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[†] and Chayma Daayeb[†,‡] and Ali Gannoun[†] and Yousri Henchiri[‡,♮,*]

August 5, 2022

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 an original dataset are imputed by multiple sets of plausible values.

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

1 Introduction

Functional data analysis (FDA) can be seen as a important field 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 (Ramsay and Silverman, 2002, 2005; Ramsay et al., 2009), recent monographs (Hsing and Eubank, 2015; Srivastava and Klassen, 2016; Kokoszka and Reimherr, 2018) and review papers (Morris, 2015; Wang et al., 2016).

[*]Corresponding author. E-mail: yousri.henchiri@umontpellier.fr. Contributing authors: christophe.crambes@umontpellier.fr, chayma.daayeb@umontpellier.fr, ali.gannoun@umontpellier.fr. [†]Institut Montpelliérain Alexander Grothendieck (IMAG), Université de Montpellier, France. [‡]Université de Tunis El Manar, Laboratoire de Modélisation Mathématique et Numérique dans les Sciences de l'Ingénieur (LAMSIN), Tunis, Tunisie. [♮]Université de la Manouba, Institut Supérieur des Arts Multimédia de la Manouba (ISAMM), Tunisie.

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 (Febrero-Bande et al., 2017).

This procedure has been considered by many authors Cardot and Sarda (2003); Hall and Hosseini-Nasab (2006); Cai and Hall (2006); Hall and Horowitz (2007) and Wang et al. (2016). Considering the functional linear regression methodology described in Ramsay and Silverman (2005, Chapter 10), we observe the sample $\mathscr{D}_n \triangleq \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where the X_i's are centered independent and identically distributed with the same law as a random function X taking values in the space $\mathbb{L}_2(\mathscr{I})$ of square integrable functions defined on an interval $\mathscr{I} \subset \mathbb{R}$, and the real responses Y_i's are generated by the regression model

$$Y_i = \alpha + \int_{\mathscr{I}} \theta(t) X_i(t) \mathrm{d}t + \varepsilon_i, \tag{1.1}$$

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 $\mathbb{L}_2(\mathscr{I})$, 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 $u \in \mathbb{L}_2(\mathscr{I})$, 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 $u \in \mathbb{L}_2(\mathscr{I})$. 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 $u \in \mathbb{L}_2(\mathscr{I})$. Considering

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

we first estimate η based on a training sample $\mathscr{D}_n$. 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

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

where the minimization is taken over

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

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

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

In general, ℓ_n is chosen such to be 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 $\widehat{\theta}$ and $\widehat{\alpha}$, as follows

$$\widehat{\theta} = \sum_{j=1}^{k_n} \widehat{\mathsf{s}}_j \widehat{\phi}_j, \quad \text{with} \quad \widehat{\mathsf{s}}_j = \frac{1}{n\widehat{\lambda}_j}\sum_{i=1}^{n}\langle X_i, \widehat{\phi}_j\rangle Y_i, \tag{1.2}$$

and $\widehat{\alpha} = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

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

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

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

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

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

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 in 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: Rubin (1987); Graham (2012); Little and Rubin (2020); He et al. (2022).

In recent years, applications producing partially observed functional data have emerged. Sometimes each individual trajectory is collected only over individual-specific subintervals, 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* (Lin and Wang, 2020; Lin et al., 2021) or for *fragmented functional data* observed on small subintervals (Delaigle et al., 2020). For densely observed partial data, existing studies have focused on estimating the unobserved part of curves (Kneip and Liebl, 2020; Kraus and Stefanucci, 2020), prediction (Goldberg et al., 2014), classification (Kraus and Stefanucci, 2018; Park and Simpson, 2019), functional regression (Gellar et al., 2014), and inferences (Kraus, 2019; Park et al., 2022).

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,

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

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 Kneip and Liebl (2020) or Kraus (2015). More precisely, for each curve X_i, $i = 1, \dots, n$, we consider the observed part $O_i \subseteq \mathscr{I}$ of X_i and the missing part $M_i = \mathscr{I} \setminus O_i$. 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. As an example, we consider a data set from energy economics presenting demand and prices of the German power market which is shown in Figure 1. The data consist of partially observed price functions. The observation period corresponds to 241 working days from March 15, 2012 to March 14, 2013. Price curves can be seen as partially observed curves, as some prices cannot be observed with respect to some residual demand values. Here, the price-demand functions are observed on different domains. This distinguishes our functional

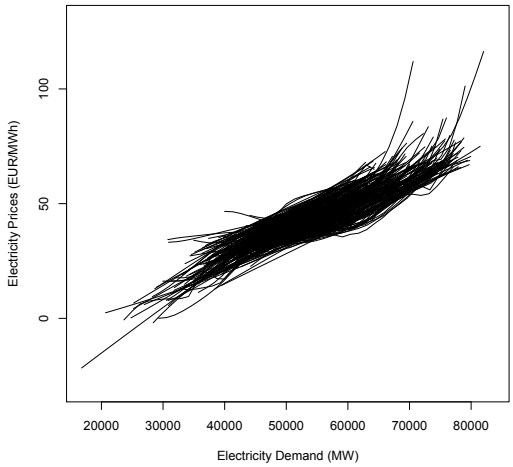

Figure 1: Daily electricity price curves in function of the residual demand.

data set from classical functional data sets, where all functions are observed on a common domain. We consider a standardized domain where the standardization can be achieved as follows: for $i = 1, \ldots, 241$, we consider a sequence from $\min_{1\leqslant j\leqslant p} t_{ij}$ to $\max_{1\leqslant j\leqslant p} t_{ij}$ with a regular step $(b-a)/p$, where $a := \min_{1\leqslant i\leqslant 241} \min_{1\leqslant j\leqslant p} t_{ij}$ and $b := \max_{1\leqslant i\leqslant 241} \max_{1\leqslant j\leqslant p} t_{ij}$.

The objective of this paper is to predict a new value of the response Y given a new test observation on the explanatory variable X once the partially observed curves X have been reconstructed and the missing data Y have been imputed with the multiple imputation method. More precisely, we want to obtain convergence rates for this prediction error, and we want to analyze how these convergence rates depend on the convergence rates of the reconstruction of the missing parts of the covariate and the convergence rates of the imputation error. We show the difference between the deterministic regression imputation, the random regression imputation and the multiple regression imputation, and its effect on the mean square error of prediction.

In the following, we give in section 2 theoretical results of the partially observed covariate. Then, in section 3, we study different methods of imputation and the prediction error when the covariate is partially observed and some observations of the real response are affected with missing data. Next, in section 4 we give theoretical results related to the prediction error. In section 5, we present some simulation results to show the behavior of the methods in practice. Section 6 is devoted to a real dataset application. Finally, all the proofs are postponed to section 7.

2 Reconstruction of partially observed covariate

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 Crambes and Henchiri (2019); Febrero-Bande et al. (2019). Other recent works explore this context but in a nonparametric setting (Wang et al., 2019; Rachdi et al., 2020) or in a functional partial linear regression setting (Ling et al., 2019; Zhou and Peng, 2020) or while the response is not missing at random (Li et al., 2018). More recently, Crambes et al. (2022) 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 (Crambes et al., 2022, Subsection 2.1 and Subsection 2.2), $\widehat{\eta}_n$ can be calculated using the curve reconstruction method of Kneip and Liebl (2020, 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

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

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 $\mathscr{Z}_i \in \mathbb{L}_2(M)$ 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 $\mathscr{L}(X_i^O)(u)$. This operator is estimated in Kneip and Liebl (2020, Section 2) by $\widehat{\mathscr{L}_{k_n}}(X_i^O)$, where the truncation parameter k_n is a positive integer that can be fixed automatically with a grid search. Note that the data structure implies that we are faced with two simultaneous estimation problems. One is efficient estimation of $\mathscr{L}(X_i^O)(u)$ for $u \in M$, the other one is the best possible estimation of the function $X_i^O(t)$ for $t \in O$ observed at p discretization points $((W_{i1}, t_{i1}), \ldots, (W_{ip}, t_{ip}))$ 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. Let κ_1 be a kernel and h_X be a bandwidth of the local linear smoothers of the curve X_i^O. Moreover, let κ_2 be a bivariate kernel and h_γ be a bandwidth of the local linear smoothers of the covariance function γ_s.

The goal is to rebuild a reconstruction function that allows us to recover the full functions from their partial observations. Coming back to the introducing example, Figure 2 shows the reconstructed curves with the method from Kneip and Liebl (2020).

In the following, we consider the whole sample $\widetilde{\mathscr{D}}_n \triangleq \left\{(X_1^\star, \delta_1^{[Y]}, Y_1), \ldots, (X_n^\star, \delta_n^{[Y]}, Y_n)\right\}$,

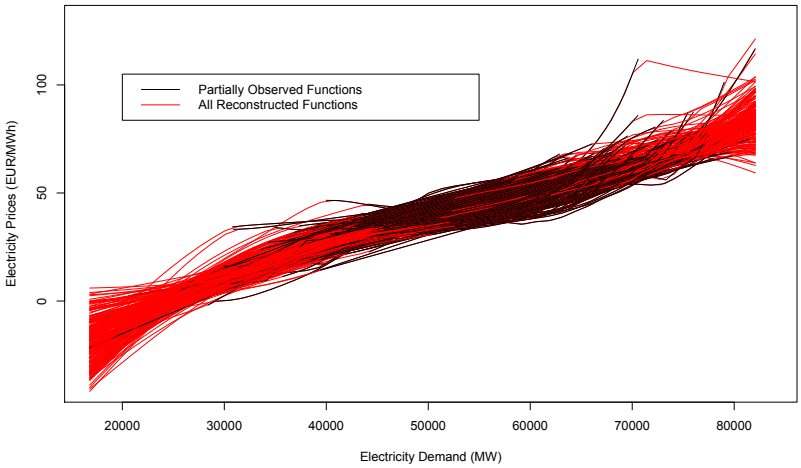

Figure 2: Reconstructed daily electricity price curves in function of the residual demand.

with possibly reconstructed explanatory curves

$$X_i^\star(t) = \begin{cases} X_i^O(t) & \text{if } t \in O, \\ \widehat{\mathscr{L}_{k_n}}(X_i^O)(t) & \text{if } t \in M. \end{cases} \tag{2.1}$$

Once the curves are reconstructed, we complete missing values in the response with deterministic and random imputation.

3 Multiple regression imputation

We may classify regression imputation methods into two classes : *deterministic* (or simple) and *random*. Deterministic regression method yields to a fixed imputed value given the observed sample if the imputation process were repeated as opposed to random methods that do not necessarily yield to the same imputed value. The deterministic method strengthens the relationships in the data and may lead to imputations which seem to be perfect for the model generated from the observed data. However, once the imputation is done, analyses then typically proceed as if the imputed values were the truth. This leads to overly optimistic measures of uncertainty and the potential for substantial bias (Buuren, 2018). To deal with this problem, we consider the random regression imputation that can be seen as a deterministic regression imputation with a random noise $\varepsilon^\star$ (Haziza, 2009, Subsection 2.2). This is a powerful concept, which also builds the basis of many modern missing values imputation approaches, as it takes into account the inherent uncertainty about missing values. The random noise, $\varepsilon^\star$, is drawn from the observed standardized residuals observed of the prediction errors.

In the following, we are interested in multiple imputation. This method consists in repeating q times the random regression imputation with $q \geqslant 2$. Multiple imputation creates multiple predictions for each missing value, the corresponding statistical analysis takes into account the uncertainty in the imputations and hence, yields to a more reliable standard error. In simple terms, if there is less information in the observed data regarding the missing values, the imputations will be more variable, leading to higher standard errors in the analysis. However, if the observed data allow to predict the missing values, the imputations will be more consistent across the multiple imputed data sets, resulting in smaller and more reliable standard errors (Greenland and Finkle, 1995). Finally, we will predict a new value under the functional linear model as the mean of all the predictive values.

3.1 Deterministic regression imputation

In this section, we follow the same steps as in Crambes et al. (2022). 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 (2.1) 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 $\mathrm{Span}(\widehat{\phi}_{1,rec}^{obs}, \ldots, \widehat{\phi}_{k_n,rec}^{obs})$ 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

$$\widetilde{\eta}_n \triangleq \mathrm{argmin}_{\widetilde{\eta}_0} \left(\widetilde{\ell}_n \left(\eta_0 \mid \widetilde{\mathscr{D}}_n \right) \right),$$

where the minimization is taken over

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

and

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

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

$$\widetilde{\alpha} = \overline{Y}_{obs} = \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \delta_i^{[Y]} Y_i, \tag{3.1}$$

and

$$\widetilde{\theta} = \sum_{j=1}^{k_n} \widetilde{\mathsf{s}}_j \widehat{\phi}_{j,rec}^{obs}, \quad \text{with} \quad \widetilde{\mathsf{s}}_j = \frac{1}{(n - m_n^{[Y]})\widehat{\lambda}_{j,rec}^{obs}} \sum_{i=1}^{n} \langle X_i^\star, \widehat{\phi}_{j,rec}^{obs} \rangle \delta_i^{[Y]} Y_i. \tag{3.2}$$

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

$$\widehat{Y}_i \triangleq \widetilde{\alpha} + \int_{\mathscr{I}} \widetilde{\theta}(t) X_i^\star(t) \mathrm{d}t. \tag{3.3}$$

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, we obtain the complete sample $(X_i^\star, Y_i^\star)$ for $i = 1, \ldots, n$, with

$$Y_i^\star = \delta_i^{[Y]} Y_i + \left(1 - \delta_i^{[Y]}\right) Y_{i,imp}. \tag{3.4}$$

The imputation accuracy is measured by the excess risk

$$\mathscr{E}(\widetilde{\eta}_n)(X_\ell) = \mathbb{E}^\star \left(\widetilde{\alpha} + \langle \widetilde{\theta}, X_\ell^\star \rangle - \alpha - \langle \theta, X_\ell^\star \rangle \right)^2,$$

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

3.2 Random regression imputation

We define the missing value Y_ℓ

$$\widetilde{Y}_\ell = \breve{\eta}_n(X_\ell^\star) \triangleq Y_{\ell,imp} + \varepsilon_\ell^\star, \tag{3.5}$$

where $\varepsilon_\ell^\star$ is drawn in the set

$$\left\{ e_i \mid e_i = \widetilde{e}_i - \overline{e}, i = 1, \ldots, n, \delta_i^{[Y]} = 1 \right\}, \tag{3.6}$$

using (3.3) and (3.4), we have

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

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

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.
Finally, we obtain the complete sample $(X_i^\star, \breve{Y}_i^\star)$ for $i = 1, \ldots, n$, with

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

Here, the imputation accuracy is measured by the excess risk

$$\mathscr{E}(\breve{\eta}_n)(X_\ell) = \mathbb{E}^\star \left(\widetilde{\alpha} + \langle \widetilde{\theta}, X_\ell^\star \rangle + \varepsilon_\ell^\star - \alpha - \langle \theta, X_\ell^\star \rangle\right)^2.$$

3.3 Multiple regression imputation

Let i be an index for the observed cases and ℓ be an index for the incomplete cases. The multiple imputation algorithm is sketched as follows:

Algorithm 1 The multiple imputation algorithm

Step 1. Estimating parameters $\widetilde{\alpha}$ and $\widetilde{\theta}$ from the functional linear model using complete sample $(X_i^\star, Y_i, \delta_i^{[Y]} = 1)$, for $i = 1, \ldots, n$, as in (3.1) and (3.2).
Step 2. Drawing $\varepsilon_\ell^{\star(w)}$ from the set of $\left\{e_i \mid e_i = \widetilde{e}_i - \overline{e}, i = 1, \ldots, n, \delta_i^{[Y]} = 1\right\}$, as in (3.6), for $\ell \in \widetilde{\mathscr{D}}_m$, where $\widetilde{\mathscr{D}}_m$ is the set of missing responses of size $m_n^{[Y]}$.
Step 3. Drawing the imputed values of missing data, as in (3.5), from

$$\widetilde{Y}_\ell^{(w)} = \widetilde{\alpha} + < \widetilde{\theta}, X_\ell^\star > + \varepsilon_\ell^{\star(w)},$$

for $\ell \in \widetilde{\mathscr{D}}_m$.
Step 4. Repeat Steps 2 to 3 independently q times to create multiple sets of imputations $(w = 1, ..., q)$.

Finally, we obtain the multiple sets of complete data $(X_i^\star, Y_i^{\star(w)})$, for $w = 1, \ldots, q$, with

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

Here, the imputation accuracy is measured by the excess risk

$$\mathscr{E}(\breve{\eta}_n)(X_\ell) = \mathbb{E}^\star \left(\frac{1}{q} \sum_{w=1}^{q} \left(\widetilde{\alpha} + < \widetilde{\theta}, X_\ell^\star > + \varepsilon_\ell^{\star(w)}\right) - \alpha - \langle \theta, X_\ell^\star \rangle\right)^2.$$

3.4 Prediction

Once the whole database has been reconstructed, we obtain estimates for both the intercept and slope, denoted by $(\widehat{\alpha}^\star, \widehat{\theta}^\star)$ and $(\breve{\alpha}^\star, \breve{\theta}^\star)$ respectively after deterministic regression imputation and after random regression imputation such that

$$\widehat{\alpha}^\star = \frac{1}{n} \sum_{i=1}^{n} Y_i^\star,$$

$$\widehat{\theta}^{\star} = \sum_{j=1}^{k_n} \widehat{s}_j^{\star} \widehat{\phi}_{j,rec}^{\star}, \quad \text{with} \quad \widehat{s}_j^{\star} = \frac{1}{n\widehat{\lambda}_{j,rec}^{\star}} \sum_{i=1}^{n} \langle X_i^{\star}, \widehat{\phi}_{j,rec}^{\star} \rangle Y_i^{\star}, \tag{3.7}$$

$$\breve{\alpha}^{\star} = \frac{1}{n} \sum_{i=1}^{n} \breve{Y}_i^{\star},$$

$$\breve{\theta}^{\star} = \sum_{j=1}^{k_n} \breve{s}_j^{\star} \widehat{\phi}_{j,rec}^{\star}, \quad \text{with} \quad \breve{s}_j^{\star} = \frac{1}{n\widehat{\lambda}_{j,rec}^{\star}} \sum_{i=1}^{n} \langle X_i^{\star}, \widehat{\phi}_{j,rec}^{\star} \rangle \breve{Y}_i^{\star}, \tag{3.8}$$

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

In multiple regression imputation setting, for $w = 1, \ldots, q$, given either the observed values or the random imputations $Y_1^{\star(w)}, \ldots, Y_n^{\star(w)}$, we estimate the parameters α and θ in model (1.1) with

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

and

$$\widehat{\theta}^{(w)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_n} \frac{\langle X_i^{\star}, \widehat{\phi}_{j,rec}^{\star} \rangle Y_i^{\star(w)}}{\widehat{\lambda}_{j,rec}^{\star}} \widehat{\phi}_{j,rec}^{\star} = \sum_{j=1}^{k_n} \widehat{\mathsf{s}}_j^{(w)} \widehat{\phi}_{j,rec}^{\star}, \tag{3.9}$$

with

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

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

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

where

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

An asymptotic behavior of the prediction error is given in Crambes et al. (2022) when the missing parts of the covariate are reconstructed and the missing values on the response are imputed by deterministic regression imputation. In the next section, we will study the convergence rate of this prediction error with multiple regression imputation.

4 Theoretical results

4.1 Assumptions

In this subsection, we give the assumptions needed for our theoretical results. Some assumptions are used in Kneip and Liebl (2020) and Crambes et al. (2022) in order to control the curve reconstruction of the covariate.

(A.1) 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.2) For any subinterval $O \subseteq \mathscr{I}$, 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) $\lambda_k^O - \lambda_{k+1}^O \geqslant c_O k^{-a_O - 1}$, (ii) $\lambda_k^O = \mathscr{O}(k^{-a_O})$, (iii) $1/\lambda_k^O = \mathscr{O}(k^{a_O})$ as $k \to \infty$.

(A.3) For any subinterval $O \subseteq \mathscr{I}$, we assume that there exists $0 < D_O < \infty$ such that the eigenfunctions satisfy $\sup_{t \in \mathscr{I}} \sup_{k \geqslant 1} \left\| \widetilde{\phi}_k^O(t) \right\| \leqslant D_O$, where $\widetilde{\phi}_k^O(s) = \langle \phi_k^O, \gamma_s \rangle / \lambda_k^O$.

(A.4) 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.5) 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.6) The random variables X and Y are almost surely bounded, respectively in $\mathbb{L}_2(\mathscr{I})$ and $\mathbb{R}$.

Assumption **(A.1)** is mild and can be satisfied even if the number of observation points p does not go fast to infinity. Assumptions **(A.2)** and **(A.3)**, related to eigenvalues and eigenfunctions of the covariance operator of X, are given in Kneip and Liebl (2020) 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.4)** and **(A.5)** are classic in the context of local polynomials smoothers. For Assumption **(A.6)**, we can find in practice a large enough interval such that it is satisfied.

4.2 Asymptotic result

To start this subsection, we give the main result from Crambes et al. (2022) for the prediction error when the missing parts of the covariate are reconstructed and the completion of the missing data in the response is done by deterministic imputation. Let Y_{new} be the predicted value of the response given a new observation X_{new} of the covariate.

Proposition 4.1. *Under assumptions* **(A.1)**-**(A.6)**, *and* $k_n \sim p^{1/(a_O+2)}$ *and* $p \sim n^{\eta_1}$ *with* $\eta_1 \leqslant 1/2$, *the prediction error, based on the deterministic regression imputation, is*

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

In the particular case where $\eta_1 = 1/2$, *the first term in the convergence rate is* $\mathscr{O}_p\big(n^{-(a_O-1)/(4(a_O+2))}\big)$.

This result shows that the prediction error rate with the deterministic regression imputation in the response is subordinate to the reconstruction error of the covariate. We now give our main result.

Theorem 4.2. *Under assumptions* **(A.1)**-**(A.6)**, *if we additionally take* $k_n \sim p^{1/(a_O+2)}$ *and* $p \sim n^{\eta_1}$ *with* $\eta_1 \leqslant 1/2$, *as well as* $m_n^{[Y]} = o\big(n^{1-\eta_1(a_O+3)/4(a_O+2)}\big)$, *the prediction error, based on the multiple regression imputation, is*

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

This result, giving the convergence rate of the prediction error after q random imputations, is asymptotically comparable to the convergence rate obtained in Proposition 4.1 in the case of a deterministic regression imputation. We let the value of q appear in the convergence rate to highlight the fact that the constant when the convergence rate should be better in the case of several random imputations instead of a single deterministic one.

Remark 4.3. *Theoretical results are generally obtained under assumptions concerning the rate of convergence of the integer* k_n. *In practice, this integer is selected by minimizing a certain empirical criterion. We chose the Generalized Cross Validation (GCV) procedure, known to be computationally fast. The GCV criterion is given below for imputation*

$$GCV(k_n) = \frac{(n - m_n^{[Y]}) \sum_{i=1}^{n} \left(\widetilde{\alpha} + \langle \widetilde{\theta}, X_i^{\star} \rangle - \alpha - \langle \theta, X_i^{\star} \rangle\right)^2 \delta_i^{[Y]}}{\left((n - m_n^{[Y]}) - k_n\right)^2},$$

and the analogous criterion for prediction

$$GCV(k_n) = \frac{n \sum_{i=1}^{n} \left(\widehat{\alpha}^{(w)} + \langle \widehat{\theta}^{(w)}, X_i^{\star} \rangle - \alpha - \langle \theta, X_i \rangle\right)^2}{\left(n - k_n\right)^2}, \quad \text{for } w = 1, \ldots, q.$$

5 Simulations

5.1 Methodology

We generated the functional covariate in a similar way to that adopted in Hall and Horowitz (2007). More specifically, the functional covariates were identically and independently generated as:

$$X_i(t) = \sum_{j=1}^{150} \zeta_{ij} \varrho_j \phi_j(t), \quad i = 1, \ldots, N,$$

where $\phi_1 \equiv 1$, $\phi_{j+1} = \sqrt{2}\cos(j\pi t)$, for $j \geqslant 2$, the ϱ_j's are defined by $\varrho_j = (-1)^{j+1}(j)^{-2}$ and the ζ_j's are independently sampled from the uniform distribution on $[-\sqrt{3}, \sqrt{3}]$. The covariance function writes

$$cov(X(t), X(s)) = \sum_{j=1}^{150} \frac{2}{j^4} \cos(j\pi t) \cos(j\pi s).$$

These covariates are sampled at $p = 100$ equally spaced points between 0 and 1. The responses are generated from (1.1), where $\alpha = 3$ and θ defined, for all $t \in [0, 1]$, by

$$\theta(t) = \sum_{j=1}^{50} b_j \phi_j(t),$$

where $b_1 = 0.3$ and $b_j = 4(-1)^{j+1} j^{-2}$ for all $j > 1$. The random errors, ε_i's, are generated as $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = 0.2$. In each simulation replicate we randomly generate $n = \frac{4}{5}N$ independent copies of (X_i, Y_i) for training and $n_1 = \frac{1}{5}N$ copies for testing, with $N = 1400$. To better assess prediction performance of model, we repeat the simulation procedure $\boldsymbol{S} = 250$ times.

To deal with partially observed curves for the covariate, we adopted the missing data simulation scenario from Crambes et al. (2022) such that

- 70% (respectively 55%) of the curves are fully observed on $[0, 1]$,
- for the 30% (respectively 45%) of partially observed curves, the curve X_i is fully observed on $[A_i, B_i] \subset [0, 1]$ with A_i drawn with uniform law on the interval $[0, A]$ and $B_i = A_i + B$, with $A = 3/8$ and $B = 6/8$.

We simulate the number of missing data on the response Y and the indicator $\delta^{[Y]}$ by the logistic functional regression. The variable δ follows the Bernoulli law with parameter $p(X)$ such that

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \langle c, X \rangle + ct,$$

where $c = \sin(2\pi t)$ for all $t \in [0, 1]$ and ct is a constant allowing to take different levels of missing data. For exemple $ct = 1$ for around 26.903% of missing data, $ct = 0.2$ for around 44.941% of missing data and $ct = -0.2$ for around 54.793% of missing data.

We estimate the parameters of the model and we obtain the predicted values of the response with imputation methods. Notice that, we use a smoothed version of the different estimators (1.2), (3.2), (3.7), (3.8) and (3.9) based on the Smooth Principal Components Regression (SPCR). Let us remark that, with appropriate conditions, all the theoretical results obtained in our work will also apply when using the SPCR estimation. We use a regression spline basis with 20 knots, a degree 3 and the order of derivation 2. The choice of these parameters is not crucial in our study, especially in comparison with the choice of the number of principal components. The choice of this optimal tuning parameter is made on a growing sequence of dimension $k_n = 2, \ldots, 22$.

5.2 Criteria

Our objective is to predict the response in the test samples. We use two criteria.

- Criterion 1: the average mean square prediction error

$$\overline{MSPE} = \frac{1}{\boldsymbol{S}} \sum_{j=1}^{\boldsymbol{S}} MSPE(j),$$

  where $MSPE(j) = \frac{1}{n_1} \sum_{\ell=n+1}^{n+n_1} \left(\widehat{\alpha} + \langle \widehat{\theta}, X_\ell^{\star,j} \rangle - \alpha - \langle \theta, X_\ell^{\star,j} \rangle \right)^2$ is the mean square prediction error computed on the j^{th} simulated sample, $j \in \{1, \ldots, \boldsymbol{S}\}$. The criterion $\overline{MSPE}$ tends to zero when the sample size tends to infinity.

- Criterion 2: the average ratio respect to truth, based on a deterministic regression imputation,

$$\overline{RT} = \frac{1}{\boldsymbol{S}} \sum_{j=1}^{\boldsymbol{S}} RT(j),$$

  where $RT(j) = \frac{\sum_{\ell=n+1}^{n+n_1} \left(\widehat{\alpha} + \langle \widehat{\theta}, X_\ell^j \rangle - Y_\ell^j \right)^2}{\sum_{\ell=n+1}^{n+n_1} \left(\epsilon_\ell^j \right)^2}$ is the ratio between the mean square prediction error and the mean square prediction error when the true parameters are known, computed on the j^{th} simulated sample. The criterion $\overline{RT}$ tends to one when the sample size tends to infinity.

5.3 Results

Tables (1) and (2) presents the criteria for the complete dataset (**FULL**) and the imputation methods presented in this paper, with reconstructed curves :

- **DETER_IM** : Deterministic regression imputation, as described in subsection 3.1.
- **RAND_IM** : Random regression imputation, as described in subsection 3.2.

Table 1: Mean and standard deviation errors for the predicted values based on 250 simulation replications with different levels of missing data and a sample size $N = 1400$. Partially observed curves are fully observed on $[3/8, 6/8]$ and the error ε is a Gaussian noise: $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.2$.

Rate of missing	26.903	26.877	44.941	45.218	54.793	55.109
data in Y in %	(1.298)	(1.409)	(1.563)	(1.515)	(1.337)	(1.460)
Rate of missing	30.047	44.952	29.995	45.030	30.086	45.164
data in X in %	(1.112)	(1.230)	(1.238)	(1.280)	(1.216)	(1.317)
(**FULL**) $\overline{MSPE} \times 10^3$	17.602	16.785	18.145	16.960	19.150	18.055
	(16.058)	(15.640)	(15.990)	(13.681)	(16.149)	(15.709)
$\overline{RT} \times 10$	14.421	14.231	14.639	14.144	14.733	14.580
	(4.337)	(3.887)	(4.134)	(3.405)	(4.042)	(4.096)
(**DETER_IM**) $\overline{MSPE} \times 10^3$	30.786	29.748	51.942	48.223	66.907	70.525
	(28.722)	(27.327)	(47.172)	(44.261)	(57.530)	(67.268)
$\overline{RT} \times 10$	17.751	17.540	23.320	21.925	26.695	27.758
	(7.624)	(6.914)	(12.195)	(10.902)	(14.482)	(16.921)
(**RAND_IM**) $\overline{MSPE} \times 10^3$	45.463	45.833	67.350	65.721	85.999	90.581
	(36.723)	(39.229)	(50.395)	(49.144)	(66.256)	(73.653)
$\overline{RT} \times 10$	2.138	2.160	2.716	2.623	3.139	3.286
	(0.959)	(1.018)	(1.304)	(1.219)	(1.664)	(1.878)
(**RAND_NORM_IM**) $\overline{MSPE} \times 10^3$	30.732	29.927	52.298	48.412	67.055	70.693
	(28.284)	(27.798)	(47.330)	(44.427)	(58.103)	(67.278)
$\overline{RT} \times 10$	17.735	17.589	23.411	21.981	26.721	27.799
	(7.505)	(7.024)	(12.237)	(10.972)	(14.605)	(16.941)
(**MUL_IM** (q=5)) $\overline{MSPE} \times 10^3$	34.405	31.449	55.165	52.568	69.329	74.675
	(29.976)	(28.133)	(48.511)	(44.218)	(56.368)	(69.281)
$\overline{RT} \times 10$	18.663	17.978	24.166	23.058	27.310	28.763
	(7.875)	(7.147)	(12.561)	(10.911)	(14.194)	(17.467)
(**MUL_NORM_IM** (q=5)) $\overline{MSPE} \times 10^3$	30.819	29.698	51.988	48.054	66.978	70.689
	(28.606)	(27.233)	(47.249)	(44.095)	(57.747)	(67.771)
$\overline{RT} \times 10$	17.756	17.526	23.332	21.885	26.713	27.797
	(7.603)	(6.895)	(12.206)	(10.829)	(14.550)	(17.032)
(**MUL_IM** (q=10)) $\overline{MSPE} \times 10^3$	30.998	30.255	53.437	49.395	68.639	73.111
	(28.554)	(27.931)	(47.390)	(44.601)	(56.621)	(67.923)
$\overline{RT} \times 10$	17.807	17.667	23.692	22.224	27.125	28.386
	(7.640)	(7.053)	(12.223)	(10.934)	(14.196)	(17.100)
(**MUL_NORM_IM** (q=10)) $\overline{MSPE} \times 10^3$	30.680	29.627	51.890	48.178	66.629	70.699
	(28.664)	(27.221)	(47.210)	(44.347)	(57.206)	(67.557)
$\overline{RT} \times 10$	17.721	17.510	23.304	21.915	26.620	27.801
	(7.601)	(6.891)	(12.200)	(10.926)	(14.392)	(16.982)

Table 2: Mean and standard deviation errors for the predicted values based on 250 simulation replications with different levels of missing data and a sample size $N = 1400$. Partially observed curves are fully observed on $[3/8, 6/8]$ and the error ε is a Gaussian noise: $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.2$.

(**MUL_IM** (q=30)) $\overline{MSPE} \times 10^3$	30.326	29.207	**51.259**	48.253	**66.384**	71.054
	(28.978)	(26.953)	**(47.133)**	(44.117)	**(56.415)**	(67.691)
$\overline{RT} \times 10$	17.627	17.419	**23.149**	21.919	**26.547**	27.884
	(7.713)	(6.837)	**(12.183)**	(10.857)	**(14.189)**	(17.048)
(**MUL_NORM_IM** (q=30)) $\overline{MSPE} \times 10^3$	30.695	29.731	51.951	48.029	66.812	70.553
	(28.701)	(27.490)	(47.299)	(44.102)	(57.459)	(67.418)
$\overline{RT} \times 10$	17.726	17.538	23.324	21.876	26.669	27.765
	(7.619)	(6.954)	(12.226)	(10.867)	(14.467)	(16.957)
(**MUL_IM** (q=100)) $\overline{MSPE} \times 10^3$	**30.114**	**29.130**	51.605	**47.395**	66.611	**70.374**
	(28.662)	**(27.545)**	(47.355)	**(43.987)**	(57.918)	**(67.553)**
$\overline{RT} \times 10$	**17.57**4	**17.392**	23.225	**21.719**	26.620	**27.730**
	(7.614)	**(6.953)**	(12.234)	**(10.835)**	(14.573)	**(16.988)**
(**MUL_NORM_IM**(q=100)) $\overline{MSPE} \times 10^3$	30.700	29.693	51.913	48.110	66.742	70.507
	(28.663)	(27.370)	(47.154)	(44.196)	(57.487)	(67.309)
$\overline{RT} \times 10$	17.727	17.527	23.314	21.897	26.652	27.755
	(7.608)	(6.924)	(12.192)	(10.884)	(14.474)	(16.929)
(**MEAN_IM**) $\overline{MSPE} \times 10^3$	30.913	30.746	54.404	53.923	74.127	74.266
	(24.018)	(23.151)	(37.310)	(37.364)	(44.807)	(45.275)
$\overline{RT} \times 10$	17.755	17.820	23.961	23.374	28.672	28.768
	(6.199)	(6.153)	(9.777)	(9.314)	(11.481)	(11.631)
(**RANDO_IM**) $\overline{MSPE} \times 10^3$	30.870	30.909	54.121	54.127	73.924	73.478
	(24.108)	(23.192)	(37.878)	(37.973)	(45.867)	(44.904)
$\overline{RT} \times 10$	17.740	17.852	23.885	23.433	28.622	28.568
	(6.214)	(6.312)	(9.897)	(9.480)	(11.771)	(11.563)
(**ZERO_IM**) $\overline{MSPE} \times 10^2$	72.025	71.648	194.283	195.935	283.638	287.324
	(8.039)	(7.951)	(14.874)	(14.811)	(15.420)	(17.570)
$\overline{RT} \times 10$	191.134	190.669	501.892	501.050	728.003	736.894
	(24.954)	(28.526)	(55.713)	(58.105)	(71.678)	(70.364)
(**REM_Y**) $\overline{MSPE} \times 10^3$	40.047	37.844	78.278	72.577	94.632	100.989
	(34.437)	(32.908)	(58.280)	(61.814)	(71.559)	(81.000)
$\overline{RT} \times 10$	20.052	19.568	29.985	28.085	33.687	35.381
	(8.923)	(8.399)	(15.204)	(15.412)	(18.126)	(20.334)
(**REM_X,Y**) $\overline{MSPE} \times 10^3$	48.448	60.808	91.500	90.137	117.749	135.675
	(47.901)	(61.016)	(74.047)	(81.080)	(94.352)	(126.053)
$\overline{RT} \times 10$	22.284	25.280	33.257	32.779	39.728	44.123
	(13.086)	(15.150)	(18.983)	(20.142)	(24.442)	(31.855)

- **RAND_NORM_IM** : Parametric approach of random regression imputation, where the error term $\varepsilon^\star$ is drawn from the distribution of the residuals, here assuming the residuals are normally distributed, thus $\varepsilon^\star \sim N(0, \widehat{\sigma}^2_{\varepsilon^\star})$, with $\widehat{\sigma}^2_{\varepsilon^\star}$ being estimated from the residuals of the formerly fitted functional linear model. This parametric method is easy to implement. It seems natural to test the performance of this method on simulations.

- **MUL_IM** : Multiple regression imputation with different values of q ($q = 5, 10, 30, 100$), as described in subsection 3.3.

- **MUL_NORM_IM** : Parametric approach of multiple regression imputation with different values of q ($q = 5, 10, 30, 100$). Here, the error term $\varepsilon^\star$ is drawn as described above, thus $\varepsilon^\star \sim N(0, \widehat{\sigma}^2_{\varepsilon^\star})$.

- **MEAN_IM** : Mean imputation,

- **RANDO_IM** : Random imputation (imputation by a random response drawn in the set of observed values),

- **ZERO_IM** : Zero imputation (imputation by zero).

Moreover, we propose two other cases :

- **REM_Y** : Reconstruct X and remove all the missing values in Y from the sample,

- **REM_X,Y** : Either a partially observed curve or a missing response are removed from the sample.

As it can be expected, the errors increase as the percentage of missing values in X and Y increase. Moreover, when the number of iterations q increases, we recover the $\overline{MSPE}$ and $\overline{RT}$ of the deterministic imputation (**DETER_IM**). Furthermore, when q is large enough ($q = 30$ and $q = 100$), our method (**MUL_IM**) behaves better than the other imputation methods, specially where we delete the missing values (**REM_Y** and **REM_X,Y**). Comparing (**MUL_IM**) and (**MUL_NORM_IM**), we notice that (**MUL_NORM_IM**), behaves better for small values of q while (**MUL_IM**) behaves better for larger values of q.

6 Real dataset study

Our experimental study is based on two steps. In the first treatment step, we do not observe the price-demand functions directly but we have to estimate each price-demand function by a local polynomial smoother estimator. Here, we choose the Gaussian kernel and we consider a cross validation criterion to select the optimal tuning bandwidth parameter from a grid of parameter values in the interval [1070,35000]. In the second step, we reconstructed the missing parts of the differents curves. Now, X_i, $i = 1, \ldots, 241$, is the daily electricity price

curve on day i (function of the residual demand), and Y_i is the value of electricity production (in MWh) on day i. The production data come from https://www.agora-energiewende.de[1]. Only a graphic (with numerical values marked at the observation points) was available on this website to collect a data (neither a table nor an Excel file). It can be possible to use a software to get numerical values from a graphic (see https://automeris.io[2]). However, this software is not completely reliable and some numerical values, being not possible, can be considered as missing data for the response variable. In our case, the percentage of missing data is 13.26%.

We split the initial sample into a learning sample (the index set is denoted I_L) with size 181 and a test sample with size 60 (the index set is denoted I_T). Firstly, we reconstructed the missing parts of the differents curves and, on the learning sample, we imputed the missing values on the response. We tested the residuals normality, the shapiro test gives a p-value equal to 0.905, hence the normality of the residuals cannot be rejected. Then, on the test sample, we computed the prediction values for the response. In order to evaluate the quality of the prediction, we calculated, for $q = 100$, the mean squared prediction error $MSPE = \frac{1}{60}\sum_{i\in I_T}(Y_i - \widehat{Y}_i)^2 = 40.440$ and the mean absolute prediction error $MAPE = \frac{1}{60}\sum_{i\in I_T}|Y_i - \widehat{Y}_i| = 5.349$. Table (3) gives the $MSPE$ and the $MAPE$ for different imputation methods.

Comparing (**MUL_IM**) and (**MUL_NORM_IM**), we notice that (**MUL_NORM_IM**) behaves better for larger values of q, even if the differences are sometimes slight, because the normality of the residuals. Notice finally that, in this situation, the method (**REM_X,Y**) would not be possible since all the curves are partially observed and this would cause removing all individuals in the sample.

Missing values are imputed directly from the regression model, reducing the prediction error with respect to the missing rate but not taking into account the uncertainty of missing values or unseen data. Multiple regression imputation takes this into account by adding a random error term from the regression model residual distribution. This does not reduce the mean square prediction error but when the number of iteration increases, we can recover that of the deterministic regression imputation. Furthermore, multiple imputations are more realistic depending on the quality of the training data set the regression model was trained under.

[1] https://www.agora-energiewende.de/en/service/recent-electricity-data/chart/power_generation/15.03.2012/14.03.2013/
[2] https://automeris.io/WebPlotDigitizer/

Table 3: The mean square prediction error and the mean absolute prediction error with standard deviation errors for deterministic, random and multiple imputation methods.

Imputation methods	$MSPE$	$MAPE$
DETER_IM	40.443 (45.615)	5.354 (3.461)
RAND_IM	40.468 (45.662)	5.356 (3.462)
RAND_NORM_IM	40.533 (46.097)	5.363 (3.463)
MUL_IM ($q = 5$)	40.452 (45.613)	5.355 (3.461)
MUL_NORM_IM ($q = 5$)	40.479 (45.577)	5.357 (3.460)
MUL_IM ($q = 50$)	40.448 (45.624)	5.354 (3.461)
MUL_NORM_IM ($q = 50$)	40.269 (45.474)	5.345 (3.450)
MUL_IM ($q = 100$)	40.440 (45.625)	5.349 (3.461)
MUL_NORM_IM ($q = 100$)	40.211 (45.363)	5.343 (3.443)
REM_Y	40.543 (45.947)	5.354 (3.475)

7 Proof of Theorem 4.2

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

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

hence

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

We obtain from Crambes et al. (2022) the convergence rate for the first term of the decomposition

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

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

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

As a consequence, with the assumptions

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

we get

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

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

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

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

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

References

Buuren, S. V. (2018). *Flexible imputation of missing data (Second edition)*. New York: Chapman and Hall.

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

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

Crambes, C., C. Daayeb, A. Gannoun, and Y. Henchiri (2022). Functional linear model with partially observed covariate and missing values in the response. Preprint at https://hal.archives-ouvertes.fr/hal-03083293v3.

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

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

Febrero-Bande, M., P. Galeano, and W. Gonzalez-Manteiga (2019). 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.

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

Gellar, J. E., E. Colantuoni, D. M. Needham, and C. M. Crainiceanu (2014). Variable-Domain functional regression for modeling ICU Data. *Journal of the American Statistical Association 109*, 1425–1439.

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

Graham, J. W. (2012). *Missing data analysis and design*. New York: Springer Verlag.

Greenland, S. and W. Finkle (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology 142*, 1255–1264.

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

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

Haziza, D. (2009). Imputation and inference in the presence of missing data. In D. Pfeffermann and C. Rao (Eds.), *Handbook of Statistics: Sample Surveys: Design, Methods and Applications*, Volume 29 A, pp. 215–256.

He, Y., G. Zhang, and C. Hsu (2022). *Multiple imputation of missing data in practice: Basic theory and analysis strategies*. New York: John Wiley and Sons.

Hsing, T. and R. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley and Sons: Wiley series in probability and statistics.

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

Kokoszka, P. and M. Reimherr (2018). *Introduction to functional data analysis*. New York: Chapman and Hall.

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

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

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

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

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

Lin, Z. and J.-L. Wang (2020). Mean and covariance estimation for functional snippets. *Journal of the American Statistical Association 117*, 348–360.

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

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

Little, R. J. A. and D. B. Rubin (2020). *Statistical analysis with missing data (Third edition)*. New York: John Wiley and Sons.

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

Park, Y., X. Chen, and D. S. Simpson (2022). Robust inference for partially observed functional response data. *Statistica Sinica 32*, 1–29.

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

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

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

Ramsay, J. O. and B. W. Silverman (2002). *Applied Functional data analysis: Methods and case studies*. New York: Springer Verlag.

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

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.

Srivastava, A. and E. P. Klassen (2016). *Functional and Shape Data Analysis*. New York: Springer Verlag.

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

Wang, L., R. Cao, J. Du, and Z. Zhang (2019). A nonparametric inverse probability weighted estimation for functional data with missing response data at random. *Journal of the Korean Statistical Society 48*, 537–546.

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