cdots, \tilde{\mu}_{Y,z^{(p)}}(T_{iN_i}))^T$, $\mathbf{V}_i = (V_{i1}, \cdots, V_{iN_i})^T$, $\tilde{\boldsymbol{\phi}}_{z^{(p)},k,i} = (\tilde{\phi}_{z^{(p)},k}(T_{i1}), \cdots, \tilde{\phi}_{z^{(p)},k}(T_{iN_i}))^T$, and with estimated principal components

$$\tilde{\xi}^*_{z^{(p)},k,i} = \tilde{\lambda}_{z^{(p)},k} \tilde{\boldsymbol{\phi}}^T_{z^{(p)},k,i} \tilde{\Sigma}^{-1}_{Y,z^{(p)},i} (\mathbf{V}_i - \tilde{\boldsymbol{\mu}}_{Y,z^{(p)},i}),$$

where $\tilde{\Sigma}_{Y,z^{(p)},i}$ is a N_i-by-N_i matrix whose (j,k)-element is $\tilde{G}_{Y,z^{(p)}}(T_{ij}, T_{ik}) + \tilde{\sigma}^2_{Y,z^{(p)}} \delta_{jk}$. Analogous criteria are used for predictor process X, selecting K by minimizing $AIC(K)$, and $BIC(K)$. Marginal versions of these criteria are also available.

In Step [2], for each bin width h, we first select the best smoothing bandwidth $b^*(h)$ based on AIC or BIC, and then select the final bin width h^* by a second application of AIC or BIC, plugging in $b^*(h)$ in this selection, as follows. For a given bin width h, define the P-by-P smoothing matrix $\mathbf{S}_{0,2}$ whose (p_1, p_2)-th element is $\omega_{0,2}(z^{(p_1)}, z^{(p_2)}, b)$. Then the effective number of parameters of the smoothing matrix is the trace of $\mathbf{S}^T_{0,2}\mathbf{S}_{0,2}$ (cf Wahba, 1990). This suggests minimization of

$$\mathrm{AIC}(b|h) = \sum_{i=1}^{n} \left\{ \frac{1}{\hat{\sigma}^2_Y} \hat{\boldsymbol{\epsilon}}^T_i \hat{\boldsymbol{\epsilon}}_i + N_i \log(2\pi) + N_i \log \hat{\sigma}^2_Y \right\} + 2\mathrm{tr}(\mathbf{S}^T_{0,2}\mathbf{S}_{0,2}),$$

leading to $b^*(h)$, where

$$\hat{\boldsymbol{\epsilon}}_i = \mathbf{V}_i - \hat{\boldsymbol{\mu}}_{Y,z_i,i} - \sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z_i, b) \sum_{m,k=1}^{M,K} \frac{\tilde{\sigma}_{z^{(p)},mk}}{\tilde{\rho}_{z^{(p)},m}} \hat{\zeta}^*_{z^{(p)},m,i} \tilde{\boldsymbol{\phi}}_{z^{(p)},k,i}$$

with $\hat{\boldsymbol{\mu}}_{Y,z_i,i} = (\hat{\mu}_{Y,z_i}(T_{i1}), \cdots, \hat{\mu}_{Y,z_i}(T_{iN_i}))^T$, and estimated principal component scores

$$\hat{\zeta}^*_{z^{(p)},m,i} = \tilde{\rho}_{z^{(p)},k} \tilde{\boldsymbol{\psi}}^T_{z^{(p)},m,i} \tilde{\Sigma}^{-1}_{X,z^{(p)},i} (\mathbf{U}_i - \hat{\boldsymbol{\mu}}_{X,z_i,i}).$$

The definition of pseudo-BIC scores is analogous.

In Step [3], to select the bin width h^*, we minimize

$$\mathrm{AIC}(h, b^*(h)) = \sum_{i=1}^{n} \left\{ \frac{1}{\hat{\sigma}^2_Y} \hat{\boldsymbol{\epsilon}}^T_i \hat{\boldsymbol{\epsilon}}_i + N_i \log(2\pi) + N_i \log \hat{\sigma}^2_Y \right\} + 2MKP$$

or the analogous BIC score, using $b^*(h)$ for each h, as determined in the previous step.

5. Simulation study

We compare global functional linear regression and varying-coefficient functional linear regression through simulated examples with a functional response. For the case of a scalar response, the proposed varying-coefficient functional linear regression approach achieves similar performance improvements (results not reported). For the finite-sample case, there are several parameters to be selected (see Section 4). In the simulations we use pseudo-AIC to select bin width h and pseudo-BIC to select the smoothing bandwidth b and the number of regularization terms $M(n)$ and $K(n)$.

The domains of predictor and response trajectories are chosen as $\mathcal{S} = [0, 10]$ and $\mathcal{T} = [0, 10]$, respectively. The predictor trajectories X are generated as $X(s) = \mu_X(s) + \sum_{m=1}^{3} \zeta_m \psi_m(s)$ for $s \in \mathcal{S}$, with mean predictor trajectory $\mu_X(s) = (s + \sin(s))$, and the three eigenfunctions are $\psi_1(s) = -\sqrt{\frac{1}{5}}\cos(\pi s/5)$, $\psi_2(s) = \sqrt{\frac{1}{5}}\sin(\pi s/5)$, $\psi_3(s) = -\sqrt{\frac{1}{5}}\cos(2\pi s/5)$, and their corresponding functional principal components are independently distributed as $\zeta_1 \sim N(0, 2^2)$, $\zeta_2 \sim N(0, \sqrt{2}^2)$, $\zeta_3 \sim N(0, 1^2)$. The additional covariate Z is uniformly distributed over $[0, 1]$. For $z \in [0, 1]$, the slope function is linear in z, $\beta(z, s, t) = (z + 1)(\psi_1(s)\psi_1(t) + \psi_2(s)\psi_2(t) + \psi_3(s)\psi_3(t))$ and the conditional response trajectory is $E(Y(t)|X, Z = z) = \mu_{Y,z}(t) + \int_0^{10} \beta(z, s, t)(X(s) - \mu_X(s))ds$, where $\mu_{Y,z}(t) = (1 + z)(t + \sin(t))$. We consider the following two cases.

Example 5.1 Regular case

The first example focuses on the regular case with dense measurement design. Observations on the predictor and response trajectories are made at $s_j = (j - 1)/3$ for $j = 1, 2, \cdots, 31$ and $t_j = (j - 1)/3$ for $j = 1, 2, \cdots, 31$, respectively. We assume the measurement errors on both the predictor and response trajectories are distributed as $N(0, 1^2)$, *i.e.*, $\sigma_X^2 = 1$ and $\sigma_Y^2 = 1$.

Example 5.2 Sparse and irregular case

In this example, we make a random number of measurements on each trajectory in the training data set, chosen with equal probability from $\{2, 3, \cdots, 10\}$. We note that, for the same subject, the number of measurements on the predictor and the number of measurements on the response trajectory are independent. For any trajectory, given the number of measurements, the measurement times are uniformly distributed over the

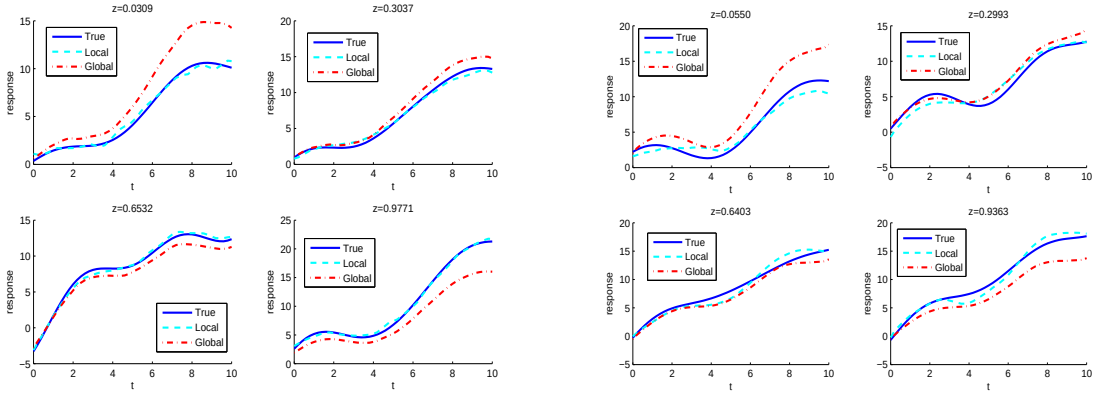
TABLE 1

Simulation results: Mean and standard deviation of MISPE for global and varying-coefficient functional linear regression with a functional response, for both regular and sparse cases.

	functional linear regression	varying-coefficient functional linear regression
Regular	4.0146 (1.6115)	0.7836 (0.4734)
Sparse and irregular	4.0013 (0.8482)	1.0637 (0.3211)

corresponding trajectory domain. The measurement error is distributed as $N(0, 1^2)$ for both the predictor and the response trajectories.

FIG 1. *In one random repetition, the true (solid) conditional expected response trajectories and predicted conditional expected response trajectories via the global functional linear regression (dot-dashed) and the varying-coefficient functional linear regression (dashed) are plotted for four randomly selected subjects in the independent test set with median Integrated Squared Prediction Error. The left four panels and the right four correspond to the regular and sparse irregular cases, respectively.*



In both examples, the training sample size is 400. An independent test set of size 1000 is generated with the predictor and response trajectories fully observed. We compare performance using Mean Integrated Squared Prediction Error (MISPE)

$$\frac{1}{1000}\sum_{j=1}^{1000}\int_{\mathcal{T}}\left[E(Y_j^*(t)|X_j^*, Z_j^*) - \left(\hat{\mu}_{Y,Z_j^*}(t) + \int_{\mathcal{S}}\hat{\beta}(Z_j^*, s, t)(X_j^*(s) - \hat{\mu}_{X,Z_j^*}(s))ds\right)\right]^2 dt/|\mathcal{T}|,$$

analogously for the global functional linear regression, where (X_j^*, Y_j^*, Z_j^*) denotes the data of the j-th subject in the independent test set. In Table 1, we report the mean and standard deviation (in parenthesis) of the MISPE of the global and the varying-coefficient functional linear regression over 100 repetitions for each case. This shows that in this simulation setting, the proposed varying-coefficient functional linear regression approach reduces MISPE drastically, compared with the global functional linear regression, both for regular and sparse irregular designs.

To visualize the differences between predicted conditional expected response trajectories, for a small random sample in both the regular and sparse and irregular design cases, we randomly choose four subjects from the test set with median values of the Integrated Squared Prediction Error (ISPE) for the varying-coefficient functional linear regression. The true and predicted conditional expected response trajectories are plotted in Figure 1, where the left four panels correspond to the regular design case and the right four to the sparse irregular case. Clearly, the locally varying method is seen to be superior.

6. Applications

We illustrate the comparison of the proposed varying-coefficient functional linear model with the global functional linear regression in two applications.

FIG 2. *The slope functions estimated by the varying-coefficient functional linear regression at different levels of the additional covariate z for the egg laying data.*

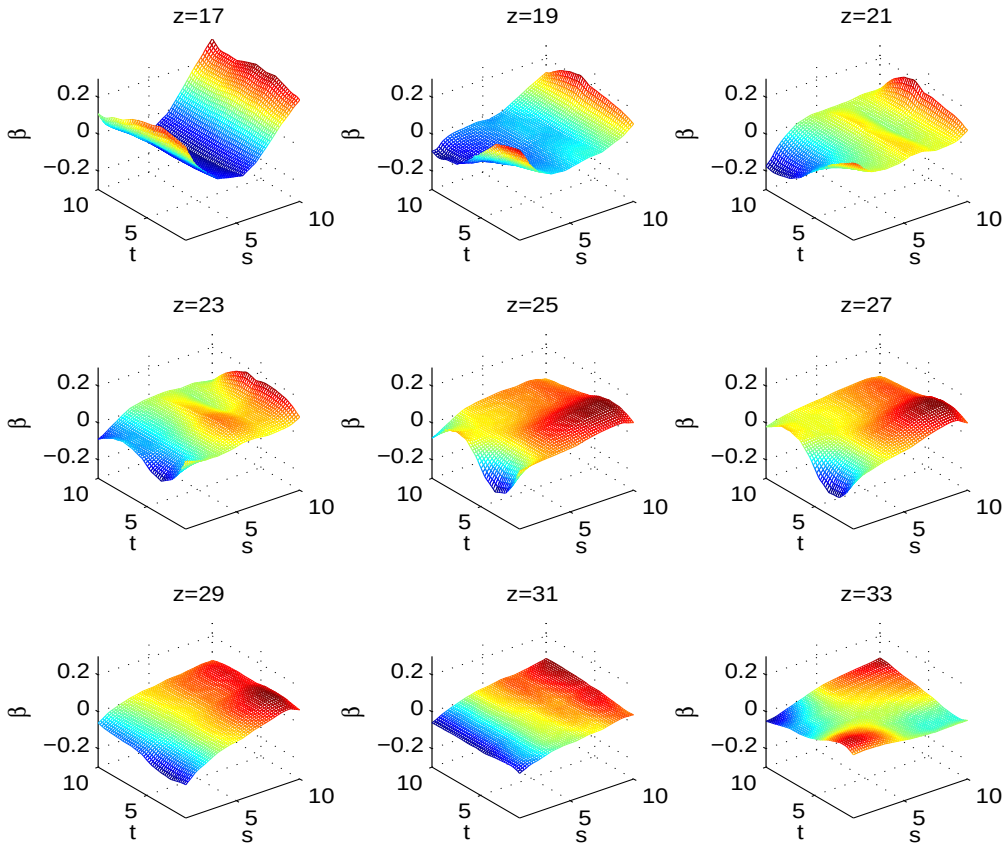
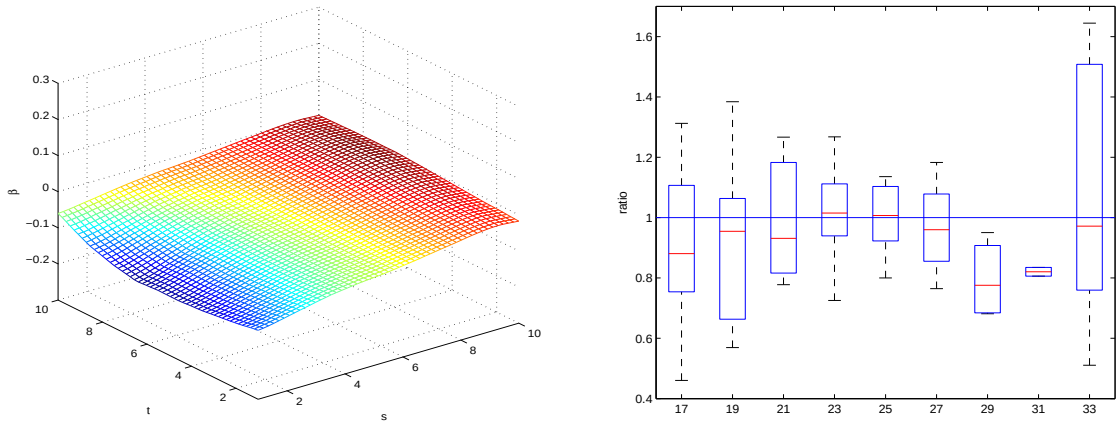


FIG 3. *The left panel plots the slope function estimated by the global functional linear regression for the egg laying data and the right panel corresponds to boxplots of the ratios of MSPE of the varying-coefficient functional linear regression to that of the global functional linear regression for the subjects in the test data set for different levels of the additional covariate Z.*



6.1. Egg laying data

The egg laying data represent the entire reproductive history of one thousand Mediterranean fruit flies (medflies for short), where daily fecundity, quantified by the number of eggs laid per day, was recorded for each fly during its lifetime. See Carey et al. (1998) for details of this data set and experimental background.

We are interested in predicting future egg-laying patterns over an interval of fixed length but with potentially different starting time, based on the daily fecundity information during a fixed earlier period. The predictor trajectories were chosen as daily fecundity between day 8 and day 17. This interval covers the tail of an initial rapid rise to peak egg-laying and the initial part of the subsequent decline and generally the egg-laying behavior at and near peak egg-laying is included. It is of interest to study in what form the intensity of peak egg-laying is associated with subsequent egg-laying behavior, as trade-offs may point to constraints that may play a role in the evolution of longevity.

While the predictor process is chosen with a fixed domain, the response process has a moving domain, with fixed length of ten days, but different starting age for each subject, which serves as the additional covariate Z. Due to the limited number of subjects in this study, we use a pre-specified discrete set for the values of Z: $\mathcal{Z} = \{17, 19, 21, 23, 25, 27, 29, 31, 33\}$ with a pre-specified bin width $h = 2$. For subject i with $z_i \in \mathcal{Z}$, measurements U_{ij} on the predictor trajectory are the daily numbers of eggs on day $j + 7$ and measurements

V_{ik} on the response trajectory correspond to the daily number of eggs on day $k + z_i$ for $j = 1, 2, \cdots, 10$ and $k = 1, 2, \cdots, 10$. The number of subjects in these bins is 30, 29, 18, 29, 22, 19, 19, 17, 36, respectively. For each bin, we randomly select 15 subjects as the training set and the remaining subjects are used to evaluate the prediction performance, comparing the performance of the global and the varying-coefficient functional linear regression. The prediction performance is quantified by Mean Squared Prediction Error (MSPE), defined for each subject i in the test set as

$$\text{MSPE}_g(i) = \frac{1}{10}\sum_{k=1}^{10}(\hat{y}_{ik}^g - V_{ik})^2 \text{ and } \text{MSPE}_l(i) = \frac{1}{10}\sum_{k=1}^{10}(\hat{y}_{ik}^l - V_{ik})^2,$$

where $\hat{y}_{ik}^g$ and $\hat{y}_{ik}^l$ denote the predicted daily fecundities corresponding to V_{ik} using the global, resp. the proposed varying-coefficient (local) functional linear regression.

Through pseudo-AIC, the global functional linear regression selects 2 and 3 principal components for the predictor and response trajectories, respectively, while the varying-coefficient functional linear regression uses 2 principal components for both trajectories. After smoothing, the slope functions estimated by the varying-coefficient models are plotted in Figure 2 for different values of Z and the estimated slope function for the global functional linear regression in the left panel of Figure 3. Boxplots of the ratio $\text{MSPE}_l(i)/\text{MSPE}_g(i)$ for subjects in the test data set are shown in the right panel of Figure 3 for different levels of the covariate Z. There is one outlier above the maximum value for $Z = 18$ which is not shown. For most bins, the median ratios are seen to be smaller than one, indicating improvement of our new varying-coefficient functional linear regression. Denoting the average MSPE (over the independent test data set) of the global and the varying-coefficient functional linear regression by $\overline{\text{MSPE}}_g$ and $\overline{\text{MSPE}}_l$, respectively, the relative performance gain $(\overline{\text{MSPE}}_l - \overline{\text{MSPE}}_g)/\overline{\text{MSPE}}_g$ is found to be -0.0810, so that the prediction improvement of the varying-coefficient method is 8.1%.

Besides prediction, it is of interest to study the dependency of the future egg-laying behavior on peak egg-laying. From the changing slope functions in Figure 2, we find that for the segments close to the peak segments, the egg laying pattern is inverting the peak pattern, meaning that sharper and higher peaks are associated with sharp downturns, pointing to a near-future exhaustion effect of peak egg-laying. In contrast, the shape of egg-laying segments farther into the future is predicted by the behavior of the first

derivative over the predictor segment, so that slow declines near the end of peak egg-laying are harbingers of future robust egg-laying. This is in concordance with a model of exponential decline in egg-laying that has been proposed by Müller et al. (2001).

6.2. BLSA data with scalar response

As a second example, we use a subset of data from the Baltimore Longitudinal Study of Aging (BLSA), a major longitudinal data of human aging (Shock et al., 1984; Pearson et al., 1997). The data consist of 1590 male volunteers who were scheduled to be seen twice per year However, many participants missed scheduled visits or were seen at other than scheduled times, so that the data are sparse and irregular with unequal numbers of measurements and different measurement times for each subject. For each subject, current age and systolic blood pressure (SBP) were recorded at each visit. We quantify how the SBP trajectories of a subject available in a middle age range between age 48 and age 53 affect the average of the SBP measurements made during the last five years included in this study, at an older age. So the predictor domain is of length five years and the response is scalar. The additional covariate for each subject is the beginning age of the last five-year interval included in the study. After excluding subjects with less than two measurements in the predictor, 214 subjects were included for whom the additional covariate ranged between 55 and 75. We bin the data according to the additional covariate with bin centers at ages 56.0, 59.0, 62.0, 65.0, 68.5, 73.0 years, and the numbers of subjects in each of these bins are 38, 33, 38, 32, 39, 34.

We randomly selected 25 subjects from each bin for model estimation and the remaining subjects to evaluate the prediction performance. In contrast to the egg-laying data, the predictor measurements in this longitudinal study are sparse and irregular. Pseudo-BIC selects two principal components for the predictor trajectories for both global and varying-coefficient functional linear regressions. Using the same criterion for relative performance gain as in the previous example, the varying-coefficient functional linear regression achieves 11.8% improvement compared to the global functional linear regression. Estimated slope functions are shown in Figure 5 and predictor trajectories in Figure 4.

The shape changes of the slope functions with changing covariate indicate that the negative derivative of SBP during the middle-age period is associated with near-future

FIG 4. *Plots of predictor processes: the left panel for the global functional linear regression and the right panel for different bins according to the additional covariate in the varying-coefficient functional linear regression.*

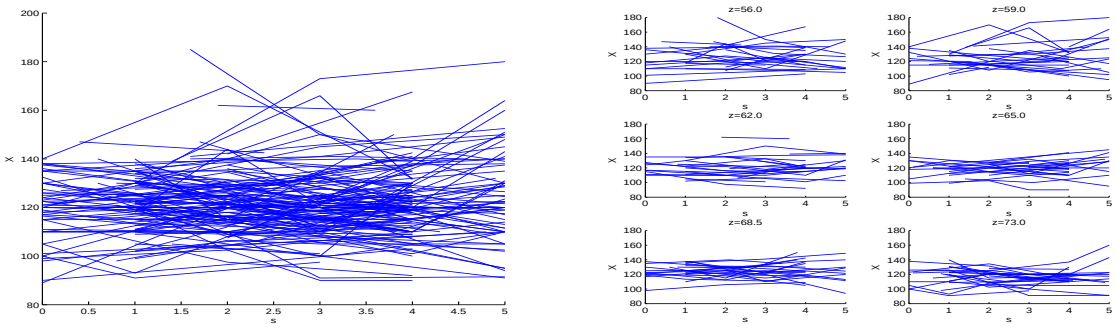
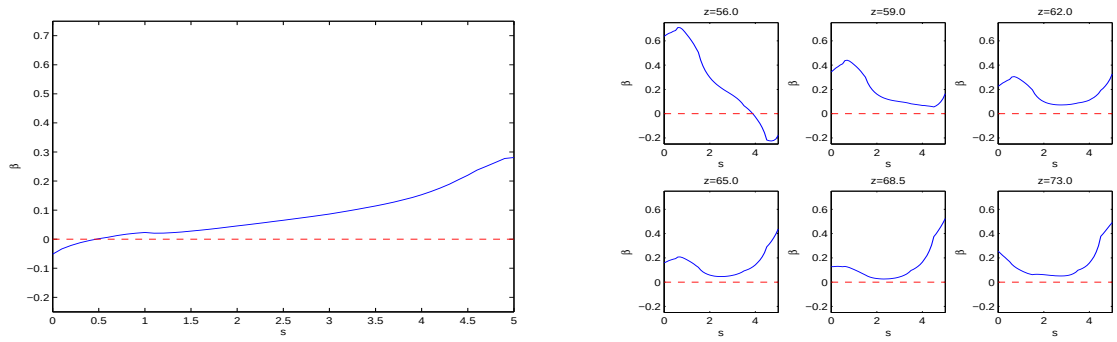


FIG 5. *The estimated slope function via the global functional linear regression and the new proposed varying-coefficient functional linear regression (for different levels of Z) are plotted as the solid lines in the left and right panels, respectively.*



SBP, and further into the future this pattern is reversed and an SBP increase near the right end of the initial period is becoming predictive.

7. Concluding remarks

Our results indicate that established functional linear regression models can be improved when an available covariate is incorporated. We implement this idea by extending the functional linear model to a varying-coefficient version, inspired by the analogous highly successful extension of classical regression models. In both application examples the increased flexibility, that is inherent in this extension, leads to clear gains in prediction error. In addition, it is often of interest to ascertain the effect of the additional covariate. This can be done by plotting the regression slopes for each bin defined by the covariate and observing the dependency of this function or surface on the value of the covariate.

Further extensions that are of interest in many applications concern the case of multivariate covariates. If the dimension is low, the smoothing methods and binning methods that we propose here can be extended to this case. For higher-dimensional covariates or covariates that are not continuous, one could form a single index to summarize the covariates and thus create a new one-dimensional covariate, which then enters the functional regression model in the same way as the one-dimensional covariate that we consider.

As seen in the data applications, the major applications of the proposed methodology are expected to come from longitudinal studies with sparse and irregular measurements where the presence of additional non-functional covariates is common.

Acknowledgements

We wish to thank two referees for helpful comments. Yichao Wu's research has been supported in part by NIH grant R01-GM07261. Jianqing Fan's research has been supported in part by National Science Foundation (NSF) grants DMS-03-54223 and DMS-07-04337. Hans-Georg Müller has been supported in part by National Science Foundation (NSF) grants DMS-03-54223, DMS-05-05537 and DMS-08-06199.

References

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, **51** 285–311.

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics*, **34** 2159–2179.

Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian Journal of Statistics*, **34** 317–335.

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

Cardot, H. and Sarda, P. (2005). Varying-coefficient functional linear regression models. Technical Report LSP, Universite Paul Sabatier.

Carey, J., Liedo, P., Müller, H., Wang, J. and Chiou, J. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of mediterranean fruit fly females. *Journal of Gerontology –Biological Sciences*, **53** 245–251.

Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics*. 1989th ed. Wiley, New York.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association*, **35** 632–641.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of Royal Statistical Society Series B*, **62** 303–322.

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

Hall, P., Müller, H. and Wang, J. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, **34** 1493–1517.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, **87** 587–602.

Liu, B. and Müller, H. (2009). Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics. *Journal of the American Statistical Association*. To appear.

Müller, H., Carey, J., Wu, D., Liedo, P. and Vaupel, J. (2001). Reproductive potential predicts longevity of female mediterranean fruit flies. *Proceedings of the Royal Society B*, **268** 445–450.

Pearson, J. D., Morrell, C. H., Brant, L. J., Landis, P. K. and Fleg, J. L. (1997). Gender differences in a longitudinal study of age associated changes in blood pressure. *Journal of Gerontology: Medical Sciences*, **52** 177–183.

Ramsay, J. and Dalzell, C. J. (1991). Some tools for functional data analysis (with discussion). *Journal of Royal Statistical Society Series B*, **53** 539–572.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis:*

Methods and Case Studies. Sprinter, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Sprinter, New York.

Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57** 253–259.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of Royal Statistical Society Series B*, **53** 233–243.

Shock, N. W., Greulich, R. C., Andres, R., Lakatta, E. G., Arenberg, D. and Tobin, J. D. (1984). Normal human aging: The baltimore longitudinal study of aging. NIH Publication No. 84-2450. Washington, D.C.: U.S. Government Printing Office.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, **24** 1–24.

Staniswalis, J.-G. and Lee, J.-J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of The American Statistian Association*, **93** 1403–1418.

Wahba, G. (1990). *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100** 577–590.

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

Zhang, C., Lu, Y., Johnstone, T., Oaks, T. and Davidson, R. (2007). Efficient modeling and inference for event-related functional mri data. Technical Report.

APPENDIX: AUXILIARY RESULTS AND PROOFS

We note that further details such as omitted proofs can be found in a technical report that is available at `http://www4.stat.ncsu.edu/~wu/WuFanMueller.pdf`.

A bivariate kernel function $\kappa_2(\cdot,\cdot)$ is said to be of order (ν,ℓ) with $\nu=(\nu_1,\nu_2)$ if it satisfies

$$\int u^{\ell_1} v^{\ell_2} \kappa_2(u,v) du dv = \begin{cases} 0 & 0 \le \ell_1 + \ell_2 < \ell, \ell_1 \neq \nu_1, \ell_2 \neq \nu_2 \\ \nu! & \ell_1 = \nu_1, \ell_2 = \nu_2 \\ \neq 0 & \ell_1 + \ell_2 = \ell \end{cases} \tag{A.1}$$

and

$$\int \left| u^{\ell_1} v^{\ell_2} \kappa_2(u,v) \right| du dv < \infty, \text{ for any } \ell_1 + \ell_2 = \ell, \tag{A.2}$$

where $\nu! = \nu_1! \cdot \nu_2!$. Similarly, a univariate kernel function $\kappa_1(\cdot)$ is of order (ν, ℓ) for a univariate $\nu = \nu_1$ when (A.1) and (A.2) hold for $\ell_2 \equiv 0$ on the right hand side while integrating over the univariate argument u on the left.

We enforce the following technical conditions.

(i) The variable S has compact domain $\mathcal{S}$. Given $Z = z$, S has conditional density $f_{S,z}(s)$. Assume, uniformly in $z \in \mathcal{Z}$, that $\frac{\partial^\ell}{\partial s^\ell} f_{S,z}(s)$ exists and is continuous for $\ell = 2$ on $\mathcal{S}$, and further $\inf_{s \in \mathcal{S}} f_{S,z}(s) > 0$, analogously for T.

(ii) Denote the conditional density functions of (S,U) and (T,V) by $g_{X,z}(s,u)$ and $g_{Y,z}(t,v)$, respectively. Assume that the derivative $\dfrac{\partial^\ell}{\partial s^\ell} g_{X,z}(s,u)$ exists for all arguments (s,u), is uniformly continuous on $\mathcal{S} \times I\!R$, and is Lipschitz continuous in z, for $\ell = 2$, analogously for $g_{Y,z}(t,v)$.

(iii) Denote the conditional density functions of quadruples (S_1, S_2, U_1, U_2) and (T_1, T_2, V_1, V_2) by $g_{2X,z}(s_1, s_2, u_1, u_2)$ and $g_{2Y,z}(t_1, t_2, v_1, v_2)$; for simplicity, the corresponding marginal conditional densities of (S_1, S_2) and (T_1, T_2) are also denoted by $g_{2X,z}(s_1, s_2)$ and $g_{2Y,z}(t_1, t_2)$, respectively. Denote the conditional density of (S, T, U, V) given $Z = z$ by $g_{XY,z}(s,t,u,v)$; and similarly its corresponding conditional marginal density of (S,T) by $g_{XY,z}(s,t)$. Assume that the derivatives $\dfrac{\partial^\ell}{\partial s_1^{\ell_1} \partial s_2^{\ell_2}} g_{2X,z}(s_1, s_2, u_1, u_2)$ exist for all arguments (s_1, s_2, u_1, u_2), are uniformly continuous on $\mathcal{S}^2 \times I\!R^2$, and are Lipschitz continuous in z for $\ell_1 + \ell_2 = \ell$, $0 \le \ell_1, \ell_2 \le \ell = 2$, analogously for $g_{2Y,z}(t_1, t_2, v_1, v_2)$ and $g_{XY,z}(s,t,u,v)$.

(iv) For every $p = 1, 2, \cdots, P$, $b_{X,z^{(p)}} \to 0$, $n_{z^{(p)},h} b^4_{X,z^{(p)}} \to \infty$, $n_{z^{(p)},h} b^6_{X,z^{(p)}} < \infty$, $b_{Y,z^{(p)}} \to 0$, $n_{z^{(p)},h} b^4_{Y,z^{(p)}} \to \infty$, and $n_{z^{(p)},h} b^6_{Y,z^{(p)}} < \infty$, as $n \to \infty$.

(v) For every $p = 1, 2, \cdots, P$, $h_{X,z^{(p)}} \to 0$, $n_{z^{(p)},h} h^6_{X,z^{(p)}} \to \infty$, $n_{z^{(p)},h} h^8_{X,z^{(p)}} < \infty$, $h_{Y,z^{(p)}} \to 0$, $n_{z^{(p)},h} h^6_{Y,z^{(p)}} \to \infty$, and $n_{z^{(p)},h} h^8_{Y,z^{(p)}} < \infty$, as $n \to \infty$.

(vi) For every $p = 1, 2, \cdots, P$, $h_{1,z^{(p)}}/h_{2,z^{(p)}} \to 1$, $h_{1,z^{(p)}} \to 0$, $n_{z^{(p)},h} h_{1,z^{(p)}}^6 \to \infty$, and $n_{z^{(p)},h} h_{1,z^{(p)}}^8 < \infty$ as $n \to \infty$.

(vii) For every $p = 1, 2, \cdots, P$, $b_{X,z^{(p)},V} \to 0$, $n_{z^{(p)},h} b_{X,z^{(p)},V}^4 \to \infty$, $n_{z^{(p)},h} b_{X,z^{(p)},V}^6 < \infty$, $b_{Y,z^{(p)},V} \to 0$, $n_{z^{(p)},h} b_{Y,z^{(p)},V}^4 \to \infty$, and $n_{z^{(p)},h} b_{Y,z^{(p)},V}^6 < \infty$, as $n \to \infty$.

(viii) Univariate kernel κ_1 and bivariate κ_2 are compactly supported, absolutely integrable, and of order $(\nu, \ell) = (0, 2)$ and $((0, 0), 2)$, respectively.

(ix) Assume $\sup_{(z,s)\in\mathcal{Z}\times\mathcal{S}} E(E(X(s) - \mu_{X,Z}(s))^4 | Z = z)) < \infty$ and analogously for Y.

(x) The slope function $\beta(z, s, t)$ is twice differentiable in z, *i.e.*, for any $(s, t) \in \mathcal{S} \times \mathcal{T}$, $\frac{\partial^2}{\partial z^2}\beta(z, s, t)$ exists and is continuous in z.

(xi) The bin width h and smoothing bandwidth b satisfy that $b/h < \infty$ as $n \to \infty$. The bin width h is chosen such that $P \propto n^{1/8}$.

Proposition 1. *For E_n defined in (3.1), under (xi) it holds that $P(E_n) \to 1$ as $n \to \infty$.*

Proof of Proposition 1. Note first that $P(\min n_{z^{(p)},h} > \tilde{n}) \geq 1 - \sum_{p=1}^P P(n_{z^{(p)},h} < \tilde{n})$. Consider the pth bin, denote $\pi_p = P(Z \in [z^{(p)} - \frac{h}{2}, z^{(p)} - \frac{h}{2}))$. Then $n_{z^{(p)},h}$ is asymptotically distributed as $N(n\pi_p, n\pi_p(1-\pi_p))$ due to the normal approximation to a binomial random variable. Thus $P(n_{z^{(p)},h} > \tilde{n}) \to f_{N(0,1)}(a_p)/a_p$ with $a_p = -(\tilde{n} - n\pi_p)/\sqrt{n\pi_p(1-\pi_p)}$, where $f_{N(0,1)}(\cdot)$ is the probability density function of the standard normal distribution. Due to [A1], π_p is bounded between $\underline{f_Z}/(\underline{f_Z} + (P-1)\bar{f_Z})$ and $\bar{f_Z}/((P-1)\underline{f_Z} + \bar{f_Z})$. It follows that $P(E_n) \to 1$ as $n \to \infty$ by noting that $\tilde{n} \propto \sqrt{n}$, $P \propto n^{1/8}$, and $f_{N(0,1)}(x)/x$ decays exponentially in x. □

We next prove the consistency of raw estimate of the mean functions of predictor and response trajectories within each bin. Consider a generic bin $[z - h/2, z + h/2)$, with bin center z and bandwidth h and let $b_{X,z}$ and $b_{Y,z}$ be smoothing bandwidths used to estimate $\mu_{X,z}(s)$ and $\mu_{Y,z}(t)$, $h_{X,z}$ and $h_{Y,z}$ for $G_{X,z}(s_1, s_2)$ and $G_{Y,z}(t_1, t_2)$, $h_{1,z}$ and $h_{2,z}$ for $C_{XY,z}(s, t)$, $b_{X,z,V}$ and $b_{Y,z,V}$ for $V_{X,z}(s) = G_{X,z}(s, s) + \sigma_X^2$ and $V_{Y,z}(t) = G_{Y,z}(t, t) + \sigma_Y^2$.

For a positive integer $l \geq 1$, let $\{\psi_p(t, v), p = 1, 2, \cdots, l\}$ be a collection of real functions $\psi_p : \mathbb{R}^2 \to \mathbb{R}$ satisfying the following conditions:

[C1.1a] The derivative functions $\dfrac{\partial^\ell}{\partial t^\ell}\psi_p(t, v)$ exist for all arguments (t, v) and are uniformly continuous on $\mathcal{T} \times \mathbb{R}$.

[C1.2a] Assume that $\int\int \psi_p^2(t, v) g_{Y,z}(t, v) dv dt < \infty$.

[C2.1a] Uniformly in $z \in \mathcal{Z}$, bandwidths $b_{Y,z}$ for one dimensional smoothers satisfy that $b_{Y,z} \to 0$, $n_{z,h} b_{Y,z}^{\nu+1} \to \infty$, and $n_{z,h} b_{Y,z}^{2\ell+2} < \infty$, as $n \to \infty$.

Define $\mu_{p\psi,z} = \mu_{p\psi,z}(t) = \frac{d^\nu}{dt^\nu} \int \psi_p(t,v) g_{Y,z}(t,v) dv$ and

$$\Psi_{pn,z} \;=\; \Psi_{pn,z}(t) = \frac{1}{n_{z,h} b_{Y,z}^{\nu+1}} \sum_{i \in \mathcal{N}_{z,h}} \frac{1}{EN} \sum_{j=1}^{N_i} \psi_p(T_{ij}, V_{ij}) \kappa_1(\frac{T_{ij}-t}{b_{Y,z}}),$$

where $g_{Y,z}(t,v)$ is the conditional density of (T,V), given $Z=z$.

Lemma 1. *Under Conditions [A0-A3] (i), (ii), (viii), [C1.1a], [C1.2a], and [C2.1a], we have* $\tau_{pn} = \sup_{(z,t) \in \mathcal{Z} \times \mathcal{T}} \mid \Psi_{pn,z}(t) - \mu_{p\psi,z}(t) \mid / (h + (\sqrt{n_{z,h}} b_{Y,z}^{\nu+1})^{-1}) = O_p(1)$.

Proof. Note that $\mid \Psi_{pn,z}(t) - \mu_{p\psi,z}(t) \mid \leq \mid \Psi_{pn,z}(t) - E\Psi_{pn,z}(t) \mid + \mid E\Psi_{pn,z}(t) - \mu_{p\psi,z}(t) \mid$ and $E \mid \tau_{pn} \mid = O(1)$ implies $\tau_{pn} = O_p(1)$. Standard conditioning techniques lead to

$$E\Psi_{pn,z}(t) \;=\; \frac{1}{b_{Y,z}^{\nu+1}} E(E(\psi_p(T_{i1}, V_{i1}) \kappa_1(\frac{T_{i1}-t}{b_{Y,z}}) \mid z - \frac{h}{2} \leq Z_i < \frac{h}{2})).$$

For $Z_i = z_i \in [z-h/2, z+h/2)$, perform a Taylor expansion of order ℓ on the integrand,

$$\begin{aligned} & E[\psi_p(T_{i1}, V_{i1}) \kappa_1(\frac{T_{i1}-t}{b_{Y,z}})] = \int\int \psi_p(t_1, v_1) g_{Y,z_i}(t_1, v_1) \kappa_1(\frac{t_1-t}{b_{Y,z}}) dt_1 dv_1 \\ = \;& \int\int \left(\frac{\partial^\nu}{\partial t^\nu} (\psi_p(t, v_1) g_{Y,z_i}(t, v_1)) \right) \frac{(t_1-t)^\nu}{\nu!} \kappa_1(\frac{t_1-t}{b_{Y,z}}) dt_1 dv_1 \\ & + \int\int \left(\frac{\partial^\ell}{\partial t^\ell} (\psi_p(t, v_1) g_{Y,z_i}(t, v_1)) \right) \mid_{t=t^*} \frac{(t_1-t)^\ell}{\ell!} \kappa_1(\frac{t_1-t}{b_{Y,z}}) dt_1 dv_1, \end{aligned}$$

where t^* is between t and t_1. Hence, $\left| E[\psi_p(T_{i1}, V_{i1}) \kappa_1(\frac{T_{i1}-t}{b_{Y,z}})] - \mu_{p\psi,z_i}(t) b_{Y,z}^{\nu+1} \right| \leq c_0 \frac{b_{Y,z}^{\ell+1}}{\ell!} \int \left| u^\ell \kappa_1(u) \right| du$ due to [C1.2a] and the assumption that the kernel function $\kappa_1(\cdot)$ is of type (ν, ℓ), where c_0 is bounded according to [C1.1a], $c_0 \leq \sup_{(z_i,t) \in \mathcal{Z} \times \mathcal{T}} \left| \frac{\partial^\ell}{\partial t^\ell} \int \psi_p(t, v_1) g_{Y,z_i}(t, v_1) dv_1 \right| < \infty$. Furthermore, using (ii), we may bound

$$\begin{aligned} & \sup_{t \in \mathcal{T}} \mid E\Psi_{pn,z}(t) - \mu_{p\psi,z}(t) \mid \\ \leq \;& c_0 b_{Y,z}^{\ell-\nu} / (\ell!) \int \left| u^\ell \kappa_1(u) \right| du + E \left\{ E \left[\sup_{t \in \mathcal{T}} |\mu_{p\psi,Z_i}(t) - \mu_{p\psi,z}(t)| \mid z - \frac{h}{2} \leq Z_i < \frac{h}{2} \right] \right\} \\ \leq \;& c_0 (\int \left| u^\ell \kappa_1(u) \right| du) b_{Y,z}^{\ell-\nu} / (\ell!) + c_1 h, \end{aligned} \tag{A.3}$$

where the constants do not depend on z. To bound $E \sup_{t \in \mathcal{T}} \mid \Psi_{pn,z}(t) - E\Psi_{pn,z}(t) \mid$, we denote Fourier transform of $\kappa_1(\cdot)$ by $\zeta_1(t) = \int e^{-iut} \kappa_1(u) du$ and letting

$\varphi_{pn,z}(u) = \frac{1}{n_{z,h}} \sum_{m\in\mathcal{N}_{z,h}} \frac{1}{EN} \sum_{j=1}^{N_m} e^{iuT_{mj}} \psi_p(T_{mj}, Y_{mj})$,

$$\Psi_{pn,z} = \frac{1}{n_{z,h} b_{Y,z}^{\nu+1}} \sum_{m\in\mathcal{N}_{z,h}} \frac{1}{EN} \sum_{j=1}^{N_m} \kappa_1(\frac{T_{mj}-t}{b_{Y,z}}) \psi_p(T_{mj}, Y_{mj}) = \frac{1}{2\pi b_{Y,z}^{\nu}} \int \varphi_{pn,z}(u) e^{-itu} \zeta_1(u b_{Y,z}) du,$$

and $\sup_{t\in\mathcal{T}} \mid \Psi_{pn,z}(t) - E\Psi_{pn,z}(t) \mid \leq \frac{1}{2\pi b_{Y,z}^{\nu}} \int \mid \varphi_{pn,z}(u) - E\varphi_{pn,z}(u) \mid \cdot \mid \zeta_1(u b_{Y,z}) \mid du$.

Decomposing $\varphi_{pn,z}(\cdot)$ into real and imaginary parts,

$$\begin{aligned} \varphi_{pn,z,R}(u) &= \frac{1}{n_{z,h}} \sum_{m\in\mathcal{N}_{z,h}} \frac{1}{EN} \sum_{j=1}^{N_m} \cos(uT_{mj}) \psi_p(T_{mj}, Y_{mj}) \\ \varphi_{pn,z,I}(u) &= \frac{1}{n_{z,h}} \sum_{m\in\mathcal{N}_{z,h}} \frac{1}{EN} \sum_{j=1}^{N_m} \sin(uT_{mj}) \psi_p(T_{mj}, Y_{mj}), \end{aligned}$$

we obtain $E \mid \varphi_{pn,z}(u) - E\varphi_{pn,z}(u) \mid = E \mid \varphi_{pn,z,R}(u) - E\varphi_{pn,z,R}(u) \mid + E \mid \varphi_{pn,z,I}(u) - E\varphi_{pn,z,I}(u) \mid$. Note the inequality $E \mid \varphi_{pn,z,R}(u) - E\varphi_{pn,z,R}(u) \mid \leq \sqrt{E \mid \varphi_{pn,z,R}(u) - E\varphi_{pn,z,R}(u) \mid^2}$ and the fact that $\{[Z_i, N_i, (T_{ij}, Y_{ij})_{j=1}^{N_i}] : i \in \mathcal{N}_{z,h}\}$ are *i.i.d.* implies that

$$\mathrm{var}(\varphi_{pn,z,R}(u)) \leq \frac{1}{n_{z,h}} E\{E(\psi_p^2(T_{m1}, Y_{m1}) | z - h/2 \leq Z_m < z + h/2)\}$$

where $m \in N_{z,h}$, analogously for the imaginary part. As a result, we have

$$E \sup_{t\in\mathcal{T}} \mid \Psi_{pn,z}(t) - E\Psi_{pn,z}(t) \mid \leq \frac{2\sqrt{E\{E(\psi_p^2(T_{m1}, Y_{m1}) | z - h/2 \leq Z_m < z + h/2)\}} \int \mid \zeta_1(u) \mid du}{2\pi \sqrt{n_{z,h}} b_{Y,z}^{\nu+1}}.$$

Note that $E(\psi_p^2(T_{m1}, Y_{m1}))$ as a function of Z_m is continuous over the compact domain $\mathcal{Z}$ and as a result bounded. Denote $c_2 = 2 \sup_{Z_m\in\mathcal{Z}} \sqrt{E(\psi_p^2(T_{m1}, Y_{m1}))} < \infty$. Hence we have

$$E \sup_{t\in\mathcal{T}} \mid \Psi_{pn,z}(t) - E\Psi_{pn,z}(t) \mid \leq \frac{c_2 \int |\zeta_1(u)| \, du}{2\pi} (\sqrt{n_{z,h}} b_{Y,z}^{\nu+1})^{-1}, \tag{A.4}$$

where the constant $c_2(\int |\zeta_1(u)| \, du)/(2\pi)$ does not depend on z.

The result follows as Condition [A1] implies that $n_{z,h}$ goes to infinity uniformly for $z \in \mathcal{Z}$ as $n \to \infty$, $n_{z,h} b_{Y,z}^{2\ell+2} < \infty$ implies that $b_{Y,z}^{\ell-\nu} = O(1/(\sqrt{n_{z,h}} b_{Y,z}^{\nu+1}))$. We next extend Theorem 1 in Yao et al. (2005a) under some additional conditions. □

[C3] Uniformly in $z \in \mathcal{Z}$, $b_{X,z} \to 0$, $n_{z,h} b_{X,z}^4 \to \infty$, $n_{z,h} b_{X,z}^6 < \infty$, $b_{Y,z} \to 0$, $n_{z,h} b_{Y,z}^4 \to \infty$, and $n_{z,h} b_{Y,z}^6 < \infty$, as $n \to \infty$.

Lemma 2. *Under Conditions [A0-A3], (i), (ii), (viii), (ix), and [C3], we have*

$$\sup_{(z,s)\in\mathcal{Z}\times\mathcal{S}} \frac{\mid \tilde{\mu}_{X,z}(s) - \mu_{X,z}(s) \mid}{h + (\sqrt{n_{z,h}} b_{X,z})^{-1}} = O_p(1) \text{ and } \sup_{(z,t)\in\mathcal{Z}\times\mathcal{T}} \frac{\mid \tilde{\mu}_{Y,z}(t) - \mu_{Y,z}(t) \mid}{h + (\sqrt{n_{z,h}} b_{Y,z})^{-1}} = O_p(1) \tag{A.5}$$

Proof. The proof is similar to the proof of Theorem 1 in Yao et al. (2005a). □

Our next two lemmas concern the consistency for estimating the covariance functions, based on the observations in the generic bin $[z-h/2, z+h/2)$. Let $\{\theta_p(r_1, r_2, v_1, v_2), p = 1, 2, \cdots, l\}$ be a collection of real functions $\theta_p : I\!R^4 \to I\!R$ with the following properties:

[C1.1b] The derivatives $\dfrac{\partial^\ell}{\partial r_1^{\ell_1} \partial r_2^{\ell_2}} \theta_p(r_1, r_2, v_1, v_2)$ exist for all arguments (r_1, r_2, v_1, v_2) and are uniformly continuous on $\mathcal{R}_1 \times \mathcal{R}_2 \times I\!R^2$, for $\ell_1 + \ell_2 = \ell$, $0 \le \ell_1, \ell_2 \le \ell$, $\ell = 0, 1, 2$.

[C1.2b] The expectation $\int\int\int\int \theta_p^2(r_1, r_2, v_1, v_2) g(r_1, r_2, v_1, v_2) dr_1 dr_2 dv_1 dv_2$ exists and is finite, uniformly bounded on $\mathcal{Z}$.

[C2.1b] Uniformly in $z \in \mathcal{Z}$, bandwidths $h_{Y,z}$ for the two-dimensional smoother satisfy $h_{Y,z} \to 0$, $n_{z,h} h_{Y,z}^{|\nu|+2} \to \infty$, $n_{z,h} h_{Y,z}^{2\ell+4} < \infty$, as $n \to \infty$.

Define $\varrho_{p\theta,z} = \varrho_{p\theta,z}(t_1, t_2) = \frac{\partial^{|\nu|}}{\partial t_1^{\nu_1} \partial t_2^{\nu_2}} \int\int \theta_p(t_1, t_2, v_1, v_2) g_{2Y,z}(t_1, t_2, v_1, v_2) dv_1 dv_2$ and

$$\Theta_{pn,z}(t_1, t_2) = \frac{1}{n_{z,h} h_{Y,z}^{|\nu|+2}} \sum_{i \in \mathcal{N}_{z,h}} \frac{1}{EN(EN-1)} \sum_{1 \le j \ne k \le N_i} \theta_p(T_{ij}, T_{ik}, V_{ij}, V_{ik}) \kappa_2\left(\frac{T_{ij} - t_1}{h_{Y,z}}, \frac{T_{ik} - t_2}{h_{Y,z}}\right).$$

Lemma 3. *Under Conditions [A0-A3], (i), (ii), (iii), (viii), [C1.1b] with $\mathcal{R}_1 = \mathcal{T}$ and $\mathcal{R}_2 = \mathcal{T}$, [C1.2b] with $g(\cdot, \cdot, \cdot, \cdot) = g_{2Y,z}(\cdot, \cdot, \cdot, \cdot)$, and [C2.1b], we have*

$$\vartheta_{pn} = \sup_{(z,t_1,t_2) \in \mathcal{Z} \times \mathcal{T} \times \mathcal{T}} \frac{| \Theta_{pn,z} - \varrho_{p\theta,z} |}{h + (\sqrt{n_{z,h}} h_{Y,z}^{|\nu|+2})^{-1}} = O_p(1).$$

Proof. Analogous to the proof of Lemma 1. □

[C4] Uniformly in $z \in \mathcal{Z}$, $h_{X,z} \to 0$, $n_{z,h} h_{X,z}^6 \to \infty$, $n_{z,h} h_{X,z}^8 < \infty$, $h_{Y,z} \to 0$, $n_{z,h} h_{Y,z}^6 \to \infty$, and $n_{z,h} h_{Y,z}^8 < \infty$, as $n \to \infty$.

The proof of the next result is omitted.

Lemma 4. *Under Conditions [A0-A3], (i-iii), (viii), (ix), [C3], and [C4], we have*

$$\sup_{(z,s_1,s_2) \in \mathcal{Z} \times \mathcal{S}^2} \frac{| \tilde{G}_{X,z}(s_1, s_2) - G_{X,z}(s_1, s_2) |}{(h + (\sqrt{n_{z,h}} h_{X,z}^2)^{-1})} = O_p(1) \tag{A.6}$$

$$\sup_{(z,t_1,t_2) \in \mathcal{Z} \times \mathcal{T}^2} \frac{| \tilde{G}_{Y,z}(t_1, t_2) - G_{Y,z}(t_1, t_2) |}{(h + (\sqrt{n_{z,h}} h_{Y,z}^2)^{-1})} = O_p(1) \tag{A.7}$$

To estimate variance of the measurement errors, as in Yao et al. (2005a), we first estimate $G_{X,z}(s, s) + \sigma_X^2$ (resp. $G_{Y,z}(t, t) + \sigma_Y^2$) using a local linear smoother based on $G_{X,i,z}(S_{il}, S_{il})$ for $l = 1, 2, \cdots, L_i$, $i \in \mathcal{N}_{z,h}$ (resp. $G_{Y,i,z}(T_{ij}, T_{ij})$ for $j = 1, 2, \cdots, N_i$, $i \in$

$\mathcal{N}_{z,h}$) with smoothing bandwidth $b_{X,z,V}$ (resp. $b_{Y,z,V}$) and denote the estimates by $\tilde{V}_{X,z}(s)$ (resp. $\tilde{V}_{Y,z}(t)$), removing the two ends of the interval $\mathcal{S}$ (resp. $\mathcal{T}$) to get more stable estimates of σ_X^2 (res. σ_Y^2). Denote the estimates based on the generic bin $[z-h/2, z+h/2)$ by $\tilde{\sigma}_{X,z}^2$ and $\tilde{\sigma}_{Y,z}^2$, $|\mathcal{S}|$ the length of $\mathcal{S}$ and $\mathcal{S}_1 = [\inf\{s : s \in \mathcal{S}\} + |\mathcal{S}|/4, \sup\{s : s \in \mathcal{S}\} - |\mathcal{S}|/4]$. Then

$$\tilde{\sigma}_{X,z}^2 = \frac{2}{|\mathcal{S}|}\int_{\mathcal{S}_1}[\tilde{V}_X(s) - \tilde{G}_{X,z}(s,s)]ds.$$

and analogously for $\tilde{\sigma}_{Y,z}^2$. Lemmas 2 and 4 imply the convergence of $\tilde{\sigma}_{X,z}^2$ and $\tilde{\sigma}_{Y,z}^2$, as stated in Corollary 1.

[C5] Uniformly in $z \in \mathcal{Z}$, $b_{X,z,V} \to 0$, $n_{z,h}b_{X,z,V}^4 \to \infty$, $n_{z,h}b_{X,z,V}^6 < \infty$, $b_{Y,z,V} \to 0$, $n_{z,h}b_{Y,z,V}^4 \to \infty$, and $n_{z,h}b_{Y,z,V}^6 < \infty$, as $n \to \infty$.

Corollary 1. *Under Condition [C5] and the conditions of Lemmas 2 and 4,*

$$\sup_{z\in\mathcal{Z}}\left|\tilde{\sigma}_{X,z}^2 - \sigma_X^2\right|/(h + (\sqrt{n_{z,h}}b_{X,z,V})^{-1} + (\sqrt{n_{z,h}}h_{X,z}^2)^{-1}) = O_p(1)$$ *and analogously for* $\tilde{\sigma}_{X,z}^2$.

Proposition 2. *Under Conditions [A0-A3] in Section 2 and (i-ix), the final estimates of* σ_X^2 *and* σ_Y^2 *(2.8) converge in probability to their corresponding true counterparts, i.e.,*

$$\hat{\sigma}_X^2 \xrightarrow{P} \sigma_X^2, \qquad \hat{\sigma}_Y^2 \xrightarrow{P} \sigma_Y^2.$$

Proof of Proposition 2. The result follows straightforwardly from Corollary 1. □

While Lemma 3 implies consistency of the estimator of the variance, we also require an extension regarding estimation of the cross-covariance function. Let $\{\tilde{\theta}_p(s,t,u,v), p = 1, 2, \cdots, l\}$ be a collection of real functions $\tilde{\theta}_p : \mathbb{R}^4 \to \mathbb{R}$.

[C2.1c] For $\ell \geq |\nu|+2$ and any pair of ℓ_1 and ℓ_2 such that $\ell = \ell_1+\ell_2$, $\ell_1 \geq \nu_1+1$, $\ell_2 \geq \nu_2+1$. Uniformly in $z \in \mathcal{Z}$, bandwidth $h_{1,z}$ and $h_{2,z}$ satisfy $h_{1,z} \to 0$, $h_{1,z}/h_{2,z} \to 1$, $n_{z,h}h_{1,z}^{|\nu|+2} \to \infty$, $n_{z,h}h_{1,z}^{2\ell+4} < \infty$, as $n \to \infty$.

Define $\varrho_{p\tilde{\theta},z} = \varrho_{p\tilde{\theta},z}(s,t) = \frac{\partial^{|\nu|}}{\partial s^{\nu_1}\partial t^{\nu_2}}\int\int \tilde{\theta}_p(s,t,u,v)g_{XY,z}(s,t,u,v)dudv$ and

$$\tilde{\Theta}_{pn,z} \;=\; \tilde{\Theta}_{pn,z}(s,t) = \frac{1}{n_{z,h}h_{1,z}^{\nu_1+1}h_{2,z}^{\nu_2+1}}\sum_{i\in\mathcal{N}_{z,h}}\frac{1}{EN}\sum_{1\leq j\leq N_i}\tilde{\theta}_p(S_{ij},T_{ij},U_{ij},V_{ij})\kappa_2(\frac{S_{ij}-s}{h_{1,z}},\frac{T_{ij}-t}{h_{2,z}}).$$

Lemma 5. *Under Conditions [A0-A3], (i), (ii), (iii), (viii), [C1.1b] with* $\mathcal{R}_1 = \mathcal{S}$ *and* $\mathcal{R}_2 = \mathcal{T}$*, [C1.2b] with* $g(\cdot,\cdot,\cdot,\cdot) = g_{XY,z}(\cdot,\cdot,\cdot,\cdot)$*, and [C2.1c] (with* $\ell_1 = \ell_2 = 1$ *and* $\nu_1 = \nu_2 = 0$*), we have* $\tilde{\vartheta}_{pn} = \sup_{(z,s,t)\in\mathcal{Z}\times\mathcal{S}\times\mathcal{T}} \mid \tilde{\Theta}_{pn,z}(s,t) - \varrho_{p\tilde{\theta},z}(s,t) \mid /(h + (\sqrt{n_{z,h}} h_{Y,1}^{\nu_1+1} h_{Y,2}^{\nu_2+1})^{-1}) = O_p(1)$.

Proof. The proof is analogous to that of Lemmas 1 and 3. □

[C6] Uniformly in $z \in \mathcal{Z}$, bandwidth $h_{1,z}$ and $h_{2,z}$ satisfy $h_{1,z} \to 0$, $h_{1,z}/h_{2,z} \to 1$, $n_{z,h} h_{1,z}^6 \to \infty$, $n_{z,h} h_{1,z}^8 < \infty$, as $n \to \infty$.

Lemma 6 (Convergence of convergence of the cross-covariance function between X and Y)**.** *Under Condition [A0-A3], (i), (ii), (iii), (viii), (ix), [C3], and [C6]*

$$\sup_{(z,s,t)\in\mathcal{Z}\times\mathcal{S}\times\mathcal{T}} \mid \tilde{C}_{XY,z}(s,t) - C_{XY,z}(s,t) \mid /(h + (\sqrt{n_{z,h}} h_{1,z} h_{2,z})^{-1}) = O_p(1).$$

Proof. The proof is similar to that of Lemma 4. □

Consider the real separable Hilbert space $L_Y^2(\mathcal{T}) \equiv H_Y$ (resp. $L_X^2(\mathcal{S}) \equiv H_X$) endowed with inner product $\langle f, g\rangle_{H_Y} = \int_{\mathcal{T}} f(t)g(t)dt$ (resp. $\langle f, g\rangle_{H_X} = \int_{\mathcal{S}} f(s)g(s)ds$) and norm $\| f \|_{H_X} = \sqrt{\langle f, f\rangle_{H_X}}$ (resp. $\| f \|_{H_Y} = \sqrt{\langle f, f\rangle_{H_Y}}$) (Courant and Hilbert, 1953). Let $\mathcal{I}'_{Y,z}$ (resp. $\mathcal{I}'_{X,z}$) be the set of indices of the eigenfunctions $\phi_{z,k}(t)$ (resp. $\psi_{z,m}(s)$) corresponding to eigenvalues $\lambda_{z,k}$ (resp. $\rho_{z,m}$) of multiplicity one. We obtain the consistency of $\tilde{\lambda}_{z,k}$ (resp. $\tilde{\rho}_{z,m}$) for $\lambda_{z,k}$ (resp. $\rho_{z,m}$), the consistency of $\tilde{\phi}_{z,k}(t)$ (resp. $\tilde{\psi}_{z,m}(s)$) for $\phi_{z,k}(t)$ (resp. $\psi_{z,m}(s)$) in the L_Y^2 (resp. L_X^2) norm $\| \cdot \|_{H_X}$ (resp. $\| \cdot \|_{H_Y}$) when $\lambda_{z,k}$ (resp. $\rho_{z,m}$) is of multiplicity one, and the uniform consistency of $\tilde{\phi}_{z,k}(t)$ (resp. $\tilde{\psi}_{z,m}(s)$) for $\phi_{z,k}(t)$ (resp. $\psi_{z,m}(s)$) as well.

For $f, g, h \in H_Y$, define the rank one operator $f \otimes g : h \to \langle f, h\rangle g$. Denote the separable Hilbert space of Hilbert-Schmidt operators on H_Y by $F_Y \equiv \sigma_2(H_Y)$, endowed with $\langle T_1, T_2\rangle_{F_Y} = tr(T_1 T_2^*) = \sum_j \langle T_1 u_j, T_2 u_j\rangle_{H_Y}$ and $\| T \|_{F_Y}^2 = \langle T, T\rangle_{F_Y}$, where T_1, T_2, $T \in F_Y$, T_2^* is the adjoint of T_2, and $\{u_j : j \geq 1\}$ is any complete orthonormal system in H_Y. The covariance operator $\mathbf{G}_{Y,z}$ (resp. $\tilde{\mathbf{G}}_{Y,z}$) is generated by the kernel $G_{Y,z}$ (resp. $\tilde{G}_{Y,z}$), *i.e.*, $\mathbf{G}_{Y,z}(f) = \int_{\mathcal{T}} G_{Y,z}(t_1, t) f(t_1) dt_1$ (resp. $\tilde{\mathbf{G}}_{Y,z}(f) = \int_{\mathcal{T}} \tilde{G}_{Y,z}(t_1, t) f(t_1) dt_1$). Obviously, $\mathbf{G}_{Y,z}$ and $\tilde{\mathbf{G}}_{Y,z}$ are Hilbert-Schmidt operators. As a result of (A.7), we have $\sup_{z\in\mathcal{Z}} \| \tilde{\mathbf{G}}_{Y,z} - \mathbf{G}_{Y,z} \|_{F_Y} /(h + (\sqrt{n_{z,h}} h_{Y,z}^2)^{-1}) = O_p(1)$.

Let $\mathcal{I}_{Y,z,i} = \{j : \lambda_{z,j} = \lambda_{z,i}\}$ and $\mathcal{I}'_{Y,z} = \{i : |\mathcal{I}_{Y,z,i}| = 1\}$, where $|\mathcal{I}_{Y,z,i}|$ denotes the number of elements in $\mathcal{I}_{Y,z,i}$. Denote $\mathbf{P}^Y_{z,j} = \sum_{k\in\mathcal{I}_{Y,z,j}} \phi_{z,k}\otimes\phi_{z,k}$ and $\tilde{\mathbf{P}}^Y_{z,j} = \sum_{k\in\mathcal{I}_{Y,z,j}} \tilde{\phi}_{z,k}\otimes \tilde{\phi}_{z,k}$ to be the true and estimated orthogonal projection operators from H_Y to the subspace spanned by $\{\phi_{z,k} : k \in \mathcal{I}_{Y,z,j}\}$. Set $\delta^Y_{z,j} = \frac{1}{2}\min\{|\lambda_{z,l} - \lambda_{z,j}| : l \notin \mathcal{I}_{Y,z,j}\}$ and $\mathbf{\Lambda}_{\delta^Y_{z,j}} = \{c \in \mathcal{C} : |c - \lambda_{z,j}| = \delta^Y_{z,j}\}$, where $\mathcal{C}$ stands for the complex numbers. Let $\mathbf{R}_{Y,z}$ (resp. $\tilde{\mathbf{R}}_{Y,z}$) to be the resolvent of $\mathbf{G}_{Y,z}$ (resp. $\tilde{\mathbf{G}}_{Y,z}$), *i.e.*, $\mathbf{R}_{Y,z}(c) = (\mathbf{G}_{Y,z} - cI)^{-1}$ (resp. $\tilde{\mathbf{R}}_{Y,z}(c) = (\tilde{\mathbf{G}}_{Y,z} - cI)^{-1}$). Denote $A_{\delta^Y_{z,j}} = \sup\{\| \mathbf{R}_{Y,z}(c) \|_{F_Y} : c \in \mathbf{\Lambda}_{\delta^Y_{z,j}}\}$ and

$$\alpha_X = \Big(\delta^X_{z,j}(A_{\delta^X_{z,j}})^2\Big)/\Big((h + (\sqrt{n_{z,h}}h^2_{X,z})^{-1})^{-1} - A_{\delta^X_{z,j}}\Big). \tag{A.8}$$

Parallel notations are made for the Y process.

Proposition 3. *Under Conditions [A0-A3] in Section 2, and Conditions (i-iii), (viii), (ix), [C3], [C4], and [C6], it holds that*

$$\begin{aligned}
|\tilde{\rho}_{z,m} - \rho_{z,m}| &= O_p(\alpha_X) && && \text{(A.9)}\\
\| \tilde{\psi}_{z,m} - \psi_{z,m} \|_{H_X} &= O_p(\alpha_X), && m \in \mathcal{I}'_{X,z} && \text{(A.10)}\\
\sup_{s\in\mathcal{S}} \left|\tilde{\psi}_{z,m}(s) - \psi_{z,m}(s)\right| &= O_p(\alpha_X), && m \in \mathcal{I}'_{X,z} && \text{(A.11)}
\end{aligned}$$

$$\begin{aligned}
\left|\tilde{\lambda}_{z,k} - \lambda_{z,k}\right| &= O_p(\alpha_Y) && && \text{(A.12)}\\
\| \tilde{\phi}_{z,k} - \phi_{z,k} \|_{H_Y} &= O_p(\alpha_Y), && k \in \mathcal{I}'_{Y,z} && \text{(A.13)}\\
\sup_{t\in\mathcal{T}} \left|\tilde{\phi}_{z,k}(t) - \phi_{z,k}(t)\right| &= O_p(\alpha_Y), && k \in \mathcal{I}'_{Y,z}, && \text{(A.14)}
\end{aligned}$$

$$|\tilde{\sigma}_{z,mk} - \sigma_{z,mk}| = O_p(\max(\alpha_X, \alpha_Y, h + (\sqrt{n_{z,h}}h_{1,z}h_{2,z})^{-1})), \tag{A.15}$$

where the norms on H_X and H_Y are defined on page 29, both α_X, α_Y are defined in (A.8) and converge to zero as $n \to \infty$, and the above O_p terms are uniform in $z \in \mathcal{Z}$.

Proof of Proposition 3. The proof is similar to the proof of Theorem 2 in Yao et al. (2005a). The uniformity result follows from that of Lemmas 4 and 6. □

Note that

$$\beta(z, s, t) = \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \frac{E(\zeta_{z,m}\xi_{z,k})}{E(\zeta^2_{z,m})}\psi_{z,m}(s)\phi_{z,k}(t). \tag{A.16}$$

To define the convergence of the right hand side of (A.16) in the L_2 sense in (s,t) and uniformly in z, we require that

[A4] $\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sigma_{z,mk}^2/\rho_{z,m}^2 < \infty$ uniformly for $z \in \mathcal{Z}$.

The proof of the following result is straightforward.

Lemma 7. *Under Condition [A4], uniformly in $z \in \mathcal{Z}$, the right hand side of (A.16) converges in the L_2 sense.*

The next result is stated without proof and requires assumptions [A4] and the following

$$[A5] \qquad \begin{array}{rcll} \sum_{m=1}^{M(n)} \dfrac{\delta_{z,m}^X (A_{\delta_{z,m}^X})^2}{(h+(\sqrt{n_{z,h}}h_{X,z}^2)^{-1})^{-1} - A_{\delta_{z,m}^X}} & \to & 0 & \\ \sum_{k=1}^{K(n)} \dfrac{\delta_{z,k}^Y (A_{\delta_{z,k}^Y})^2}{(h+(\sqrt{n_{z,h}}h_{Y,z}^2)^{-1})^{-1} - A_{\delta_{z,k}^Y}} & \to & 0 & \text{uniformly in } z \in \mathcal{Z}. \\ MK(h+(\sqrt{n_{z,h}}h_{1,z}h_{2,z})^{-1}) & \to & 0 & \end{array}$$

Lemma 8. *Under conditions of Proposition 3, [A4], and [A5],*

$$\lim_{n\to\infty}\sup_{z\in\mathcal{Z}}\int_{\mathcal{S}}\int_{\mathcal{T}}[\tilde{\beta}(z,s,t)-\beta(z,s,t)]^2 = 0, \qquad \textit{in probability.} \tag{A.17}$$

Proof of Theorem 1. We consider only the convergence of $\hat{\beta}(z,s,t)$. The consistency of $\hat{\mu}_{X,z}(s)$ and $\hat{\mu}_{Y,z}(t)$ is analogous. Note first that

$$\begin{aligned} & \int_{\mathcal{T}}\int_{\mathcal{S}}(\hat{\beta}(z,s,t)-\beta(z,s,t))^2 dsdt \\ \le\ & 2(2b/h+1)\sum_{p=1}^{P}\omega_{0,2}(z^{(p)},z,b)^2\int_{\mathcal{T}}\int_{\mathcal{S}}(\tilde{\beta}(z^{(p)},s,t)-\beta(z^{(p)},s,t))^2 dsdt \\ & +2\int_{\mathcal{T}}\int_{\mathcal{S}}(\sum_{p=1}^{P}\omega_{0,2}(z^{(p)},z,b)\beta(z^{(p)},s,t)-\beta(z,s,t))^2 dsdt, \end{aligned} \tag{A.18}$$

where $2b/h+1$ of the very last inequality is due to the fact that the kernel function $K(\cdot)$ is of bounded support $[-1,1]$. Denote $a(k)=\sum_{p=1}^{P}K_b(z^{(p)}-z)(z^{(p)}-z)^k$, $b(k)=\sum_{p=1}^{P}K_b(z^{(p)}-z)^2(z^{(p)}-z)^k$, $\mu_k=\int K(u)u^k du$ and $\nu_k=\int(K(u))^2u^k du$. Then we have

$$a(k)=\mu_k\frac{b^k}{h}(1+o(1)),\ \text{and}\ b(k)=\nu_k\frac{b^{k-1}}{h}(1+o(1))$$

for small h (large $P\propto 1/h$) and small b. Moreover the usual boundary techniques can be applied near the two end points. Consequently

$$\begin{aligned} & \sum_{p=1}^{P}\omega_{0,2}(z^{(p)},z,b)^2 = \mathbf{e}_{1,2}^T(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}(\mathbf{C}^T\mathbf{W}\mathbf{W}\mathbf{C})(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}\mathbf{e}_{1,2} \\ =\ & \mathbf{e}_{1,2}^T\begin{pmatrix} a(0) & a(1) \\ a(1) & a(2)\end{pmatrix}^{-1}\begin{pmatrix} b(0) & b(1) \\ b(1) & b(2)\end{pmatrix}\begin{pmatrix} a(0) & a(1) \\ a(1) & a(2)\end{pmatrix}^{-1}\mathbf{e}_{1,2} = (\frac{\mu_2^2\nu_0-2\mu_1\mu_2\nu_1+\mu_1^2\nu_2}{\mu_0\mu_2-\mu_1^2})(\frac{b}{h})(1+o(1)), \end{aligned}$$

due to the compactness of $\mathcal{Z}$, the above o-term is uniform in $z \in \mathcal{Z}$, implying

$$\int_{\mathcal{Z}} \sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)^2 dz \;=\; (\frac{\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2}{\mu_0\mu_2 - \mu_1^2})(\frac{b}{h})\,|\mathcal{Z}|\,(1+o(1)) \quad \text{(A.19)}$$

for small h and b, where $|\mathcal{Z}|$ denotes the Lebesgue measure of $\mathcal{Z}$. Hence, (A.19) and the consistency of $\tilde{\beta}(z,s,t)$ in the L_2 sense in (s,t) and uniformly in z due to (A.17) imply

$$\int_{\mathcal{Z}} \left[\sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)^2 \int_{\mathcal{T}} \int_{\mathcal{S}} ((\tilde{\beta}(z^{(p)}, s, t) - \beta(z^{(p)}, s, t)))^2 ds dt \right] dz \xrightarrow{P} 0. \quad \text{(A.20)}$$

For the second part in (A.18), applying Taylor expansion on $\beta(z^{(p)}, s, t)$ at each z,

$$\begin{aligned}
&\sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)\beta(z^{(p)}, s, t) \\
=\;& \mathbf{e}_{1,2}^T \begin{pmatrix} a(0) & a(1) \\ a(1) & a(2) \end{pmatrix}^{-1} \begin{pmatrix} a(0) \\ a(1) \end{pmatrix} \beta(z,s,t) + \mathbf{e}_{1,2}^T \begin{pmatrix} a(0) & a(1) \\ a(1) & a(2) \end{pmatrix}^{-1} \begin{pmatrix} a(1) \\ a(2) \end{pmatrix} \frac{\partial}{\partial z}\beta(z,s,t) \\
&+ \frac{1}{2}\mathbf{e}_{1,2}^T \begin{pmatrix} a(0) & a(1) \\ a(1) & a(2) \end{pmatrix}^{-1} \begin{pmatrix} a(2) \\ a(3) \end{pmatrix} \frac{\partial^2}{\partial z^2}\beta(z,s,t) + \text{higher order terms} \\
=\;& \beta(z,s,t) + \frac{1}{2}b^2 \frac{\mu_2^2 - \mu_1\mu_3}{\mu_0\mu_2 - \mu_1^2} \frac{\partial^2}{\partial z^2}\beta(z,s,t) + \text{higher order terms.}
\end{aligned}$$

Hence $\sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)\beta(z^{(p)}, s, t) - \beta(z,s,t) = \frac{1}{2}b^2 \frac{\mu_2^2 - \mu_1\mu_3}{\mu_0\mu_2 - \mu_1^2} \frac{\partial^2}{\partial z^2}\beta(z,s,t)(1+o(1))$, and

$$\begin{aligned}
&\int_{\mathcal{Z}} \int_{\mathcal{T}} \int_{\mathcal{S}} (\sum_{p=1}^{P} \omega_{0,2}(z^{(p)}, z, b)\beta(z^{(p)}, s, t) - \beta(z,s,t))^2 ds dt dz \\
=\;& \frac{1}{2}b^2 \frac{\mu_2^2 - \mu_1\mu_3}{\mu_0\mu_2 - \mu_1^2} (\int_{\mathcal{Z}} \int_{\mathcal{T}} \int_{\mathcal{S}} \frac{\partial^2}{\partial z^2}\beta(z,s,t) ds dt dz)(1+o(1)) \to 0. \quad \text{(A.21)}
\end{aligned}$$

Combining (A.20) and (A.21) completes the proof by noting further Condition (xi). □

Proof of Theorem 2. Note that

$$\begin{aligned}
Y^*(t) - \hat{Y}^*(t) \;=\;& \mu_{Y,Z^*}(t) - \hat{\mu}_{Y,Z^*}(t) + \int_{\mathcal{S}} (\beta(Z^*, s, t) - \hat{\beta}(Z^*, s, t))(X^*(s) - \mu_{X,Z^*}(s)) ds \\
&- \int_{\mathcal{S}} \hat{\beta}(Z^*, s, t)(\mu_{X,Z^*}(s) - \hat{\mu}_{X,Z^*}(s)) ds.
\end{aligned}$$

The convergence results in Theorem 1 imply that $\int_{\mathcal{T}} (Y^*(t) - \hat{Y}^*(t))^2 dt \xrightarrow{P} 0$ as desired. □