

Research Repository

Functional linear regression for partially observed functional data

Accepted for publication in Wenqing He, Liqun Wang, Jiahua Chen, Chunfang Devon Lin (eds.) 2022. Advances and Innovations in Statistics and Data Science. Springer. Cham.

Research Repository link: <https://repository.essex.ac.uk/34002/>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the published version if you wish to cite this paper.

http://doi.org/10.1007/978-3-031-08329-7_7

www.essex.ac.uk

Functional linear regression for partially observed functional data

Yafei Wang^{a,b}, Tingyu Lai^a, Bei Jiang^b, Linglong Kong^{b†}, Zhongzhan Zhang^{a*}

Abstract In functional linear regression model, many methods have been proposed and studied to estimate the slope function while the functional predictor was observed in the entire domain. However, works on functional linear regression model with partially observed trajectories have received less attention. In this paper, to fill the literature gap we consider the scenario where individual functional predictor may be observed only on part of the domain. Depending on whether measurement error is presented in functional predictors, two methods are developed, one is based on linear functionals of the observed part of the trajectory and the other one uses conditional principal component scores. We establish the asymptotic properties of the two proposed methods. Finite sample simulations are conducted to verify their performance. Diffusion tensor imaging (DTI) data from Alzheimer's Disease Neuroimaging Initiative (ADNI) study is analyzed.

Keywords: Functional linear model; Partially observed functional data; Principal components; Measurement error; ADNI.

1 Introduction

With the advance in technology, it is increasingly common to encounter data that are functions or curves in nature (see Ramsay (2005)). Functional linear regression models provide a framework for modeling the dynamic relationship between response and functional predictors, which was first introduced by Ramsay and Dalzell (1991). One of the primary goals for functional linear model (FLM) is to get an estimator of functional coefficient. And many procedures have been proposed to approximate functional coefficient, for example, functional principal component analysis (FPCA)

a: Beijing University of Technology, Beijing, 100124, China.

b: University of Alberta, Edmonton, T6G 2G1, Canada.

†: Corresponding author: lkong@ualberta.ca

*: Corresponding author: zzhang@bjut.edu.cn

based approaches (Cardot et al. (1999), Hall and Horowitz (2007), Yao et al. (2005b)), spline-based approaches (Crambes et al. (2009), Marx and Eilers (1999)), wavelet-based approaches (Zhao et al. (2012), Wang et al. (2019)), and others. We refer to Morris (2015) and Reiss et al. (2017) for more informative and extensive reviews on such functional linear models.

Among the different based methods in functional data analysis, FPCA based approaches for capturing the information of covariates are popular (Hall et al. (2006), Che et al. (2017)). In the setting where trajectories are observed on dense and regular grid on the entire domain, the existing works can be found in Besse and Ramsay (1986), Rice and Silverman (1991), Cardot et al. (1999), Shin (2009), Horváth and Kokoszka (2012), to name a few. Yao et al. (2005a) emphasizes the case where the functional predictors are observed with irregularly sparse measurements which is often referred to as sparse functional data, and proposes a nonparametric method to perform FPCA. For general review on FPCA, see Shang (2014). In this paper, we prefer to use FPCA method to get an estimator of the functional coefficient.

Sparse functional data addresses the case where each trajectory is observed at a small number of points that are distributed randomly on the domain which is different from the partially observed functional data (or incomplete or fragmentary functional data) which was first introduced in Liebl (2013). Partially observed functional data addresses each trajectory is observed at points that cover a subset of the domain in such a way that trajectories can be reasonably treated as fragments of curves (Delaigle and Hall (2016)) that has great implication in applications, such as in biomedicine, economics (see Kraus (2015), Kneip and Liebl (2020)). Considering the partially observed functional data can be treated as missing data for functional curves over the domain, two missing mechanisms are introduced in the existing works: one is missing completely at random (MCAR), that is, the missing data mechanism is independent from other stochastic components (Delaigle and Hall (2016), Goldberg et al. (2014)); the other one is the missing mechanism in which depends on systematic strategies, such as missing parts of the trajectories only occur at the upper interval of the domain (see Liebl and Rameseder (2019)). In the setting of MCAR, Delaigle and Hall (2016), Goldberg et al. (2014) and Kraus (2015) address the problem for recovering the missing parts of trajectories. Kraus (2015) and Kneip and Liebl (2020) model the functional principal (FPC) scores of an incomplete trajectory. In the scenario where missing data mechanism depends on systematic strategies, Liebl and Rameseder (2019) establishes estimators for the mean and the covariance function of the incomplete functional data via the fundamental theorem of calculus. To the best of our knowledge, no work exists focusing on estimating functional coefficient of FLM with partially observed trajectories.

In this paper, we address the problem of getting an estimator of functional coefficient for the case of partially observed functional data without and with measurement error. In the scenario that trajectories observed without measurement error, instead of deleting the incomplete trajectories, we get estimators of FPC scores for each incomplete trajectory by modeling it as linear functionals of the observed parts of that trajectory. In the setting where trajectories observed with measurement error,

we use local linear smoother methods to estimate mean and covariance function of the functional predictor, followed by getting FPC scores via conditional expectation.

The contributions of this paper are as follows. First, we extend FLM approach to partially observed functional data without measurement error, which leads to an improved estimator for functional coefficient comparing with the one obtained through deleting the incomplete trajectories for given dataset. Second, we develop an estimate method for functional coefficient in FLM for incomplete trajectories with measurement error. We illustrate its usefulness by comparing with another two methods: one is based on integration method to get the FPC scores of the functional predictor instead of using conditional expectations; the other estimator is obtained by ignoring the measurement error of the trajectories in the dataset. Third, in both scenarios, we obtain the rate of convergence for the proposed estimators. Overall, the methodological and numerical developments in this paper can provide a practically useful way in analyzing FLM with partially observed functional data.

The rest of this paper is organized as follows. In Section 2, we introduce functional linear models. In Section 3.1, we develop an estimator for functional coefficient with incomplete trajectories observed without measurement error, and establish theoretical properties for the proposed estimator. An estimator and theoretical properties in the scenario that incomplete trajectories observed with measurement error is introduced in Section 3.2. Section 4 illustrates the finite sample performance of our proposed estimators through simulation studies, followed by a real data analysis in Section 5. Discussion is presented in Section 6. Proofs of theorems are given in the Appendix.

2 Functional Linear Model

Consider a functional linear model, in which the scalar response Y_i is linearly related to the functional covariate X_i ,

$$Y_i = \alpha + \int_{\mathcal{T}} \gamma(t) X_i(t) dt + \epsilon_i, \quad (1)$$

where α is the intercept, $\{X_i(t) : t \in \mathcal{T}, i = 1, \dots, n\}$ are the functional predictors, sampled from the stochastic process $\{X(t) : t \in \mathcal{T}\}$ with mean function μ , domain \mathcal{T} is bounded and closed, γ is the slope function to be estimated, ϵ_i are random errors satisfying $E[\epsilon_i] = 0$, $E[\epsilon_i^2] = \sigma^2 < \infty$. We can easily get an estimator of intercept once we get an estimator of γ . So we focus on estimating γ in the following (Hall and Horowitz (2007)). Let $\langle \cdot, \cdot \rangle$, $\|\cdot\|$ be the inner product and norm on $L^2(\mathcal{T})$, the set of all square integrable functions on \mathcal{T} , with $\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t)dt$, $\|f\| = \langle f, f \rangle^{1/2}$ for any $f, g \in L^2(\mathcal{T})$.

We first recall the method FPCA in estimating the slope function for model (1) with the functional predictor X_i observed on the entire domain \mathcal{T} . For the stochastic process $X \in L^2(\mathcal{T})$, denote its mean function as μ : $\mu = E(X)$, and its covariance

function as $c_X(s, t)$: $c_X(s, t) = \text{cov}(X(s), X(t))$. Assume c_X is continuous on $\mathcal{T} \times \mathcal{T}$. The expression $c_X(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t)$ exists by the Mercer Lemma (Riesz and Nagy (1955)), where $\lambda_1 > \lambda_2 > \dots > 0$; ϕ_1, ϕ_2, \dots are the eigenvalue sequence and the continuous orthonormal eigenfunction sequence of the linear operator C_X : $(C_X \phi)(\cdot) = \int_{\mathcal{T}} c_X(\cdot, t) \phi(t) dt$, $\phi \in L^2(\mathcal{T})$, with the kernel c_X . On the other hand, by the Karhunen-Loëve (K-L) expansion, one has $X_i(t) = \sum_{j=1}^{\infty} U_{ij} \phi_j(t)$, where the random variables $U_{ij} = \langle X_i - \mu, \phi_j \rangle$ are uncorrelated with $E[U_{ij}] = 0$, $E[U_{ij}^2] = \lambda_j$, and $\gamma(t) = \sum_{j=1}^{\infty} \gamma_j \phi_j(t)$ with $\gamma_j = \langle \gamma, \phi_j \rangle$.

The full model (1) is then equivalent to $Y_i - EY_i = \sum_{j=1}^{\infty} \gamma_j U_{ij} + \epsilon_i$ based on K-L expansion, which can be approximated by $\sum_{j=1}^m \gamma_j U_{ij} + \epsilon_i$ by using the first m terms. To simplify notations, we assume that $\{Y_i, i = 1, \dots, n\}$ are centered. Let $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\gamma = (\gamma_1, \dots, \gamma_m)^T$, $\hat{\mu}$ be an estimator of μ , $\{\hat{\lambda}_j\}$ and $\{\hat{\phi}_j\}$ be estimators of the sequence $\{\lambda_j\}$ and $\{\phi_j\}$ with $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > 0$. The least square estimator $\hat{\gamma}$ is then given as

$$\hat{\gamma} = (\hat{\mathbf{U}}_m^T \hat{\mathbf{U}}_m)^{-1} \hat{\mathbf{U}}_m \mathbf{Y}, \quad (2)$$

provided that $(\hat{\mathbf{U}}_m^T \hat{\mathbf{U}}_m)^{-1}$ exists with $\hat{U}_{ij} = \langle X_i - \hat{\mu}, \hat{\phi}_j \rangle$, $\hat{\mathbf{U}}_m = (\hat{U}_{ij})_{i=1, \dots, n; j=1, \dots, m}$. Moreover, for the estimator $\hat{\gamma}_j$, $j = 1, \dots, m$, it has the equivalent form as

$$\hat{\gamma}_j = \hat{\lambda}_j^{-1} \left\langle n^{-1} \sum_{i=1}^n (Y_i - \bar{Y}_0) (X_i - \hat{\mu}), \hat{\phi}_j \right\rangle.$$

Consequently, an estimator of γ is given by

$$\hat{\gamma}(t) = \sum_{j=1}^m \hat{\gamma}_j \hat{\phi}_j(t). \quad (3)$$

The number m of included eigenfunctions is chosen by fraction of variance explained criterion in practice (James et al. (2000)): $m = \min\{k : \sum_{l=1}^k \hat{\lambda}_l / \sum_{l=1}^n \hat{\lambda}_l \geq R\}$, with a given threshold R . For the asymptotic analysis, we assume m depends on sample size n such that $m \rightarrow \infty$ as $n \rightarrow \infty$.

3 Estimation Methods

The above analysis is based on the assumption the functional predictor is observed on the entire domain. We now consider the scenario that the predictor $X_i, i = 1, \dots, n$ may be available only on parts of \mathcal{T} . We first give some notations and then make further analysis. Let X_1, \dots, X_n be an independent and identically distributed samples from the random function X . We denote the observed and missing parts of X_i by O_i and M_i with $O_i \cup M_i = \mathcal{T}$. Let $O_i = [L_i, R_i] \subseteq \mathcal{T}$, and assume that it is a random subinterval independent of X_i with $R_i - L_i > 0$ almost surely. The observed data for

i th functional predictor is then given as $X_i(t), t \in O_i, i = 1, \dots, n$, denoted by X_{iO_i} . In this section, our objective interest is to develop an estimation method for model (1) with partially observed functional observations without and with measurement error respectively. And in these scenarios, our objective is to get estimators of the functional principal component scores $\{U_{ij}\}$ and the eigenfunctions $\{\phi_j\}$ as indicated in formulas (2), (3). Depending on whether measurement error is presented in partially observed functional curves, two methods are developed: one is established by applying linear functionals of the observed parts of that trajectory, while the other one is based on principal component analysis through conditional expectation.

3.1 Partially Observed Functional Data without Measurement Error

In the scenario that functional curves are partially observed on the domain without measurement error, to get an estimator of γ in model (1), we need to get estimators of U_{ij} and ϕ_j pertaining to this case. An estimator of U_{ij} is obtained based on the linear functional of the observed part X_{iO_i} , and an estimator of ϕ_j is obtained by giving estimators of mean and covariance function of X . The steps are given here.

Step 1: Estimate the mean μ and the covariance function c_X by sample mean and sample covariance.

Step 2: Estimate eigenvalues $\{\lambda_j\}$ and eigenfunctions $\{\phi_j\}$ by $\int_T \hat{c}_X(s, t) \hat{\phi}_j(s) ds = \hat{\lambda}_j \hat{\phi}_j(t)$.

Step 3: Estimate principal component scores $U_{ij} = U_{ijO_i} + U_{ijM_i}$ with $\hat{U}_{ijO_i} = \langle X_{iO_i} - \hat{\mu}_{O_i}, \hat{\phi}_{jO_i} \rangle$, and estimate U_{ijM_i} by modeling it as linear functionals of X_{iO_i} given as $\hat{U}_{ijM_i} = \langle \hat{\xi}_{ijM_i}, X_{iO_i} - \hat{\mu}_{O_i} \rangle$.

Step 4: Estimate γ based on formulas (2) and (3) for X_{iO_i} observed without measurement error.

We first address the problem of getting estimators of μ and c_X , denoted as $\hat{\mu}^{\text{NME}}$ and \hat{c}_X^{NME} respectively, followed by establishing estimators of U_{ij} and eigenfunctions ϕ_j which are denoted as $\hat{U}_{ij}^{\text{NME}}$ and $\hat{\phi}_j^{\text{NME}}$. For simplicity of presentation, we suppress the notation on "NME" in this subsection unless otherwise stated.

Let $O_i(t) = I_{O_i}(t)$ with indicator function $I_{O_i}(t)$ being 1 if $t \in O_i$, and 0 otherwise, and let $W_i(s, t) = O_i(s)O_i(t)$. The estimators of the mean function μ and the covariance function c_X of X obtained from the observed points s, t of X_i , are given by,

$$\hat{\mu}(t) = \frac{1}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) X_i(t), \quad (4)$$

$$\hat{c}_X(s, t) = \frac{1}{\sum_{i=1}^n W_i(s, t)} \sum_{i=1}^n W_i(s, t) (X_i(s) - \hat{\mu}(s)) (X_i(t) - \hat{\mu}(t)). \quad (5)$$

Therefore, we get the estimators $\{\hat{\lambda}_j\}$, $\{\hat{\phi}_j\}$ related to $\{\lambda_j\}$ and $\{\phi_j\}$ from \hat{c}_X associated with the covariance operator \hat{C}_X .

We could not get estimators \hat{U}_{ij} of FPC scores $\{U_{ij}\}$ of X_i directly from its definition if $O_i \neq \mathcal{T}$. To bridge the gap, U_{ij} is decomposed into two parts:

$$U_{ij} = \langle X_{iO_i} - \mu_{O_i}, \phi_{jO_i} \rangle + \langle X_{iM_i} - \mu_{M_i}, \phi_{jM_i} \rangle = U_{ijO_i} + U_{ijM_i}, \quad (6)$$

where μ_{O_i} and ϕ_{jO_i} denote the restriction of μ and the eigenfunction ϕ_j on O_i respectively, and the definitions of μ_{M_i} , ϕ_{jM_i} are similar. The estimator \hat{U}_{ijO_i} of U_{ijO_i} can be estimated directly from the observed part X_{iO_i} and the estimator $\hat{\phi}_j$, given as $\hat{U}_{ijO_i} = \langle X_{iO_i} - \hat{\mu}_{iO_i}, \hat{\phi}_{jO_i} \rangle$. For the term U_{ijM_i} , we consider using the linear functional form $\langle \xi_{ijM_i}, X_{iO_i} - \mu_{O_i} \rangle$ of the observed part X_{iO_i} to estimate it which is also considered in Kraus (2015), that is,

$$\hat{\xi}_{ijM_i} = \underset{\xi_{ijM_i} \in L^2}{\operatorname{argmin}} n^{-1} \sum_{i=1}^n (\hat{U}_{ijM_i} - \langle \xi_{ijM_i}, X_{iO_i} - \hat{\mu}_{iO_i} \rangle)^2$$

with $\hat{U}_{ijM_i} = \langle X_{iM_i} - \hat{\mu}_{M_i}, \hat{\phi}_{jM_i} \rangle$. The estimator $\hat{\xi}_{ijM_i}$ has the explicit form: $\hat{\xi}_{ijM_i} = \hat{C}_{O_iO_i}^{-1} \hat{C}_{O_iM_i} \hat{\phi}_{jM_i}$, where $\hat{C}_{O_iO_i}$, $\hat{C}_{O_iM_i}$ are the empirical covariance operator for $C_{O_iO_i}$, $C_{O_iM_i}$ with the kernel being the covariance function \hat{c}_X of X_i restricted to $O_i \times O_i$ and $O_i \times M_i$ respectively. To obtain a stable solution, we adopt ridge regularization, given by

$$\begin{aligned} \hat{\xi}_{ijM_i}^{(\rho)} &= (\hat{C}_{O_iO_i}^{(\rho)})^{-1} \hat{C}_{O_iM_i} \hat{\phi}_{jM_i}, \\ \hat{U}_{ijM_i}^{(\rho)} &= \langle \hat{\xi}_{ijM_i}^{(\rho)}, X_{iO_i} - \hat{\mu}_{iO_i} \rangle, i = 1, \dots, n, j = 1, \dots, m, \end{aligned} \quad (7)$$

where $\hat{C}_{O_iO_i}^{(\rho)} = \hat{C}_{O_iO_i} + \rho \mathcal{F}_{O_i}$, \mathcal{F}_{O_i} is an identity operator defined on $L^2(O_i)$, ρ is a ridge parameter; see Kraus (2015) for further details. Let $\hat{U}_{ij}^{\text{NME}} = \hat{U}_{ijO_i} + \hat{U}_{ijM_i}^{(\rho)}$. The estimator $\hat{\gamma}^{\text{NME}}$ of γ using all of the information of the dataset is then obtained through replacing \hat{U}_{ij} in (2) with $\hat{U}_{ij}^{\text{NME}}$,

$$\hat{\gamma}^{\text{NME}}(t) = \sum_{j=1}^m \hat{\gamma}_j \hat{\phi}_j(t). \quad (8)$$

To facilitate our theoretical analysis, we first impose some assumptions on observation points for partially observed functional curves, indicating the observation points asymptotically provide enough information in individual or pairwise crossover.

(A1) There exists $\delta_1 > 0$ s.t. $\sup_{t \in [0,1]} P\{n^{-1} \sum_{i=1}^n I_{O_i}(t) \leq \delta_1\} = O(n^{-2})$.

(A2) There exists $\delta_2 > 0$ s.t. $\sup_{s,t \in [0,1]^2} P\{n^{-1} \sum_{i=1}^n W_i(s,t) \leq \delta_2\} = O(n^{-2})$.

Moreover, we also introduce some regularity conditions necessary to derive theoretical properties for the estimate $\hat{\gamma}^{\text{NME}}$.

- (A3) $E\|X - \mu\|^4 < \infty$.
- (A4) $nm^{-1} \rightarrow \infty$, $n/(\sum_{j=1}^m \delta_j^{-2}) \rightarrow \infty$ with $\delta_j = \min_{j \geq 1} \{\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j\}$ and $n\lambda_m^2 \rightarrow \infty$ as $m \rightarrow \infty$.
- (A5) The ridge parameter ρ satisfies $\rho \rightarrow 0$, $n\rho^3 \rightarrow 0$, $nm^{-1}\rho^2 \rightarrow \infty$.
- (A6) $\sum_{k=1}^{\infty} [E[YU_k]]^2/\lambda_k^2 < \infty$.
- (A7) $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{r_{M_i O_i j k}^2}{\lambda_{O_i O_i k}^2} < \infty$, with $r_{M_i O_i j k} = \text{cov}(\langle X_{M_i} - \mu_{M_i}, \phi_{M_i M_i j} \rangle, \langle X_{M_i} - \mu_{M_i}, \phi_{O_i O_i k} \rangle)$.

Assumption (A3) is a common condition in the analysis of functional model by using the method of FPCA to guarantee the random functions have finite fourth moment (see Cardot et al. (1999)). Note that if the eigenvalues $\{\lambda_j\}$ are exponentially or geometrically decreasing, the assumption (A4) holds. The same kind of conditions are also introduced in Cardot et al. (1999). Assumption (A5) is used to control the size of ridge effect. To define the convergence of the right hand of the formula $\gamma(s) = \sum_{k=1}^{\infty} (E[YU_k]/\lambda_k)\phi_k(s)$, in the L^2 sense, assumption (A6) is required that is similar to the condition (A1) in Yao et al. (2005b). Assumption (A7) is used to make the solution $\hat{\xi}_{ijM_i}$ valid which is commonplace in the theory of inverse problems as Picard condition (see Hansen (1990)).

Let $\theta_n = \sum_{k=m}^{\infty} [E[YU_k]]^2/\lambda_k^2$. Then assumption (A6) indicates that $\theta_n \rightarrow 0$. Denote $\nu = \sum_{j=1}^m V_{ij}$ with $V_{ij} = \langle \phi_{jM_i}, (C_{M_i M_i} - C_{M_i O_i} C_{O_i O_i}^{-1} C_{O_i M_i}) \phi_{jM_i} \rangle$. Based on the above assumptions, Theorem 1 gives the converge rate for the estimator $\hat{\gamma}^{\text{NME}}$ in the L^2 sense.

Theorem 1 Suppose that (A1)-(A7) are satisfied. Then

$$\|\hat{\gamma}^{\text{NME}} - \gamma\|^2 = O_p(n^{-1}m\rho^{-2} + \iota_n + \theta_n + \nu).$$

with $\iota_n = n^{-1} \sum_{j=1}^m \delta_j^{-2}$.

Theorem 1 indicates that the approximation error rate of $\hat{\gamma}^{\text{NME}}$ for γ is controlled by four terms. The first term depends on sample size n , tuning parameter m , ridge parameter ρ , which is of the higher order than the one given in Hall and Horowitz (2007) that is mainly due to functional curves observed on the part of the domain. The second term is related to the spacings between adjacent eigenvalues, and its effect on convergence rate of γ is also emphasized in Hall and Horowitz (2007). The third term is related to the convergence of γ in L^2 sense, which is also show in Yao et al. (2005b) to get approximation error rate for functional coefficient. The fourth term is introduced by approximating U_{ijM_i} with \tilde{U}_{ijM_i} .

Note that in practice, the ridge parameter ρ included in the regularized estimation of the j th score of the i th functional observation is chosen by generalized cross-validation based on the set of samples observed on the entire domain (see Kraus (2015)).

3.2 Partially Observed Functional Data with Measurement Error

In this subsection, we construct an estimator for the slope function γ for partially observed trajectories with measurement error. We suppose the functional observations are:

$$Z_{il} = X_i(t_{il}) + \varepsilon_{il}, \quad t_{il} \in O_i, i = 1, \dots, n, l = 1, \dots, N_i, \quad (9)$$

where ε_{il} is independent from all the other variables $X_j, j \neq i$, with $E(\varepsilon_{il}) = 0$, $\text{var}(\varepsilon_{il}) = \sigma_X^2$.

To get an estimator of γ in (1) in the scenario that trajectories may be observed on parts of the domain with measurement error (WME), we need give estimators of FPC scores and eigenstructure pertaining to this case. Estimator of eigenstructure is established after using local linear smoothers to get estimators of mean and covariance function of X . We obtain estimators of FPC scores by using approach of principal component analysis via conditional expectation. The steps are given here.

Step 1: Estimate the mean and covariance functions by local linear smoothers.

Step 2: Estimate eigenvalues $\{\lambda_j\}$ and eigenfunctions $\{\phi_j\}$ by

$$\int_{\mathcal{T}} \hat{c}_X^{\text{WME}}(s, t) \hat{\phi}_j^{\text{WME}}(s) ds = \hat{\lambda}_j^{\text{WME}} \hat{\phi}_j^{\text{WME}}(t).$$

Step 3: Estimate FPC scores $\{U_{ij}\}$ by principal component analysis via conditional expectation (PACE): $\tilde{U}_{ij} = E[U_{ij} | \mathbf{Z}_i]$.

Step 4: Based on obtained estimators \hat{U}_{ij} and $\hat{\phi}_j^{\text{WME}}$, we get estimator γ^{WME} for X_{iO_i} observed with measurement error.

We first calculate estimators for the mean and the covariance function of X in the scenario (9), denoted as $\hat{\mu}^{\text{WME}}$ and \hat{c}_X^{WME} , that are required to derive estimators for the FPC scores $U_{ij} = \int (X_i(t) - \mu(t)) \phi_j(t) dt$. For simplicity of presentation, we suppress notation on “WME” unless otherwise stated in this subsection.

Let $K(\cdot)$ be a nonnegative univariate kernel function that is assumed to be a symmetric probability density function (pdf) with compact support $\text{supp}(K) = [-1, 1]$, and h_μ, h_c be the bandwidths for obtaining estimators of μ, c_X . Assume that the second derivatives of μ, c_X on $\mathcal{T}, \mathcal{T}^2$ respectively exist. We use local linear smoothers for the mean function μ (Yao et al. (2005a), Yao et al. (2005b), Kneip and Liebl (2020)) defined as $\hat{\mu}(t) = \hat{\beta}_0$, where

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\text{argmin}} \sum_{i=1}^n \sum_{l=1}^{N_i} K\left(\frac{t_{il} - t}{h_\mu}\right) [Z_{il} - \beta_0 - \beta_1(t - t_{il})]^2. \quad (10)$$

Let $\hat{G}_{ilk} = (Z_{il} - \hat{\mu}(t_{il}))(Z_{ik} - \hat{\mu}(t_{ik}))$ be the raw covariance points. The local linear smoother for the covariance function c_X is defined as $\hat{c}_X = \hat{\beta}_0$, where

$$\begin{aligned} (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \underset{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2}{\arg \min} & \sum_{i=1}^n \sum_{1 \leq l, k \leq N_i} K\left(\frac{t_{il} - t}{h_c}\right) K\left(\frac{t_{ik} - s}{h_c}\right) \\ & \times [\hat{G}_{ilk} - \tilde{\beta}_0 - \tilde{\beta}_1(t_{il} - t) - \tilde{\beta}_2(t_{ik} - s)]^2. \end{aligned} \quad (11)$$

Similar to the technique introduced in Yao et al. (2005a), the points \hat{G}_{ill} , $l = 1 \cdots, N_i$ are not included in (11). Let $\mathcal{T}_1 = [\inf\{L_i \in \mathcal{T}, i = 1, \cdots, n\} + |\mathcal{T}|/4, \sup\{R_i \in \mathcal{T}, i = 1, \cdots, n\} - |\mathcal{T}|/4]$ with $|\mathcal{T}|$ being the length of \mathcal{T} . The estimator of σ_X^2 is defined as $\hat{\sigma}_X^2$ if $\hat{\sigma}_X^2 > 0$, otherwise $\hat{\sigma}_X^2 = 0$ with

$$\hat{\sigma}_X^2 = 2 \int_{\mathcal{T}_1} (\hat{V}_X(t) - \tilde{G}(t)) dt / |\mathcal{T}|$$

where $\hat{V}_X(t)$ is the local linear estimator using the points $\{\hat{G}_{ill}\}$, $\tilde{G}(t)$ is the estimate $\hat{c}_X(s, t)$ restricted to $s = t$ (Staniswalis and Lee (1998), Yao et al. (2005a)). The estimators of $\{\lambda_j, \phi_j\}_{j \geq 1}$ are the corresponding solutions of the eigen-equations

$$\int_{\mathcal{T}} \hat{c}_X(s, t) \hat{\phi}_j(s) ds = \hat{\lambda}_j \hat{\phi}_j(t).$$

Based on the K-L expansion of X_i , model (9) can be rewritten as

$$Z_{il} = \mu(t_{il}) + \sum_{j=1}^{\infty} U_{ij} \phi_j(t_{il}) + \varepsilon_{il}, \quad t_{il} \in O_i, i = 1 \cdots, n, l = 1 \cdots, N_i.$$

Let $\mathbf{X}_i = (X_i(t_{i1}), \cdots, X_i(t_{iN_i}))^T$, $\mathbf{Z}_i = (Z_{i1}, \cdots, Z_{iN_i})^T$, $\mu_i = (\mu(t_{i1}), \cdots, \mu(t_{iN_i}))^T$, $\phi_{ij} = (\phi_j(t_{i1}), \cdots, \phi_j(t_{iN_i}))^T$. Assume that U_{ij} and ε_{il} are jointly Gaussian. Following Yao et al. (2005a), the best prediction of U_{ij} of the i th subject given the observations (Z_{il}, t_{il}) , $l = 1, \cdots, N_i$ is obtained as

$$\tilde{U}_{ij} = \lambda_j \phi_{ij}^T \Sigma_{\mathbf{Z}_i}^{-1} (\mathbf{Z}_i - \mu_i),$$

where $\Sigma_{\mathbf{Z}_i} = \text{cov}(\mathbf{Z}_i, \mathbf{Z}_i) = \text{cov}(\mathbf{X}_i, \mathbf{X}_i) + \sigma_X^2 \mathbf{I}_{N_i}$ with identity matrix \mathbf{I}_{N_i} . That is, the (u, v) th element of $\Sigma_{\mathbf{Z}_i}$ is $(\Sigma_{\mathbf{Z}_i})_{u,v} = c_X(t_{iu}, t_{iv}) + \sigma_X^2 I_{uv}$ with $I_{uv} = 1$ if $u = v$, and 0 otherwise. Then the estimator of U_{ij} is given through substituting μ, λ_j, ϕ_j with $\hat{\mu}, \hat{\lambda}_j, \hat{\phi}_j$ as

$$\hat{U}_{ij}^{\text{WME}} = \hat{\lambda}_j \hat{\phi}_{ij}^T \hat{\Sigma}_{\mathbf{Z}_i}^{-1} (\mathbf{Z}_i - \hat{\mu}_i), \quad (12)$$

where the (u, v) th entry of $\hat{\Sigma}_{\mathbf{Z}_i}$ is $(\hat{\Sigma}_{\mathbf{Z}_i})_{u,v} = \hat{c}_X(t_{iu}, t_{iv}) + \hat{\sigma}_X^2 I_{uv}$. Replacing \hat{U}_{ij} in (2) with $\hat{U}_{ij}^{\text{WME}}$, we then get the estimator $\hat{\gamma}^{\text{WME}}$ of γ from (3)

$$\hat{\gamma}^{\text{WME}}(t) = \sum_{j=1}^m \hat{\gamma}_j \hat{\phi}_j,$$

where $\hat{\gamma}_j$ is the j th entry of $\hat{\gamma}$ with $\hat{U}_{ij}^{\text{WME}}$ in (2).

Next, we give some theoretical results for $\hat{\gamma}^{\text{WME}}(t)$. We assume the following regularity conditions which are similar to the assumptions in Kneip and Liebl (2020), Yao et al. (2005b).

(B1) The observational points $\{t_{il}, l = 1, \dots, N_i\}$ given O_i for the i th subject, are i.i.d. random variables with pdf $f_{t|O_i}(u) > 0$ for all $u \in O_i \subseteq \mathcal{T}$ and zero else. For the marginal pdf f_t of observation times t_{ij} , $f_t(u) > 0$ for all $u \in \mathcal{T}$.

(B2) Let $N = \min\{N_i, i = 1, \dots, n\}$. $N \asymp n^r$ with $0 < r < \infty$, where $a_n \asymp b_n$ means that there exists a constant $0 < L < \infty$ such that $a_n/b_n \rightarrow L$ as $n \rightarrow \infty$.

(B3) $h_\mu \rightarrow 0$, $h_c \rightarrow 0$, $nN h_\mu \rightarrow \infty$, $nM h_c \rightarrow \infty$ as $n \rightarrow \infty$ with $M = N^2 - N$.

(B4) K is a second order kernel with compact support $[-1, 1]$.

(B5) Let $G_{ilk} = (Z_{il} - \mu(t_{il}))(Z_{ik} - \mu(t_{ik}))$. Define f_{Zt} , f_{tt} , f_{Gtt} as the joint pdf of (Z_{il}, t_{il}) on $\mathbb{R} \times \mathcal{T}$, (t_{il}, t_{il}) on \mathcal{T}^2 , $(G_{ilk}, t_{il}, t_{ik})$ on $\mathbb{R} \times \mathcal{T}^2$, respectively. All of the second derivatives of f_{Zt} , f_{tt} , f_{Gtt} are uniformly continuous and bounded. Moreover, f_t is uniformly continuous and bounded on \mathcal{T} .

(B6) Let $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_m\}$, $\Xi = (\lambda_1 \phi_{i1}, \dots, \lambda_m \phi_{im})^T$, $\Upsilon = \Lambda - \Xi \Sigma_{Z_i}^{-1} \Xi^T$ and $\varsigma_n \equiv \text{trace}(\Upsilon)$. Denote $r_\mu = h_\mu^2 + 1/\sqrt{nN h_\mu} + 1/\sqrt{n}$, $r_c = h_c^2 + 1/\sqrt{nM h_c^2} + 1/\sqrt{n}$. $\nu_n \equiv m r_\mu \rightarrow 0$, $\tau_n \equiv r_c (\sum_{j=1}^m \delta_j^{-1}) \rightarrow 0$.

Theorem 2 Under the regularity conditions (A3), (A6), (B1)-(B6), we have that

$$\|\hat{\gamma}^{\text{WME}} - \gamma\|^2 = O_p(\nu_n + \tau_n + \varsigma_n + \theta_n).$$

Theorem 2 gives the rate of convergence of the estimator $\hat{\gamma}^{\text{WME}}$ in the L^2 sense. The rate of convergence of $\hat{\gamma}^{\text{WME}}$ depends on the sample size and bandwidths which is common for estimating curves or surface by local linear smoothers for functional data analysis (see Li and Hsing (2010)). Related results of Theorem 2 can also be found in Yao et al. (2005b). The terms ν_n , τ_n are related to rates of convergence of estimators for the mean and covariance function by using local linear smoothers. The term ς_n are introduced by approximating U_{ij} with \hat{U}_{ij} .

4 Simulation Studies

In this section, we use the simulated datasets to evaluate the finite sample properties of our proposed methods in Section 3. This studies are based on $n \in \{50, 100, 200\}$ i.i.d. samples $\{X_i, Y_i\}_{i=1}^n$ and equally spaced grid $\{t_1, \dots, t_{30}\}$ on $[0, 1]$ with $t_1 = 0$, $t_{30} = 1$. For the i th functional observation $X_i(t)$, the missing interval M_i takes the form $[R_i - E_i, R_i + E_i]$, with $R_i = a_1 T_{i1}^{1/2}$, $E_i = a_2 T_{i2}$, where T_{i1} , T_{i2} are independent random variables uniformly distributed on $[0, 1]$, $a_1, a_2 \in \mathbb{R}$. We consider $(a_1, a_2) = (1.5, 0.2)$, $(a_1, a_2) = (1.5, 0.4)$ with the expected missing length over the domain being 0.4 and 0.8, respectively. We set the intercept $\alpha = 0$. To evaluate the performance of an estimator $\hat{\gamma}$ of γ , mean integrated square error (MISE) is used below as an evaluation criterion, given by,

$$\text{MISE} = \frac{1}{N} \sum_{l=1}^N \int_0^1 (\hat{\gamma}_l(t) - \gamma(t))^2 dt,$$

where N is the number of Monte Carlo replications.

For functional predictors $\{X_i\}$ without measurement error, the trajectories are generated as follows. The simulated random function X_i has zero mean, the covariance function is generated from two eigenfunctions, $\phi_1(t) = \sqrt{2}\sin(\pi t/2)$, $\phi_2(t) = \sqrt{2}\sin(3\pi t/2)$. For the eigenvalues, we take $\lambda_1 = (\pi/2)^{-2}$, $\lambda_2 = (3\pi/2)^{-2}$, $\lambda_k = 0$, for $k \geq 3$. The error ϵ_i in (1) is assumed to be standard normal. For the slope function γ in (1), we take the form $\gamma(t) = \phi_1(t) + 3\phi_2(t)$. We compare the finite sample performance of our proposed method with the method that gives an estimator for γ through formula (2), (3) with deleting the incomplete functional observations in the datasets denoted as “SUB”. Moreover, the estimator of γ based on the original complete dataset is also considered in this scenario, and denote it as “ORI”. We conduct 1000 simulation runs in each setup. Table 1 reports the results.

Table 1 MISEs of the estimators of γ under different methods with 1000 Monte Carlo replications for functional predictors without measurement error

method	(a_1, a_2)	$n = 50$	$n = 100$	$n = 200$
ORI ^a		2.0295	1.0767	0.3670
NME ^b	(1.5, 0.2)	2.8653	1.6650	0.7343
	(1.5, 0.4)	3.5650	2.4412	1.3497
SUB ^c	(1.5, 0.2)	3.5632	1.8844	0.8322
	(1.5, 0.4)	4.600	2.6664	1.4401

^a The estimator is obtained with the original dataset $\{X_i, Y_i\}$ with functional predictors observed in entire domain $[0, 1]$ (ORI)

^b The estimator $\hat{\gamma}^{\text{NME}}$ introduced in Section 3.1 (NME)

^c The estimator is obtained by deleting the functional predictors with missing parts (SUB)

As shown in Table 1, in the scenario where incomplete functional predictors are observed without measurement error, the estimation method in Section 3.1 performs better than “SUB” method. This is because some useful information the dataset has will be lost if we delete them directly, while the “NME” method can take advantage of the whole information about the dataset. Specially, in each setting for (a_1, a_2) , MISEs from the “NME” method have smaller values relative to the “SUB” method. These simulation results also demonstrate that MISEs decrease with increasing sample size n for these three methods. And MISEs increase with longer missing length on $[0, 1]$ at fixed n indicating that a large error is introduced for the “NME” method in imputing missing scores of incomplete functional predictors through little available information from functional samples. In further, the difference of MISEs among these three methods are reduced with increasing sample size n , and the “NME” method still performs better than the “SUB” method, those imply the “NME” method is promising.

For functional predictors X_i with measurement error, they are generated according to $Z_i(t_{il}) = X_i(t_{il}) + \epsilon_{il}$, $l = 1, \dots, 30$, as follows. We take $X_i(t) = \sum_{j=1}^{50} U_{ij} \phi_j(t)$ with $U_{ij} = (-1)^{j+1} j^{-1.1/2} W_{ij}$, where W_{ij} is uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$, $\phi_1(t) = 1$, $\phi_j(t) = \sqrt{2}\cos(j\pi t)$ for $j \geq 2$. The additional random error ϵ_{il} , $l = 1 \dots, 30$ and the error ϵ_i in (1) are assumed to be normal with mean zero, variance 0.25. For

the slope function γ , we take $\gamma = \sum_{j=1}^{50} \gamma_j \phi_j(t)$ with $\gamma_1 = 0.3$, $\gamma_j = 4(-1)^{j+1} j^{-2}$ for $j \geq 2$ (Hall and Horowitz (2007)). We conduct 100 simulation runs in each setup. To demonstrate the superior performance of our proposed method in Section 3.2, we compare it with the other two methods after we get estimators of $\mu(t)$ and $c_X(s, t)$ by solving the optimization problems (10), (11) respectively: one is that an estimator of γ is established by applying integration method to get the FPC scores \hat{U}_{ij} in (2) instead of using formula (12), denoted as “IN”; the other one is that an estimator of γ is obtained by using the method in Section 3.1 with dataset $\{Z_i, Y_i\}$ with measurement error being ignored. The results are summarized in Table 2.

Table 2 MISEs of the estimators of γ under different methods with 100 Monte Carlo replications for functional predictors with measurement error

method	(a_1, a_2)	$n = 50$	$n = 100$	$n = 200$
WME ^a	(1.5, 0.2)	0.1535	0.1176	0.0753
	(1.5, 0.4)	0.2033	0.1607	0.1057
IN ^b	(1.5, 0.2)	0.1702	0.1560	0.1024
	(1.5, 0.4)	0.2671	0.2374	0.1974
NME ^c	(1.5, 0.2)	0.6312	0.4517	0.3320
	(1.5, 0.4)	0.7249	0.5086	0.3808

^a The estimator is obtained by using the method in Section 3.2 (WME)

^b The estimator is obtained by using integration method to get estimators of the principal component scores U_{ij} (IN)

^c The estimator is obtained by using the method in Section 3.1 (NME)

We find from Table 2 that the “WME” method has the best performance relative to the other two methods in each setup, and the gains are dramatic when switching from the “NME” method to the “WME” method with the “NME” method ignoring observation errors for functional predictors. Specifically, for the case of $n = 100$, comparing with the “NME” method, the MISEs are reduced by 74%, 68% using the “WME” method with $(a_1, a_2) = (1.5, 0.2)$ and $(a_1, a_2) = (1.5, 0.4)$ respectively. For the “IN” method, it provides a reasonable estimator for γ and has better performance than the “NME” method, but nevertheless the “WME” method still performs better than “IN” method with improvement of 25%, 32% with respect to $(a_1, a_2) = (1.5, 0.2)$ and $(a_1, a_2) = (1.5, 0.4)$. In addition, these simulation results show that the MISEs decrease with increasing sample size n that is consistent with the derived theoretical results.

To sum up, in the scenario that incomplete functional predictors observed without measurement error, the “NME” method taking advantage of the whole information of the dataset produces a better estimator compared with the “SUB” method; in the scenario that incomplete functional predictors observed with measurement error, the “WME” method is preferred for giving the smallest MISE relative to the “IN” and “NME” methods. Both MISEs of the estimators of γ decrease with increasing sample size n , that is consistent with the derived theoretical properties.

5 Real Data Analysis

A real diffusion tensor imaging (DTI) dataset considered here is from NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study with 212 subjects, and is obtained through <http://adni.loni.usc.edu/>. The primary goal of ADNI study is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), biological markers, and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). DTI obtained using mathematical method to represent the anisotropic diffusion of the water molecule in brain organization, can be used to learn MCI and AD. The concrete measure of anisotropy include fractional anisotropy (FA), relative anisotropy (RA), Volume ratio (VR), and FA is commonly adopted for its advantage in contrast ratio of grey-white matter. More details about preprocessing and methods of this study can be found in Zhu et al. (2012), Yu et al. (2016).

Our main interest is characterizing the dynamic relationship between FA and mini-mental state examination (MMSE) score which is seen as a reliable and valid clinical measure in quantitatively assessing the severity of cognitive impairment. FA is measured at 83 equally spaced grid along the corpus callosum (CC) fiber tract that is the largest fiber tract in human brain, and is responsible for much of the communication between two hemispheres, and connects homologous areas in two cerebral hemispheres.

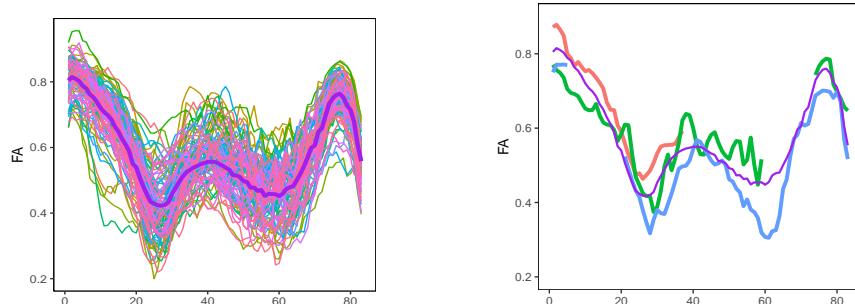


Fig. 1 A part of complete (left) and incomplete (right) FA curves with mean function (purple line)

To demonstrate the usefulness of the proposed method in Section 3.1, we artificially delete some observed points of FA, and then compare the estimator of γ obtained by using these incomplete functional observations with the estimator obtained by applying original complete dataset. For the i th FA curve, the missing domain has the same form with the interval given in Section 4 with $(a_1, a_2) = (1.5, 0.2)$ and $(a_1, a_2) = (1.5, 0.4)$. A part of complete and incomplete individual trajectories are displayed in Figure 1.

Estimators of functional coefficient obtained by both complete and incomplete FA dataset are illustrated in Figure 2. It shows that estimators obtained by incomplete

dataset with different missing domain (red line and green line) are similar to the estimator obtained from original complete dataset (blue line). This reveals that the proposed framework is useful in getting an estimator for the model with incomplete functional predictors.

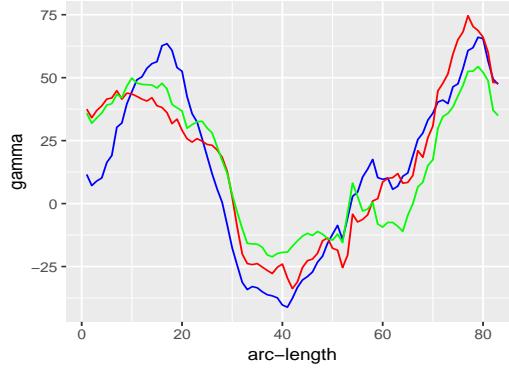


Fig. 2 Estimators of γ with different expected missing length on $[0, 1]$. Blue line: the estimator using original complete dataset; Red line: the estimator with $(a_1, a_2) = (1.5, 0.2)$; Green line: the estimator with $(a_1, a_2) = (1.5, 0.4)$

Next, we focus on the problem of recovering the missing parts X_{iM_i} of X_i . Assume that the infinite-dimensional process X_i is well approximated by the projection onto the function space $L^2(\mathcal{T})$ via the first m eigenfunctions (Yao et al. (2005a)). In practice, the prediction for the trajectory $X_i(t)$ of the i th subject using the first m eigenfunctions given in Section 3.1 can be approached by

$$\hat{X}_i(t) = \hat{\mu}^{\text{NME}}(t) + \sum_{k=1}^m \hat{U}_{ij}^{(\rho)} \hat{\phi}_j^{\text{NME}}(t).$$

We randomly select four FA curves with different missing parts. The predicted profiles for these four curves are presented in Figure 3, showing that the predicted profiles are close to the real part. This demonstrate the “NME” method by recovering the missing parts of incomplete trajectories encourages a better estimator comparing with the “SUB” method with deleting them directly.

6 Discussion

In this paper, we address the problem for getting estimators of γ in (1) with partially observed trajectories without and with measurement error. Basic elements of our approach are estimators of FPC scores for each partially observed trajectory. Specially, in the scenario that incomplete functional predictors observed without

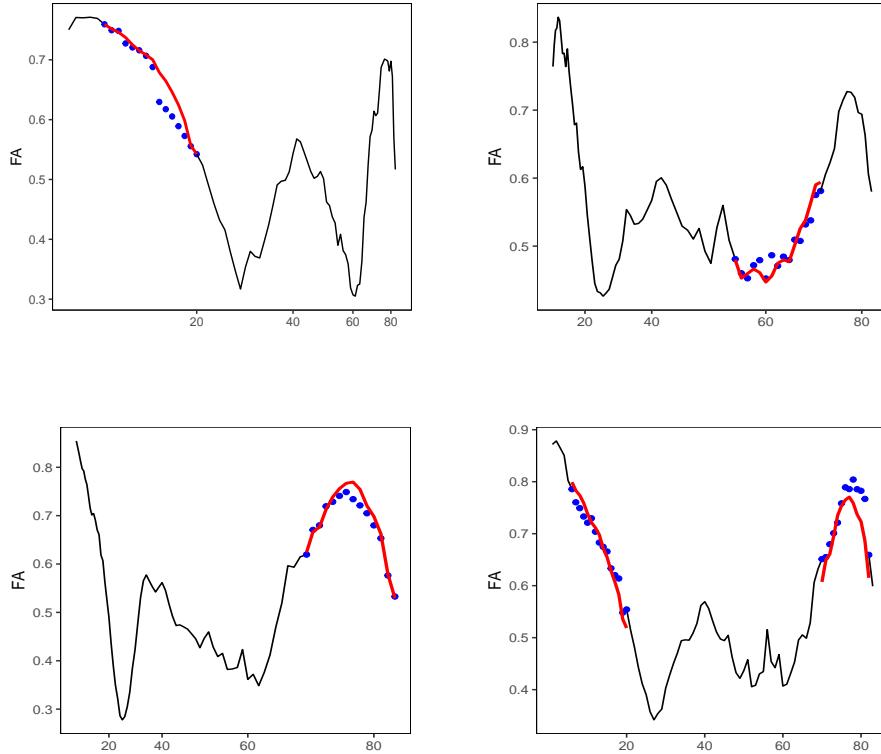


Fig. 3 Predicted profiles for four randomly chosen FA curves with different missing parts with $(a_1, a_2) = (1.5, 0.2)$. Missing parts of trajectories from left to right and top to down: missing in left side, middle side, right side, both left and right side. Blue point: real data point; Red line: predicted profile

measurement error, we achieve it by modeling FPC scores of the missing part as linear functionals of the observed part of that trajectory. In the scenario where incomplete functional data is observed with measurement error, we obtain estimators of FPC scores via conditional expectation. Rates of convergence of the proposed estimators $\hat{\gamma}^{\text{NME}}$, $\hat{\gamma}^{\text{WME}}$ under different scenarios are established. We also compare the proposed methods with the “SUB” or “IN” method. We conclude from simulation studies that both the “NME” and “WME” methods borrowing strength from entire samples to get estimators of γ in model (1) perform well in practice.

The methods proposed here can be extended to other models in terms of functional regression with partially observed trajectories, such as partial functional linear regression (see Shin (2009)). The framework established in this paper is based on the assumption that missing parts of trajectories are missing completely at random. In a number of applications, it is common to encounter that the underlying missing mechanism for dataset depends on systematic strategies (Liebl and Rameseder

(2019)) that clearly violate MCAR assumption. Extension to this scenario is also of interest and significance in practice.

Appendix

Lemma 1 (Kraus (2015), Proposition 1.)

a) Let $E\|X\|^2 < \infty$ and assumption (A1) be satisfied. Then $E(\|\hat{\mu}^{\text{NME}} - \mu\|^2) = O(n^{-1})$ for $n \rightarrow \infty$.

b) Let $E\|X\|^4 < \infty$ and observation pattern (A2) holds. Then $E(\|\hat{C}_X^{\text{NME}} - C_X\|_S^2) = O(n^{-1})$ for $n \rightarrow \infty$ (here $\|\cdot\|_S$ denotes the Hilbert-Schmidt norm).

Lemma 2 (Kneip and Liebl (2020), Theorem 4.1.)

Under the assumptions (B1)-(B5), we have that

(a) $\sup_{t \in \mathcal{T}} |\hat{\mu}^{\text{WME}}(t) - \mu(t)| = O_p(r_\mu)$ with $r_\mu = h_\mu^2 + 1/\sqrt{nNh_\mu} + 1/\sqrt{n}$.

(b) $\sup_{(s,t) \in \mathcal{T}^2} |\hat{c}^{\text{WME}}(s,t) - c_X(s,t)| = O_p(r_\mu + r_c)$ with $r_c = h_c^2 + 1/\sqrt{nMh_c^2} + 1/\sqrt{n}$.

Proof of Theorem 1. The following results can be derived from the theory developed by Bhatia et al. (1983):

$$\sup_{j \geq 1} |\hat{\lambda}_j^{\text{NME}} - \lambda_j| \leq \|\hat{C}_X^{\text{NME}} - C_X\|, \quad \sup_{j \geq 1} \delta_j \|\hat{\phi}_j^{\text{NME}} - \phi_j\| \leq 8^{1/2} \|\hat{C}_X^{\text{NME}} - C_X\|. \quad (13)$$

Therefore, we obtain from Lemma 1,

$$\begin{aligned} \sup_{j \geq 1} |\hat{\lambda}_j^{\text{NME}} - \lambda_j| &= O_p(n^{-1/2}), \\ \sup_{j \geq 1} \delta_j \|\hat{\phi}_j^{\text{NME}} - \phi_j\| &= O_p(n^{-1/2}). \end{aligned} \quad (14)$$

Note that,

$$\begin{aligned} &\int_{\mathcal{T}} (\hat{\gamma}^{\text{NME}}(s) - \gamma(s))^2 ds \\ &= \int_{\mathcal{T}} \left\{ \sum_{j=1}^{m-1} \left[\frac{n^{-1} \sum_{i=1}^n [Y_i \hat{U}_{ij}^{\text{NME}}]}{\hat{\lambda}_j^{\text{NME}}} \hat{\phi}_j^{\text{NME}}(s) - \frac{E[YU_j]}{\lambda_j} \phi_j(s) \right] \right\}^2 ds \\ &\quad + \int_{\mathcal{T}} \left\{ \sum_{j=m}^{\infty} \frac{E[YU_j]}{\lambda_j} \phi_j(s) \right\}^2 ds \\ &\quad + 2 \int_{\mathcal{T}} \left\{ \sum_{j=1}^{m-1} \left[\frac{n^{-1} \sum_{i=1}^n [Y_i \hat{U}_{ij}^{\text{NME}}]}{\hat{\lambda}_j^{\text{NME}}} \hat{\phi}_j^{\text{NME}}(s) - \frac{E[YU_j]}{\lambda_j} \phi_j(s) \right] \right\} \left\{ \sum_{j=m}^{\infty} \frac{E[YU_j]}{\lambda_j} \phi_j(s) \right\} ds \\ &:= A_1(n) + A_2(n) + A_3(n). \end{aligned} \quad (15)$$

For simplicity, we suppress the notation on “NME”. Assumption (A6) implies that $A_2(n) \rightarrow 0$ as $m \rightarrow \infty$. For $A_3(n)$, Cauchy-Schwarz inequality implies that

$A_3^2(n) \leq A_1^2(n) \times A_2^2(n) \xrightarrow{P} 0$. Combing the result (14), and the formula (15), we see that the result of the theorem follows if we can get the convergence rate of \hat{U}_{ij} of the trajectories per subject with $\hat{U}_{ij} = \hat{U}_{ijO_i} + \hat{U}_{ijM_i}^{(\alpha)}$.

Denote the estimates of U_{ijM_i} , $C_{O_iO_i}$, $C_{O_iM_i}$, ϕ_{jM_i} as $\hat{U}_{ijM_i(-i)}$, $\hat{C}_{O_iO_i(-i)}$, $\hat{C}_{O_iM_i(-i)}$, $\hat{\phi}_{jM_i(-i)}$ with deleting the i th curves $X_i(t)$. Let $\tilde{\xi}_{ijM_i}^{(\rho)} = (C_{O_iO_i}^{(\rho)})^{-1} C_{O_iM_i} \phi_{jM_i}$ with $C_{O_iO_i}^{(\rho)} = C_{O_iO_i} + \rho \mathcal{F}_{O_i}$, $\tilde{U}_{ijM_i}^{(\rho)} = \langle \tilde{\xi}_{ijM_i}^{(\rho)}, X_{iO_i} \rangle$, and the notation $\tilde{\xi}_{ijM_i} \tilde{U}_{ijM_i}$ are corresponded to the symbols $\tilde{\xi}_{ijM_i}^{(\rho)}, \tilde{U}_{ijM_i}^{(\rho)}$ with $\rho = 0$. Since

$$\begin{aligned} E\|\hat{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}\|^2 &= E\|\hat{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}^{(\rho)} + \tilde{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}\|^2 \\ &= 2E\|\hat{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}^{(\rho)}\|^2 + 2\|\tilde{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}\|^2 \\ &\leq 4E\|\hat{U}_{ijM_i}^{(\rho)} - \hat{U}_{ijM_i(-i)}^{(\rho)}\|^2 + 4E\|\hat{U}_{ijM_i(-i)}^{(\rho)} - \tilde{U}_{ijM_i}^{(\rho)}\|^2 \\ &\quad + 2\|\tilde{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}\|^2, \end{aligned} \quad (16)$$

we then analyze the terms $E\|\hat{U}_{ijM_i}^{(\rho)} - \hat{U}_{ijM_i(-i)}^{(\rho)}\|^2$, $E\|\hat{U}_{ijM_i(-i)}^{(\rho)} - \tilde{U}_{ijM_i}^{(\rho)}\|^2$, $\|\tilde{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}\|^2$ in turn. Let $\hat{\xi}_{ijM_i(-i)}^{(\rho)} = (\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1} \hat{C}_{O_iM_i(-i)} \hat{\phi}_{jM_i(-i)}$. Then

$$\begin{aligned} E\|\hat{U}_{ijM_i(-i)}^{(\rho)} - \tilde{U}_{ijM_i}^{(\rho)}\|^2 &= E\langle \hat{\xi}_{ijM_i(-i)}^{(\rho)} - \tilde{\xi}_{ijM_i}^{(\rho)}, X_{iO_i} \rangle^2 \\ &= E\{E[\langle \hat{\xi}_{ijM_i(-i)}^{(\rho)} - \tilde{\xi}_{ijM_i}^{(\rho)}, X_{iO_i} \rangle^2 | \{X_{kO_i}, k \neq i\}]\} \\ &= E\|C_{O_iO_i}^{1/2}((\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1} \hat{C}_{O_iM_i(-i)} \hat{\phi}_{jM_i(-i)} - (C_{O_iO_i}^{(\rho)})^{-1} C_{O_iM_i} \phi_{jM_i})\|^2 \\ &\leq 4 \left\{ E\|C_{O_iO_i}^{1/2}(\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1} (\hat{C}_{O_iM_i(-i)} - C_{O_iM_i})(\hat{\phi}_{jM_i(-i)} - \phi_{jM_i})\|^2 \right. \\ &\quad + E\|C_{O_iO_i}^{1/2}(\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1} C_{O_iM_i}(\hat{\phi}_{jM_i(-i)} - \phi_{jM_i})\|^2 \\ &\quad + E\|C_{O_iO_i}^{1/2}(\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1} (\hat{C}_{O_iM_i(-i)} - C_{O_iM_i})\phi_{jM_i}\|^2 \\ &\quad \left. + E\|C_{O_iO_i}^{1/2}((\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1} - (C_{O_iO_i}^{(\rho)})^{-1}) C_{O_iM_i} \phi_{jM_i}\|^2 \right\} \\ &:= B_1 + B_2 + B_3 + B_4. \end{aligned} \quad (17)$$

Let $\mathcal{F}_m = \{\frac{\lambda_m}{2} < \hat{\lambda}_m < \frac{3}{2}\lambda_m\}$. Suppose the event \mathcal{F}_m holds. Otherwise, we have $P(|\hat{\lambda}_m - \lambda_m| \geq \frac{\lambda_m}{2}) \leq P(\|\hat{C}_X^{\text{NME}} - C_X\| \geq \frac{\lambda_m}{2}) \rightarrow 0$ from assumption (A4). We have the following results for terms B_1 to B_4 with the equality

$$\left(\hat{C}_{O_iO_i(-i)}^{(\rho)}\right)^{-1} - \left(C_{O_iO_i}^{(\rho)}\right)^{-1} = (\hat{C}_{O_iO_i(-i)} - C_{O_iO_i}) \left(C_{O_iO_i}^{(\rho)}\right)^{-1} \left(\hat{C}_{O_iO_i(-i)}^{(\rho)}\right)^{-1}.$$

For the term B_1 ,

$$\begin{aligned} B_1 &\leq E[\|C_{O_iO_i}^{1/2}\|_2^2 \cdot \|(\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}\|_\infty^2 \cdot \|\hat{C}_{O_iM_i(-i)} - C_{O_iM_i}\|_2^2 \cdot \|\hat{\phi}_{jM_i(-i)} - \phi_{jM_i}\|^2] \\ &= O(n^{-2}\delta_j^{-2}) \cdot O(\rho^{-2}). \end{aligned}$$

Denote $\|\cdot\|_\infty$ as the operator norm. For the term B_2 , under the assumption (A7), $E\|C_{O_i O_i}^{1/2}\|_\infty^2 < \infty$ and the result (14), it is clear that

$$\begin{aligned} B_2 &\leq E[\|C_{O_i O_i}^{1/2}\|_\infty^2 \cdot \|(\hat{C}_{O_i O_i(-i)}^{(\rho)})^{-1} C_{O_i M_i}\|_2^2 \cdot \|\hat{\phi}_{j M_i(-i)} - \phi_{j M_i}\|^2] \\ &\leq \sum_j \sum_k \frac{r_{M_i O_i j k}^2}{(\lambda_{O_i O_i k} + \rho)^2} \cdot O(n^{-1} \delta_j^{-2}) = O(n^{-1} \delta_j^{-2}). \end{aligned}$$

For the term B_3 ,

$$\begin{aligned} B_3 &\leq E[\|C_{O_i O_i}^{1/2}\|_2^2 \cdot \|(\hat{C}_{O_i O_i(-i)}^{(\rho)})^{-1}\|_\infty^2 \cdot \|\hat{C}_{O_i M_i(-i)} - C_{O_i M_i}\|_2^2 \cdot \|\phi_{j M_i}\|^2] \\ &= O(n^{-1} \rho^{-2}). \end{aligned}$$

Note that $\frac{\rho \lambda_{O_i O_i k}}{(\lambda_{O_i O_i k} + \rho)^2} < 1$. Under the assumption (A7), we have that

$$\begin{aligned} B_4 &\leq E\left[\|C_{O_i O_i}^{1/2} \cdot (C_{O_i O_i}^{(\rho)})^{-1} \cdot (\hat{C}_{O_i O_i(-i)}^{(\rho)})^{-1} \cdot C_{O_i M_i}\|_2^2 \cdot \|\hat{C}_{O_i O_i(-i)} - C_{O_i O_i}\|_2^2 \cdot \|\phi_{j M_i}\|^2\right] \\ &\leq \left\{ \sum_j \sum_k \frac{\rho \lambda_{O_i O_i k}}{(\lambda_{O_i O_i k} + \rho)^2} \cdot \frac{r_{O_i M_i j k}^2}{(\lambda_{O_i O_i k} + \rho)^2} \cdot \rho^{-1} \right\} \cdot O(n^{-1}) \\ &= O(n^{-1}) \cdot O(\rho^{-1}). \end{aligned}$$

These results combined with (17) indicate

$$E\|\hat{U}_{ij M_i(-i)}^{(\rho)} - \tilde{U}_{ij M_i}^{(\rho)}\|^2 = O(n^{-1} \rho^{-2} + n^{-1} \delta_j^{-2}). \quad (18)$$

We then analyze $E\|\hat{U}_{ij M_i}^{(\rho)} - \hat{U}_{ij M_i(-i)}^{(\rho)}\|^2$,

$$\begin{aligned} E\|\hat{U}_{ij M_i}^{(\rho)} - \hat{U}_{ij M_i(-i)}^{(\rho)}\| &= E\langle \hat{\xi}_{ij M_i}^{(\rho)} - \hat{\xi}_{ij M_i(-i)}^{(\rho)}, X_{i O_i} \rangle \\ &\leq \{E\|\hat{\xi}_{ij M_i}^{(\rho)} - \hat{\xi}_{ij M_i(-i)}^{(\rho)}\|^2\}^{1/2} \{E\|X_{i O_i}\|^2\}^{1/2} \\ &\leq L \{E\|\hat{\xi}_{ij M_i}^{(\rho)} - \hat{\xi}_{ij M_i(-i)}^{(\rho)}\|^2\}^{1/2} \end{aligned} \quad (19)$$

where the last inequality holds from the finite second moment of X that is bounded by constant L . We also have,

$$\begin{aligned}
E\|\hat{\xi}_{ijM_i}^{(\rho)} - \hat{\xi}_{ijM_i(-i)}^{(\rho)}\|^2 &= E\|\left((\hat{C}_{O_iO_i}^{(\rho)})^{-1}\hat{C}_{O_iM_i} - (\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}\hat{C}_{O_iM_i(-i)}\right)\hat{\phi}_{jM_i(-i)}\|^2 \\
&= E\|\left((\hat{C}_{O_iO_i}^{(\rho)})^{-1} - (\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}\right)\hat{C}_{O_iM_i} \\
&\quad + (\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}(\hat{C}_{O_iM_i} - \hat{C}_{O_iM_i(-i)})\|\hat{\phi}_{jM_i(-i)}\|^2 \\
&\leq 2\{E\|\left((\hat{C}_{O_iO_i}^{(\rho)})^{-1} - (\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}\right)\hat{C}_{O_iM_i}\|^2 \\
&\quad + E\|(\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}(\hat{C}_{O_iM_i} - \hat{C}_{O_iM_i(-i)})\|^2\}. \tag{20}
\end{aligned}$$

Note that

$$\begin{aligned}
E\|\hat{C}_{O_iM_i} - \hat{C}_{O_iM_i(-i)}\|^2 &= O(n^{-2}), \\
E\|\left((\hat{C}_{O_iO_i}^{(\rho)})^{-1} - (\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}\right)\hat{C}_{O_iM_i}\|^2 &= O(n^{-2}), \\
E\|(\hat{C}_{O_iO_i(-i)}^{(\rho)})^{-1}(\hat{C}_{O_iM_i} - \hat{C}_{O_iM_i(-i)})\|^2 &= O(n^{-2}\rho^{-2}).
\end{aligned}$$

Combining formulas (19) and (20), we deduce that

$$E\|\hat{U}_{ijM_i}^{(\rho)} - \hat{U}_{ijM_i(-i)}^{(\rho)}\|^2 = O(n^{-2}\rho^{-2}). \tag{21}$$

On the other hand,

$$E\|\tilde{U}_{ijM_i}^{(\rho)} - \tilde{U}_{ijM_i}\|^2 = O(\rho), \tag{22}$$

$$\begin{aligned}
\text{var}(\tilde{U}_{ijM_i} - U_{ijM_i}) &= \langle \phi_{jM_i}, C_{M_iM_i} \phi_{jM_i} \rangle - \langle \phi_{jM_i}, C_{M_iO_i} C_{O_iO_i}^{-1} C_{O_iM_i} \phi_{jM_i} \rangle \\
&:= V_{ij}.
\end{aligned} \tag{23}$$

Therefore, with $n\rho^3 \rightarrow 0$ and the formulas (16), (18), (21)-(23), we have that

$$E\|\hat{U}_{ijM_i}^{(\rho)} - U_{ijM_i}\|^2 = O(n^{-1}\rho^{-2} + n^{-1}\delta_j^{-2} + V_{ij}).$$

Then the results is proved with $n\rho^3 \rightarrow 0$.

Proof of Theorem 2. Let $\tilde{\mathbf{U}}_i = (\tilde{U}_{i1}, \dots, \tilde{U}_{im})^T$, $\mathbf{U}_i = (U_{i1}, \dots, U_{im})^T$. The covariance matrix of $\tilde{\mathbf{U}}_i$ is $\text{var}(\mathbf{U}_i) = \Xi \Sigma_{\mathbf{Z}_i}^{-1} \Xi^T$ with $\Xi = \text{cov}(\tilde{\mathbf{U}}_i, \mathbf{Z}_i) = (\lambda_1 \phi_{i1}, \dots, \lambda_m \phi_{im})^T$. Moreover, $\text{var}(\tilde{\mathbf{U}}_i - \mathbf{U}_i) = \Lambda - \Xi \Sigma_{\mathbf{Z}_i} \Xi^T$. Combining these results with formulas (14), (12) and the results of Lemma 2, the result of Theorem 3 is obtained by replacing $\hat{U}_{ij}^{\text{NME}}$ with $\hat{U}_{ij}^{\text{WME}}$ in (15) with assumptions (B1)-(B6).

References

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, 51(2):285–311.

Bhatia, R., Davis, C., and McIntosh, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra and its Applications*, 52:45–67.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.

Che, M., Kong, L., Bell, R. C., and Yuan, Y. (2017). Trajectory modeling of gestational weight: A functional principal component analysis approach. *PLoS one*, 12(10):e0186761.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, 37(1):35–72.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of markov chains. *Biometrika*, 103(4):779–799.

Goldberg, Y., Ritov, Y., and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference*, 147:53–65.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91.

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

Hansen, P. C. (1990). The discrete picard condition for discrete ill-posed problems. *BIT Numerical Mathematics*, 30(4):658–672.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

Kneip, A. and Liebl, D. (2020). On the optimal reconstruction of partially observed functional data. *Annals of Statistics*, 48(3):1692–1717.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):777–801.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *The Annals of Applied Statistics*, 7(3):1562–1592.

Liebl, D. and Rameseder, S. (2019). Partially observed functional data: The case of systematically missing parts. *Computational Statistics and Data Analysis*, 131:104–115.

Marx, B. D. and Eilers, P. H. (1999). Generalized linear regression on sampled signals and curves: a p-spline approach. *Technometrics*, 41(1):1–13.

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

Ramsay, J. (2005). Functional data analysis. *Encyclopedia of Statistics in Behavioral Science*.

Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):539–561.

Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review*, 85(2):228–249.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243.

Riesz, F. and Nagy, S. (1955). B.(1990). functional analysis. *Dover Publications, Inc., New York. First published in*, 3(6):35.

Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference*, 139(10):3405–3418.

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 93(444):1403–1418.

Wang, Y., Kong, L., Jiang, B., Zhou, X., Yu, S., Zhang, L., and Heo, G. (2019). Wavelet-based lasso in functional linear quantile regression. *Journal of Statistical Computation and Simulation*, 89(6):1111–1130.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

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

Yu, D., Kong, L., and Mizera, I. (2016). Partial functional linear quantile regression for neuroimaging data analysis. *Neurocomputing*, 195:74–87.

Zhao, Y., Ogden, R. T., and Reiss, P. T. (2012). Wavelet-based lasso in functional linear regression. *Journal of computational and graphical statistics*, 21(3):600–617.

Zhu, H., Li, R., and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Annals of Statistics*, 40(5):2634.