Multiple imputation in the functional linear model with partially observed covariate and missing values in the response

Christophe Crambes, Chayma Daayeb, Ali Gannoun, Yousri Henchiri

Multiple imputation in the functional linear model with partially observed covariate and missing values in the response

Christophe Crambes[1†], Chayma Daayeb[1,2†], Ali Gannoun[1†] and Yousri Henchiri[2,3*†]

[1]Université de Montpellier, Institut Montpelliérain Alexander Grothendieck (IMAG), Place Eugène Bataillon, 34090, Montpellier, France.
[2]Université de Tunis El Manar, Laboratoire de Modélisation Mathématique et Numérique dans les Sciences de l'Ingénieur (ENIT-LAMSIN), B.P. 37, 1002, Tunis, Tunisie.
[3*]Université de la Manouba, Institut Supérieur des Arts Multimédia de la Manouba (ISAMM), Campus Universitaire de La Manouba, 2010, Tunis, Tunisie.

*Corresponding author(s). E-mail(s): yousri.henchiri@umontpellier.fr;
Contributing authors: christophe.crambes@umontpellier.fr; chayma.daayeb@umontpellier.fr; ali.gannoun@umontpellier.fr;
†These authors contributed equally to this work.

Abstract

Missing data problems are common and difficult to handle in data analysis. Ad hoc methods such as simply removing cases with missing values can lead to invalid analysis results. In this paper, we consider a functional linear regression model with partially observed covariate and missing values in the response. We use a reconstruction operator that aims at recovering the missing parts of the explanatory curves, then we are interested in regression imputation method of missing data on the response variable, using functional principal component regression to estimate the functional coefficient of the model. We study the asymptotic behavior of the prediction error when missing values in a original dataset are imputed by multiple sets of plausible values.

Keywords: Functional linear model, Missing data, Functional Principal Components, Missing At Random, Missing Completely At Random, Multiple imputation.

MSC Classification: 62G20 , 62G09 , 62J05 , 62F12.

1 Introduction

Functional data analysis (FDA) can be seen as a important subfield of statistics that has reached a certain maturity. FDA methods have been applied quite broadly in medicine, science, business, engineering, ..., while new theoretical and methodological developments regularly appear. For a more comprehensive treatment of FDA theory and methods, readers are referred to the classic monographs [1–3], recent monographs [4–6] and review papers [7, 8].

The functional linear model with scalar response in which a functional random variable is used to predict a real random variable has been the object of considerable attention in the literature. Several procedures have been proposed to the prediction and estimation problems under this model including, for example, functional principal component regression [9].This procedure has been considered by many authors [10–13] and [8]. Considering the functional linear regression methodology described in [2, Chapter 10], we observe the sample $\mathcal{D}_n \triangleq \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where the X_i's are independent and identically distributed with the same law as a random function X taking values in the space $\mathbb{L}_2(\mathcal{I})$ of square integrable functions defined on an interval $\mathcal{I} \subset \mathbb{R}$, and the real responses Y_i's are generated by the regression model

$$Y_i = \alpha + \int_{\mathcal{I}} \theta(t) X_i(t) \mathrm{d}t + \varepsilon_i, \tag{1}$$

for all $i = 1, \ldots, n$. Here, α is a constant corresponding to the intercept of the model, and θ is a square integrable function belonging to $\mathbb{L}_2(\mathcal{I})$, representing the slope function. It is supposed that the errors ε_i's are independent and identically distributed with finite variance and zero mean and independent from the explanatory variables X_i's.

The functional principal component regression methodology is based on spectral expansions of both the covariance operator of X and its estimator. We define the empirical cross covariance operator $\widehat{\Delta}_n$ given by $\widehat{\Delta}_n u = \frac{1}{n}\sum_{i=1}^n \langle X_i, u\rangle Y_i$ for all $u \in \mathbb{L}_2(\mathcal{I})$, the empirical covariance operator $\widehat{\Gamma}_n$ given by $\widehat{\Gamma}_n u = \frac{1}{n}\sum_{i=1}^n \langle X_i, u\rangle X_i$ for all $u \in \mathbb{L}_2(\mathcal{I})$. Denoting $(\widehat{\phi}_j)_{j=1,\ldots,k_n}$ the eigenfunctions associated to $\widehat{\Gamma}_n$ corresponding to the k_n highest eigenvalues $\widehat{\lambda}_1 > \ldots > \widehat{\lambda}_{k_n} > 0$ (where k_n is an integer depending on n), we define the orthogonal projection operator $\widehat{\Pi}_{k_n}$ onto the subspace $\mathrm{Span}(\widehat{\phi}_1, \ldots, \widehat{\phi}_{k_n})$ by $\widehat{\Pi}_{k_n} u = \sum_{j=1}^{k_n} \langle \widehat{\phi}_j, u\rangle \widehat{\phi}_j$ for all $u \in \mathbb{L}_2(\mathcal{I})$. Considering

$$\eta(X) \triangleq \alpha + \int_{\mathcal{I}} \theta(t)X(t)\mathrm{d}t, \tag{2}$$

we first estimate η based on a training sample $\mathcal{D}_n$. Let ℓ_n be a functional data fit that measures how well η fits the data. Then, the functional principal component regression estimator $\widehat{\eta}_n$ of η is given by

$$\widehat{\eta}_n \triangleq \mathrm{argmin}_{\eta_0} \left(\ell_n \left(\eta_0 \mid \mathcal{D}_n\right)\right), \tag{3}$$

where the minimization is taken over

$$\left\{\eta_0 \mid \eta_0(X) = \alpha_0 + \int_{\mathcal{I}} \theta_0(t)X(t)\mathrm{d}t : \alpha_0 \in \mathbb{R}, \theta_0 \in \mathrm{Span}\left(\widehat{\phi}_1, \ldots, \widehat{\phi}_{k_n}\right)\right\}.$$

The most common choice of the functional data fit is the mean square error

$$\ell_n(\eta_0 \mid \mathcal{D}_n) \triangleq \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \eta_0(X_i)\right)^2. \tag{4}$$

In general, ℓ_n is chosen such that it is convex in η_0 and $\mathbb{E}(\ell_n(\eta_0))$ in uniquely minimized by η. Equivalently, the minimization can be taken over (α_0, θ_0) to obtain estimates for both the intercept and slope, denoted by $\widehat{\alpha}$ and $\widehat{\theta}$, as follows

$$\widehat{\alpha} = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \widehat{\theta} = \sum_{j=1}^{k_n} \widehat{\mathsf{s}}_j \widehat{\phi}_j, \quad \text{with} \quad \widehat{\mathsf{s}}_j = \frac{1}{n\widehat{\lambda}_j}\sum_{i=1}^{n} \langle X_i, \widehat{\phi}_j\rangle Y_i. \tag{5}$$

In this work, we focus on the prediction problem. Let $\widehat{\eta}_n$ be a prediction rule given by

$$\widehat{\eta}_n(X_{new}) \triangleq \widehat{\alpha} + \int_{\mathcal{I}} \widehat{\theta}(t)X_{new}(t)\mathrm{d}t, \tag{6}$$

where X_{new} is a copy of X independent of $X_1, \ldots, X_n$. The prediction accuracy can be naturally measured by the excess risk

$$\begin{aligned} \mathcal{E}(\widehat{\eta}_n)(X_{new}) &\triangleq \mathbb{E}^{\star}\left(\widehat{\eta}_n(X_{new}) - \eta(X_{new})\right)^2 \\ &= \mathbb{E}^{\star}\left(\widehat{\alpha} + \langle\widehat{\theta}, X_{new}\rangle - \alpha - \langle\theta, X_{new}\rangle\right)^2 \end{aligned} \tag{7}$$

where $\mathbb{E}^{\star}$ stands for the expectation with respect to X_{new}.

2 Missing data mechanism

Earlier works on functional data focused in large part on regular functional data where data are fully observed. This may not always be the case, and missing data appear in many situations, for example when the measuring device breaks down. Many methods for the imputation of missing values have been developed. They can be divided into two branches, *single imputation* and *multiple imputation*. Single imputation consists in creating a single imputed value to replace a missing value. This procedure does not reflect the uncertainty about the prediction of the missing values during the imputation process. Multiple imputation is a statistical technique designed to take advantage of imputing a missing data several times. Each missing value is replaced by two or more imputed values in order to represent the uncertainty of the value to be imputed. For a comprehensive review of missing data mechanism and imputation methods, we refer the readers to a non-exhaustive list of monographs giving an overview of this topic: [14–17].

In recent years, applications producing partially observed functional data have emerged. Sometimes each individual trajectory is collected only over individual-specific subinterval(s), densely or sparsely, within the whole domain of interest. Several recent works have begun addressing the estimation of covariance functions for short functional segments observed at sparse and irregular grid points, called *functional snippets* [18–20] or for *fragmented functional data* observed on small subintervals [21]. For densely observed partial data, existing studies have focused on estimating the unobserved part of curves [22, 23], prediction [24], classification [25, 26], functional regression [27], and inferences [28, 29].

To go further, we describe two types of missing data mechanisms that will be the subject of our paper. The first one is related to the real response and the second one is related to the functional covariate. Concerning the missing data mechanism on the real response, we consider a dichotomous random variable $\delta^{[Y]}$ leading to the sample $(\delta_i^{[Y]})_{i=1,\dots,n}$ such that $\delta_i^{[Y]} = 1$ if the value Y_i is available and $\delta_i^{[Y]} = 0$ if the value Y_i is missing, for all $i = 1, \dots, n$. We consider that the data in the response is missing at random (MAR): the fact that the value Y is missing does not depend on the response of the model, but can possibly depend on the covariate, that is,

$$\mathbb{P}(\delta^{[Y]} = 1 \mid X, Y) = \mathbb{P}(\delta^{[Y]} = 1 \mid X). \tag{8}$$

MAR assumption implies that the distribution of Y is the same for units such that $\delta_i^{[Y]} = 1$ (observed units) as for those such that $\delta_i^{[Y]} = 0$ (non-observed units), conditionally on X. As a consequence, the variable $\delta^{[Y]}$ (the fact that an observation is missing or not) is independent of the error of the model ε. In the following, the number of missing values among $Y_1, \dots, Y_n$ is denoted

$$m_n^{[Y]} = \sum_{i=1}^{n} \mathbb{1}_{\left\{\delta_i^{[Y]}=0\right\}}.$$

Concerning the missing data mechanism on the functional covariate, we adopt the paradigm of partially observed functions as in [22] or [30]. More precisely, for each curve X_i, $i = 1, \ldots, n$, we consider the observed part $O_i \subseteq \mathcal{I}$ of X_i and the missing part $M_i = \mathcal{I} \setminus O_i$. The observed part O_i refers to an interval (or several intervals) where the curve X_i is observed at some measure points of O_i. Based on the punctual observations, the whole curve can be reconstructed on O_i with usual methods (e.g. smoothing splines, regression splines, local polynomial smoothing, ...). On the contrary, no information is available on the missing part M_i. For the rest of paper, we write "O" and "M" to denote a given production of O_i and M_i. In addition, we denote the observed and missing parts of X_i by X_i^O and X_i^M.

3 Multiple imputation: A deterministic and random imputation

3.1 A deterministic imputation

In this work, we have to deal with the situation in which some of the real responses of a data set generated from the functional linear model with scalar response are missing at random. This situation has been only considered in [31, 32]. Other recent works explore this context but in a nonparametric setting [33, 34] or in a functional partial linear regression setting [35, 36] or while the response is not missing at random [37]. More recently, [38] are interested in a more general case of missing data in functional linear regression: when the covariate is partially observed and when the response is affected by missing data. Following this latter paper [38, Subsection 2.1 and Subsection 2.2], $\widehat{\eta}_n$ can be calculated using the curve reconstruction method of [22, Section 2]. We give here some essential elements for our work: we consider a reconstruction problem relating the missing part of the curves to the observed part, writing

$$X_i^M(s) = L(X_i^O(t)) + \mathcal{Z}_i(s),$$

for all $t \in O$ and $s \in M$, where $L : \mathbb{L}_2(O) \to \mathbb{L}_2(M)$ is a linear reconstruction operator and $\mathcal{Z}_i \in \mathbb{L}_2(M)$ is the reconstruction error. Then, the optimal linear reconstruction operator, minimizing the following expected risk

$$\mathbb{E}\left(\left(X_i^M(u) - L(X_i^O)(u)\right)^2\right), \quad \text{for all} \quad u \in M,$$

is given by $\mathcal{L}(X_i^O)(u)$. This operator is estimated in [22, Section 2] by $\widehat{\mathcal{L}}_{k_n}(X_i^O)$, where the truncation parameter k_n is a positive integer that can be fixed automatically with a grid search. But note that the data structure implies that we are faced with two simultaneous estimation problems. One is efficient estimation of $\mathcal{L}(X_i^O)(u)$ for $u \in M$, the other one is a best possible estimation of the

function $X_i^O(t)$ for $t \in O$ from the observations $((W_{i1}, t_{i1}), \ldots, (W_{ip}, t_{ip}))$ with $W_{ij} = X_i^O(t_{ij})$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$ where $t_{ij} \in O$. In order to estimate the curve X_i^O and the covariance function $\gamma_s(t) = Cov(X_i^M(s), X_i^O(t))$ a nonparametric curve estimation by local polynomials smoothers is used [38, see Subsection 2.1 and Subsection 2.2]. In the following, we consider the whole sample $\widetilde{\mathcal{D}}_n \triangleq \left\{(X_1^\star, \delta_1^{[Y]}, Y_1), \ldots, (X_n^\star, \delta_n^{[Y]}, Y_n)\right\}$, with possibly reconstructed explanatory curves

$$X_i^\star(t) = \begin{cases} X_i^O(t) & \text{if } t \in O, \\ \widehat{\mathcal{L}}_{k_n}(X_i^O)(t) & \text{if } t \in M. \end{cases}$$

Using the exponent notation "obs" to make reference to the units for which the response is observed, we define the covariance operator with the reconstructed curves as follows

$$\widehat{\Gamma}_{n,rec}^{obs} = \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \langle X_i^\star, . \rangle \delta_i^{[Y]} X_i^\star.$$

Let $\widehat{\Pi}_{k_n,rec}^{obs}$ be the projection operator onto the subspace $\text{Span}(\widehat{\phi}_{1,rec}^{obs}, \ldots, \widehat{\phi}_{k_n,rec}^{obs})$ where $\widehat{\phi}_{1,rec}^{obs}, \ldots, \widehat{\phi}_{k_n,rec}^{obs}$ are the k_n first eigenfunctions of the covariance operator $\widehat{\Gamma}_{n,rec}^{obs}$. With analogous notations, $\widehat{\lambda}_{1,rec}^{obs}, \ldots, \widehat{\lambda}_{k_n,rec}^{obs}$ represent the k_n first eigenvalues of $\widehat{\Gamma}_{n,rec}^{obs}$.

The functional principal component regression estimator $\widetilde{\eta}_n$ of η is given by

$$\widetilde{\eta}_n \triangleq \text{argmin}_{\widetilde{\eta}_0} \left(\widetilde{\ell}_n \left(\eta_0 \mid \widetilde{\mathcal{D}}_n \right) \right), \tag{9}$$

where the minimization is taken over

$$\left\{ \eta_0 \mid \eta_0(X) = \alpha_0 + \int_{\mathcal{I}} \theta_0(t) X(t) \mathrm{d}t : \alpha_0 \in \mathbb{R}, \theta_0 \in \text{Span} \left(\widehat{\phi}_{1,rec}^{obs}, \ldots, \widehat{\phi}_{k_n,rec}^{obs} \right) \right\},$$

and

$$\widetilde{\ell}_n(\eta_0 \mid \widetilde{\mathcal{D}}_n) \triangleq \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \delta_i^{[Y]} \left(Y_i - \eta_0(X_i^\star) \right)^2. \tag{10}$$

Equivalently, the minimization can be taken over (α_0, θ_0) to obtain estimates for both the intercept and slope, for imputation, denoted by $\widetilde{\alpha}$ and $\widetilde{\theta}$ such that

$$\widetilde{\alpha} = \overline{Y}_{obs} = \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \delta_i^{[Y]} Y_i, \tag{11}$$

and

$$\widetilde{\theta} = \sum_{j=1}^{k_n} \widetilde{\mathsf{s}}_j \widehat{\phi}_{j,rec}^{obs}, \quad \text{with} \quad \widetilde{\mathsf{s}}_j = \frac{1}{(n - m_n^{[Y]})\widehat{\lambda}_{j,rec}^{obs}} \sum_{i=1}^{n} \langle X_i^\star, \widehat{\phi}_{j,rec}^{obs} \rangle \delta_i^{[Y]} Y_i. \quad (12)$$

For $i = 1, \ldots, n$, such that $\delta_i^{[Y]} = 1$, let $\widehat{Y}_i$ be the predicted value of Y_i given by

$$\widehat{Y}_i \triangleq \widetilde{\alpha} + \int_{\mathcal{I}} \widetilde{\theta}(t) X_i^\star(t) \mathrm{d}t. \quad (13)$$

Considering a missing value on the response, say Y_ℓ, such that $\delta_\ell^{[Y]} = 0$, we define the imputed value $Y_{\ell,imp}$ by

$$Y_{\ell,imp} = \widetilde{\eta}_n(X_\ell^\star) \triangleq \widetilde{\alpha} + \sum_{j=1}^{k_n} \widetilde{\mathsf{s}}_j \langle X_\ell^\star, \widehat{\phi}_{j,rec}^{obs} \rangle.$$

Finally, for $i = 1, \ldots, n$, we define

$$Y_i^\star = \delta_i^{[Y]} Y_i + \left(1 - \delta_i^{[Y]}\right) Y_{i,imp}. \quad (14)$$

3.2 A random imputation

We present in this section the random imputation which can be seen as a deterministic imputation plus a random noise (see [39]). Given an integer q, for a missing value Y_ℓ, we define

$$\widetilde{Y}_\ell^{(s)} \triangleq Y_{\ell,imp} + \varepsilon_\ell^{\star(s)},$$

for $s = 1, \ldots, q$, where $\varepsilon_\ell^{\star(s)}$ is drawn in the set

$$\left\{ e_i \mid e_i = \widetilde{e}_i - \overline{e}, i = 1, \ldots, n, \delta_i^{[Y]} = 1 \right\},$$

with

$$\widetilde{e}_i = \widetilde{\sigma}^{-1} \left(Y_i^\star - \widehat{Y}_i \right),$$

$$\widetilde{\sigma} = \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \delta_i^{[Y]} \left(Y_i^\star - \widehat{Y}_i \right)^2,$$

and

$$\overline{e} = \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \delta_i^{[Y]} \widetilde{e}_i.$$

This method is nonparametric as no distribution is assumed for the distribution of the standardized residuals observed e_i's. The imputation accuracy is measured by the excess risk

$$\mathcal{E}(\widetilde{\eta}_n)(X_\ell) = \mathbb{E}^\star\left(\widetilde{Y}_\ell^{(s)} - \alpha - \langle\theta, X_\ell^\star\rangle\right)^2 \tag{15}$$

where $\mathbb{E}^\star$ stands for the expectation with respect to X_ℓ.

Finally, for $i = 1, \ldots, n$, we define

$$Y_i^{\star(s)} = \delta_i^{[Y]} Y_i + \left(1 - \delta_i^{[Y]}\right) \widetilde{Y}_i^{(s)}. \tag{16}$$

3.3 Prediction

For $s = 1, \ldots, q$, given either the observed values or the random imputations $\widetilde{Y}_1^{\star(s)}, \ldots, \widetilde{Y}_n^{\star(s)}$, we estimate the parameters α and θ in model (1) with

$$\widehat{\alpha}^{(s)} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{\star(s)} \tag{17}$$

and

$$\widehat{\theta}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star\rangle Y_i^{\star(s)}}{\widehat{\lambda}_{j,rec}^\star}\widehat{\phi}_{j,rec}^\star = \sum_{j=1}^{k_n} \widehat{\mathsf{s}}_j^{(s)} \widehat{\phi}_{j,rec}^\star, \tag{18}$$

with

$$\widehat{\mathsf{s}}_j^{(s)} = \frac{1}{n\widehat{\lambda}_{j,rec}^\star}\sum_{i=1}^{n}\langle X_i^\star, \widehat{\phi}_{j,rec}^\star\rangle Y_i^{\star(s)},$$

and where the covariance operator is $\widehat{\Gamma}_{n,rec}^\star = \frac{1}{n}\sum_{i=1}^{n}\langle X_i^\star, .\rangle X_i^\star$, and $\widehat{\phi}_{1,rec}^\star, \ldots, \widehat{\phi}_{k_n,rec}^\star$ and $\widehat{\lambda}_{1,rec}^\star, \ldots, \widehat{\lambda}_{k_n,rec}^\star$ represent respectively the k_n first eigenfunctions and eigenvalues of the operator $\widehat{\Gamma}_{n,rec}^\star$.

For a new curve X_{new}, we predict the response value as follows

$$\widehat{Y}_{new} = \frac{1}{q}\sum_{s=1}^{q} \widehat{Y}_{new}^{\star(s)}, \tag{19}$$

where

$$\widehat{Y}_{new}^{\star(s)} = \widehat{\alpha}^{(s)} + \langle\widehat{\theta}^{(s)}, X_{new}^\star\rangle.$$

4 Theoretical results

4.1 Assumptions

In this subsection, we give the assumptions needed for our theoretical results. These assumptions are used in [22, 38] in order to control the curve reconstruction for the covariate.

(A.1) The variable X has a finite four moment order, that is $\mathbb{E}\left(\|X\|^4\right) < \infty$.

(A.2) Let $np \to \infty$ when $n \to \infty$ and $p = p(n)$. We assume $p = n^{\eta_1}$ with $0 < \eta_1 < \infty$ in the following.
(A.3) For any subinterval $O \subseteq \mathcal{I}$, we assume that the eigenvalues $\lambda_1 > \lambda_2 > \ldots > 0$ have multiplicity one. Moreover, we assume that there exist $a_O > 1$ and $0 < c_O < \infty$ such that (i) $\lambda_k^O - \lambda_{k+1}^O \geq c_O k^{-a_O - 1}$, (ii) $\lambda_k^O = \mathcal{O}(k^{-a_O})$, (iii) $1/\lambda_k^O = \mathcal{O}(k^{a_O})$ as $k \to \infty$.
(A.4) For any subinterval $O \subseteq \mathcal{I}$, we assume that there exists $0 < D_O < \infty$ such that the eigenfunctions satisfy $\sup_{t \in \mathcal{I}} \sup_{k \geq 1} \left\| \widetilde{\phi}_k^O(t) \right\|$
(A.5) The bandwidth h_X satisfies $h_X \to 0$ and $(ph_X) \to \infty$ as $p \to \infty$. For instance, we assume that $h_X = \frac{1}{n^{\eta_2}}$ with $0 < \eta_2 < \eta_1$. The bandwidth h_γ satisfies $h_\gamma \to 0$ and $(n(p^2 - p)h_\gamma) \to \infty$ as $n(p^2 - p) \to \infty$. For example, we can take $h_\gamma = \frac{1}{n^{\eta_3}}$ with $0 < \eta_3 < 2\eta_1 + 1$.
(A.6) Let κ_1 and κ_2 be nonnegative, second order univariate and bivariate kernel functions with support $[-1, 1]$. For example, we can use univariate and bivariate Epanechnikov kernel functions with compact support $[-1, 1]$, namely $\kappa_1(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{[-1,1]}(x)$ and $\kappa_2(x, y) = \frac{9}{16}(1 - x^2)(1 - y^2)\mathbb{1}_{[-1,1]}(x)\mathbb{1}_{[-1,1]}(y)$.
(A.7) The random variables X and Y are almost surely bounded, respectively in $\mathbb{L}_2(\mathcal{I})$ and $\mathbb{R}$.

Assumption **(A.1)** holds for many processes X (Gaussian processes, bounded processes). Assumption **(A.2)** is mild and can be satisfied even if the number of observation points p does not go fast to infinity. Assumptions **(A.3)** and **(A.4)**, related to eigenvalues and eigenfunctions of the covariance operator of X, are given in [22] in order to control the curve reconstruction for the covariate. In particular, a polynomial decrease of the eigenvalues is required, allowing a large class of eigenvalues for the covariance operator of X. Assumptions ***(A.5)*** and ***(A.6)*** are classic in the context of local polynomials smoothers. For Assumption ***(A.7)***, we can find in practice a large enough interval such that it is satisfied.

4.2 Asymptotic result

Theorem 1 *Under assumptions **(A.1)**-**(A.7)**, if we additionally take* $k_n \sim p^{1/(a_O+2)}$ *and* $p \sim n^{\eta_1}$ *with* $\eta_1 \leq 1/2$, *as well as* $m_n^{[Y]} = \mathcal{O}\left(n^{1-\eta_1(a_O+3)/4(a_O+2)}\right)$, *we get*

$$\mathbb{E}\left(\widehat{Y}_{new} - \alpha - \langle \theta, X_{new}^\star \rangle\right)^2 = \mathcal{O}_p\left(\frac{n^{-\eta_1(a_O-1)/(2(a_O+2))}}{q} + \frac{n^{\eta_1/(a_O+2)}}{q(n - m_n^{[Y]})}\right).$$

This result, giving the convergence rate of the prediction error after q random imputations, is asymptotically comparable to the convergence rate obtained in [38] in the case of a single deterministic imputation. We let the value of q appear in the convergence rate to highlight the fact that the constant

besides the convergence rate should be better in the case of several random imputations instead of a single deterministic one.

5 Proof of Theorem 1

Considering the decomposition of $\widehat{\theta}^{(s)}$, we write

$$\begin{aligned}
\widehat{\theta}^{(s)} &= \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i^{[Y]}=1}}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star \rangle Y_i}{\widehat{\lambda}_{j,rec}^\star} \widehat{\phi}_{j,rec}^\star \\
&\quad + \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i^{[Y]}=0}}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star \rangle \left(Y_{i,imp} + \varepsilon_i^{\star(s)} \right)}{\widehat{\lambda}_{j,rec}^\star} \widehat{\phi}_{j,rec}^\star \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star \rangle Y_i^\star}{\widehat{\lambda}_{j,rec}^\star} \widehat{\phi}_{j,rec}^\star \\
&\quad + \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i^{[Y]}=0}}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star \rangle \left(Y_{i,imp} + \varepsilon_i^{\star(s)} \right)}{\widehat{\lambda}_{j,rec}^\star} \widehat{\phi}_{j,rec}^\star,
\end{aligned}$$

hence

$$\begin{aligned}
\widehat{Y}_{new}^{\star(s)} - \alpha - \langle \theta, X_{new}^\star \rangle &= \widehat{\alpha}^{(s)} + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star \rangle Y_i^\star}{\widehat{\lambda}_{j,rec}^\star} \langle \widehat{\phi}_{j,rec}^\star, X_{new}^\star \rangle - \alpha - \langle \theta, X_{new}^\star \rangle \\
&\quad + \frac{1}{n} \sum_{\substack{i=1 \\ \delta_i^{[Y]}=0}}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi}_{j,rec}^\star \rangle \left(Y_{i,imp} + \varepsilon_i^{\star(s)} \right)}{\widehat{\lambda}_{j,rec}^\star} \langle \widehat{\phi}_{j,rec}^\star, X_{new}^\star \rangle.
\end{aligned}$$

We obtain from [38] the convergence rate for the first term of the decomposition

$$\begin{aligned}
&\mathbb{E} \left(\widehat{\alpha}^{(s)} + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^\star, \widehat{\phi^\star}_{j,rec} \rangle Y_i^\star}{\widehat{\lambda}_{j,rec}^\star} \langle \widehat{\phi}_{j,rec}^\star, X_{new}^\star \rangle - \alpha - \langle \theta, X_{new}^\star \rangle \right)^2 \\
&= \mathcal{O}_p \left(n^{-\eta_1 (a_O - 1)/(2(a_O+2))} + \frac{n^{\eta_1/(a_O+2)}}{n - m_n^{[Y]}} \right).
\end{aligned}$$

For the second term, we first use the boundedness of X and Y, which allows to bound $\varepsilon_i^{\star(s)}$, hence

$$\mathbb{E}\left(\frac{1}{n}\sum_{\substack{i=1\\ \delta_i^{[Y]}=0}}^{n}\sum_{j=1}^{k_n}\frac{\langle X_i^\star,\widehat{\phi}_{j,rec}^\star\rangle\left(Y_{i,imp}+\varepsilon_i^{\star(s)}\right)}{\widehat{\lambda}_{j,rec}^\star}\langle\widehat{\phi}_{j,rec}^\star,X_{new}^\star\rangle\right)^2$$
$$=\mathcal{O}_p\left(\frac{(m_n^{[Y]})^2k_n^2}{n^2}\right).$$

As a consequence, with the assumptions

$$k_n\sim n^{\eta_1/(a_O+2)}\text{ and }m_n^{[Y]}=\mathcal{o}\left(n^{1-\eta_1(a_O+3)/4(a_O+2)}\right),$$

we get

$$\mathbb{E}\left(\frac{1}{n}\sum_{\substack{i=1\\ \delta_i^{[Y]}=0}}^{n}\sum_{j=1}^{k_n}\frac{\langle X_i^\star,\widehat{\phi}_{j,rec}^\star\rangle\left(Y_{i,imp}+\varepsilon_i^{\star(s)}\right)}{\widehat{\lambda}_{j,rec}^\star}\langle\widehat{\phi}_{j,rec}^\star,X_{new}^\star\rangle\right)^2$$
$$=\mathcal{o}_p\left(n^{-\eta_1(a_O-1)/(2(a_O+2))}\right),$$

and the second term in the decomposition of $\widehat{Y}_{new}^{\star(s)}-\alpha-\langle\theta,X_{new}^\star\rangle$ is negligeable with respect to the first one. As a result, we obtain

$$\mathbb{E}\left(\widehat{Y}_{new}^{\star(s)}-\alpha-\langle\theta,X_{new}^\star\rangle\right)^2=\mathcal{O}_p\left(n^{-\eta_1(a_O-1)/(2(a_O+2))}+\frac{n^{\eta_1/(a_O+2)}}{n-m_n^{[Y]}}\right).$$

Finally, the mean over q iterations of the random imputation gives

$$\mathbb{E}\left(\widehat{Y}_{new}-\alpha-\langle\theta,X_{new}^\star\rangle\right)^2=\frac{1}{q^2}\sum_{s=1}^{q}\mathbb{E}\left(\widehat{Y}_{new}^{\star(s)}-\alpha-\langle\theta,X_{new}^\star\rangle\right)^2$$
$$=\mathcal{O}_p\left(\frac{n^{-\eta_1(a_O-1)/(2(a_O+2))}}{q}+\frac{n^{\eta_1/(a_O+2)}}{q(n-m_n^{[Y]})}\right).$$

References

[1] Ramsay, J.O., Silverman, B.W.: Aplied Functional Data Analysis: Methods and Case Studies. Springer, New York (2002)

[2] Ramsay, J.O., Silverman, B.W.: Functional Data Analysis (Second Edition). Springer, New York (2005)

[3] Ramsay, J.O., Hooker, G., Graves, S.: Functional Data Analysis with R and MATLAB. Springer, New York (2009)

[4] Hsing, T., Eubank, R.: Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley series in probability and statistics, John Wiley and Sons (2015)

[5] Srivastava, A., Klassen, E.P.: Functional and Shape Data Analysis. Springer, New York (2016)

[6] Kokoszka, P., Reimherr, M.: Introduction to Functional Data Analysis. Chapman and Hall, New York (2018)

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

[8] Wang, J.-L., Chiou, J.-M., Müller, H.-G.: Review of functional data analysis. Annual Review of Statistics and Its Application **3**, 257–295 (2016)

[9] Febrero-Bande, M., Galeano, P., González-Manteiga, W.: Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. International Statistical Review **85**, 61–83 (2017)

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

[11] Hall, P., Hosseini-Nasab, M.: On properties of functional principal components analysis. Journal of the Royal Statistical Society Series B (Statistical Methodology) **68**, 109–126 (2006)

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

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

[14] Rubin, D.B.: Multiple Imputation for Nonresponse in Surveys. John Wiley and Sons, New York (1987)

[15] Graham, J.W.: Missing Data Analysis and Design. Springer, New York (2012)

[16] Little, R.J.A., Rubin, D.B.: Statistical Analysis with Missing Data (Third Edition). John Wiley and Sons, New York (2020)

[17] He, Y., Zhang, G., Hsu, C.H.: Multiple Imputation of Missing Data in

Practice: Basic Theory and Analysis Strategies. John Wiley and Sons, New York (2022)

[18] Lin, Z., Wang, J.-L.: Mean and covariance estimation for functional snippets. Journal of the American Statistical Association (Just-accepted). https://doi.org/10.1080/01621459.2020.1777138

[19] Lin, Z., Wang, J.-L., Zhong, Q.: Basis expansions for functional snippets. Biometrika **108**, 709–726 (2021)

[20] Waghmare, K., Panaretos, V.M.: The completion of covariance kernels (2021). Preprint at https://arxiv.org/pdf/2107.07350.

[21] Delaigle, A., Hall, P., Huang, W., Kneip, A.: Estimating the covariance of fragmented and other related types of functional data. Journal of the American Statistical Association **116**, 35–72 (2020)

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

[23] Kraus, D., Stefanucci, M.: Ridge reconstruction of partially observed functional data is asymptotically optimal. Statistics and Probability Letters **165** (2020)

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

[25] Kraus, D., Stefanucci, M.: Classification of functional fragments by regularized linear classifiers with domain selection. Biometrika **106**, 161–180 (2018)

[26] Park, Y., Simpson, D.G.: Robust probabilistic classification applicable to irregularly sampled functional data. Computational Statistics and Data Analysis **131**, 37–49 (2019)

[27] Gellar, J.E., Colantuoni, E., Needham, D.M., Crainiceanu, C.M.: Variable- domain functional regression for modeling icu data. Journal of the American Statistical Association **109**, 1425–1439 (2014)

[28] Kraus, D.: Inferential procedures for partially observed functional data. Journal of Multivariate Analysis **173**, 583–603 (2019)

[29] Park, Y., Chen, X., Simpson, D.S.: Robust inference for partially observed functional response data (2021). Preprint at http://www3.stat.sinica.edu.tw/preprint/SS-2020-0358_Preprint.pdf.

[30] Kraus, D.: Components and completion of partially observed functional

data. Journal of the Royal Statistical Society: Series B **77**, 777–801 (2015)

[31] Crambes, C., Henchiri, Y.: Regression imputation in the functional linear model with missing values in the response. Journal of Statistical Planning and Inference **201**, 103–119 (2019)

[32] Febrero-Bande, M., Galeano, P., Gonzalez-Manteiga, W.: Estimation, imputation and prediction for the functional linear model with scalar response with responses missing at random. Computational Statistics and Data Analysis **131**, 91–103 (2019)

[33] Wang, L., Cao, R., Du, J., Zhang, Z.: A nonparametric inverse probability weighted estimation for functional data with missing response data at random. Journal of the Korean Statistical Society (2019)

[34] Rachdi, M., Laksaci, A., Kaid, Z., Benchiha, A., Al-Awadhi, F.A.: kNN local linear regression for functional and missing data at random. Statistica Neerlandica **28**, 1867–1886 (2020)

[35] Ling, N., Kan, R., Vieu, P., Meng, S.: Semi-functional partially linear regression model with responses missing at random. Metrika **82**, 39–70 (2019)

[36] Zhou, J., Peng, Q.: Estimation for functional partial linear models with missing responses. Statistics and Probability Letters **156** (2020)

[37] Li, T., Xie, F., Feng, X., Ibrahim, J.G., Zhu, H., the Alzheimers Disease Neuroimaging Initiative: Functional linear regression models for nonignorable missing scalar responses. Statistica Sinica **28**, 1867–1886 (2018)

[38] Crambes, C., Daayeb, C., Gannoun, A., Henchiri, Y.: Functional linear model with partially observed covariate and missing values in the response (2021). Preprint at https://hal.archives-ouvertes.fr/hal-03083293v2.

[39] Haziza, D.: Imputation and inference in the presence of missing data. In: Pfeffermann, D., Rao, C.R. (eds.) Handbook of Statistics: Sample Surveys: Design, Methods and Applications, vol. 29 A, pp. 215–256 (2009)