mathematics

MDPI

Article

Local Linear Regression for Functional Ergodic Data with Missing at Random Responses

Yassine Baghli [1,†], Oussama Bouanani [2,†] and Salim Bouzebda [3,*,†]

[1] Laboratoire de Mathématiques pour L'Intelligence Artificielle et Sciences du Vivant, Faculty of Exact Sciences and Computer Science, University of Mostaganem, Mostaganem 27000, Algeria
[2] LMSSA (Laboratory of Stochastic Models, Statistics and Applications), Faculty of Exact Sciences and Computer Science, University of Mostaganem, Mostaganem 27000, Algeria
[3] Université de Technologie de Compiègne, LMAC (Laboratory of Applied Mathematics of Compiègne), CS 60 319, 60 203 Compiègne, France
* Correspondence: salim.bouzebda@utc.fr
† These authors contributed equally to this work.

Abstract

In this article, we develop a novel kernel-based estimation framework for functional regression models in the presence of missing responses, with particular emphasis on the Missing At Random (MAR) mechanism. The analysis is carried out in the setting of stationary and ergodic functional data, where we introduce apparently for the first time a local linear estimator of the regression operator. The principal theoretical contributions of the paper may be summarized as follows. First, we establish almost sure uniform rates of convergence for the proposed estimator, thereby quantifying its asymptotic accuracy in a strong sense. Second, we prove its asymptotic normality, which provides the foundation for distributional approximations and subsequent inference. Third, we derive explicit closed-form expressions for the associated asymptotic variance, yielding a precise characterization of the limiting law. These results are obtained under standard structural assumptions on the relevant functional classes and under mild regularity conditions on the underlying model, ensuring broad applicability of the theory. On the methodological side, the asymptotic analysis is exploited to construct pointwise confidence regions for the regression operator, thereby enabling valid statistical inference. Furthermore, a comprehensive set of simulation experiments is conducted, demonstrating that the proposed estimator exhibits superior finite-sample predictive performance when compared to existing procedures, while simultaneously retaining robustness in the presence of missingness governed by MAR mechanisms.

Keywords: functional data analysis; kernel method; local linear estimation method; uniform almost complete convergence; normality; missing responses at random

MSC: 60F05; 60F15; 60F17; 62G07

check for updates

Academic Editor: Heng Lian

Received: 13 September 2025
Revised: 3 October 2025
Accepted: 9 December 2025
Published: 10 December 2025

Citation: Baghli, Y.; Bouanani, O.; Bouzebda, S. Local Linear Regression for Functional Ergodic Data with Missing at Random Responses. *Mathematics* **2025**, *13*, 3941. https://doi.org/10.3390/math13243941

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Over recent years, Functional Data Analysis (FDA) has emerged as a significant research area, focusing on variables that take values in infinite-dimensional spaces, such as Hilbert or Banach spaces. This framework is particularly useful for analyzing curves, surfaces, or other continuously indexed data. For an overview, we mention some pioneering contributions on this subject, such as [1] for linear models concerning random variables

assuming values in a Hilbert space and [2] for scalar-on-function linear models, functional principal component analysis, and parametric discriminant analysis. Ref. [3] primarily delved into nonparametric techniques, particularly kernel-based estimation, for scalar-on-function nonlinear regression models, which were further extended to encompass classification and discrimination analysis. Furthermore, ref. [4] explored the extension of various statistical concepts to the functional data framework, including goodness-of-fit tests, portmanteau tests, and change detection. Ref. [5] conducted a study on the analysis of variance for functional data, while ref. [6] primarily investigated regression analysis for Gaussian processes. Various semi-parametric models have been examined in the literature, such as functional single-index models [7], projection pursuit models [8], partial linear models [9], and functional sliced inverse regression [10]. Further studies on functional data modeling and analysis are documented in sources such as Rice [11–14] and recent references [15]. In this article, our focus lies on the examination of local linear regression, particularly when the regressor is a functional type and its response variable displays specific observations.

Local linear fitting offers several advantages over standard kernel methods, notably its ability to adapt to the design and to reduce boundary bias. It has been shown to be an effective nonparametric smoothing technique. These advantages are particularly significant in multidimensional settings, where managing boundary effects becomes more challenging. The method proposed by [16] involves the use of a local higher-order polynomial fitting to estimate the function m and its partial derivatives. The author demonstrated the convergence of mean square and joint asymptotic normality, as also mentioned in [17]. This research has significant implications for the estimation of widely used ARCH time series. Furthermore, ref. [18] established the strong consistency of the regression estimator and its partial derivatives up to a fixed total order p, along with obtaining sharp rates of almost sure convergence, uniform over compact sets. In [19], the local polynomial estimation of regression functions and their derivatives was studied, with the joint asymptotic normality of relevant estimates established for both strongly mixing and ρ-mixing processes. Moreover, ref. [20] considered the nonparametric estimation of a multivariate regression function and its derivatives for a regression model with long-range dependent errors, adopting a local linear fitting approach and establishing joint asymptotic distributions for the estimators of the regression function and its derivatives.

It is important to mention that all the contributions mentioned above are applicable only when the samples are being observed completely. However, functional statistics with missing data is a frequent problem in data analysis. Note that when data are missing completely at random (MAR), it means that the probability of missing data is not associated with any observed or unobserved variable. There are several approaches to address this issue, such as complete case analysis (ignoring observations with missing data) or imputation (replacing missing values with estimated values). This field recognizes a diversity of applications in reality, such as market surveys, medical investigations, and reliability tests; there may be instances where some pairs of observations are incomplete. Performing statistical analysis with missing data is a complex and demanding operation due to the lack of information about the missing data mechanism in most circumstances, as noted by [21]. The assumption of random missing responses is a fundamental and commonly utilized method. Several instances of this scenario and its statistical implications for a regression model can be found in the statistical literature when the explanatory variables are of finite dimension. For more information, refer to [22–28] and the references provided in those works. There are only a limited number of studies that have reported on statistical models with missing data when the explanatory variables are infinite or contain a functional feature. For instance, ref. [29] was the first to suggest a method for

estimating the average value of a single outcome using a functional sample where the independent variables are fully observed and the dependent variables are randomly absent. In their study, the authors of [30] examined the asymptotic characteristics of the regression operator estimate for functional stationary ergodic data that include missing responses occurring randomly. Ref. [31] examined the semi-functional partially linear regression model and the functional linear model with randomly missing scalar responses in the case of an independent and identically distributed (i.i.d.) sample. Ref. [32] explored the statistical procedures for partially observed functional data. The main objective of the present investigation is to establish a comprehensive framework for linear fitting when there are missing responses that occur randomly, extending our previous study in [33]. The central aim of this paper is to establish comprehensive theoretical validation for the kernel linear estimator in the MAR setting. However, it becomes evident that the solution to this problem transcends a mere amalgamation of ideas from disparate domains. Instead, intricate mathematical derivations are imperative to contend with the challenges posed by estimators in our specific context. This necessitates the adept application of techniques rooted in large sample theory in the functional data framework.

The article is structured as follows. Section 2 provides the necessary notation and definitions. The main results are outlined in Section 3. The asymptotic distribution is discussed in Section 3.4. The establishment of a uniform almost complete convergence is detailed in Section 3.1. In Section 4, we illustrate our methodology through a simulation study designed to compare the classical nonparametric functional regression model with our proposed estimator. Section 5 summarizes the findings and highlights the remaining open issues. All proofs are deferred to Appendix A, with a focus on the most crucial arguments due to the lengthiness of the proofs. In addition, relevant technical results are provided in Appendix B.

2. Modeling with Responses: MAR and Estimator

In the context of missing data with a Missing at Random (MAR) mechanism for the response variable, an available incomplete sample of size n, drawn from the joint distribution of (X, Y, ϱ), is given by

$$\{(X_i, Y_i, \varrho_i), 1 \leq i \leq n\},$$

where X_i is observed completely, $\varrho_i = 1$ if Y_i is observed, and $\varrho_i = 0$ otherwise. Meanwhile, the Bernoulli random variable ϱ satisfies

$$\mathbb{P}(\varrho = 1 \mid X = x, Y = y) = \mathbb{P}(\varrho = 1 \mid X = x) = p(x),$$

where $p(x)$ is a function, called the conditional probability of observing the response given the predictor, which is often unknown. This mechanism shows that ϱ and Y are conditionally independent given X.

We consider a sequence $\{(X_i, Y_i, \varrho_i) : i \geq 1\}$ of strictly stationary process ergodic random copies of the random vector [rv] (X, Y), where X takes its values in some abstract space $\mathfrak{F}$ and Y in $\mathbb{R}$. (In the case of Hilbert space-valued elements, strict stationarity is not necessarily required; second-order stationarity suffices. A Hilbert space-valued sequence $\{X_t\}_{t\in\mathbb{Z}}$ is second-order (also called weakly stationary) if $\mathbb{E}\|X_t\|^2 < \infty$, $\mathbb{E}X_t = \mu$, and $\mathbb{E}[(X_s - \mu) \otimes (X_t - \mu)] = \mathbb{E}[(X_{s-t} - \mu) \otimes (X_0 - \mu)]$ for all $s, t \in \mathbb{Z}$. We say that $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary if the joint distribution of $\{X_{t_1}, \ldots, X_{t_n}\}$ is the same as the joint distribution of $\{X_{t_1+h}, \ldots, X_{t_n+h}\}$ for all $t_1, \ldots, t_n \in \mathbb{Z}$, $n \geq 1$, and $h \geq 1$.) Suppose that $\mathfrak{F}$ is endowed with a semi-metric $d(\cdot, \cdot)$ (a semi-metric (sometimes called pseudo-metric) $d(\cdot, \cdot)$ is a metric that allows $d(x_1, x_2) = 0$ for some $x_1 \neq x_2$l) defining a topology to measure

the proximity between two elements of $\mathfrak{F}$, which is assumed to be independent of the construction of X to avoid measurability problems. We focus on the conditional expectation $\psi(Y)$ given $X = x$l,

$$r(x) = r_\psi(x) = \mathbb{E}(\psi(Y) \mid X = x), \tag{1}$$

whenever this regression function is meaningful. Here and elsewhere, $\psi(\cdot)$ denotes a specified measurable function, which is assumed to be bounded at each compact subinterval of $\mathbb{R}$. Note that we can write

$$\psi(Y) = r(X) + \epsilon, \qquad \text{with} \qquad \mathbb{E}(\epsilon \mid X) = 0, \qquad \mathbb{E}(\epsilon^2 \mid X) = \sigma^2(X).$$

In this work, we study the problem of nonparametric estimation of the regressor function $r(\cdot)$ using the local linear fitting such that the regression function lies in an infinite-dimensional space. From now on, for ease of notation, we set $\psi(y) = y$. Recall that in functional statistics there are several approaches to developing the concept of local linear methods. For example, in a Hilbert space like $L^2([0,1])$ [34], the linear approximation of $r(\cdot)$ for any z in the neighborhood of x is given by

$$r(z) = r(x) + \langle b, z - x \rangle + o(||z - x||),$$

where $b = b(x) \in L^2([0,1])$ and $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2([0,1])$. However, in the Banach space [35] the linear approximation of $r(\cdot)$ for any z in the neighborhood of x is given by

$$r(z) = r(x) + c_1 \beta(z, x) + o(\beta(z, x)). \tag{2}$$

Based on the local linear approximation stated in (2), the local linear estimator adapted to MAR (Missing at Random) responses for the regression function $r(x)$ denoted by $\widehat{r}(x)$ is defined as the first component of the vector estimate $\widehat{c}(x) = (\widehat{c}_0(x), \widehat{c}_1(x))^\top$. Within this framework, $\widehat{c}_1(x)$ serves as an estimator for the first derivative of the regression function $r(\cdot)$. These quantities are obtained by solving the following local linear minimization problem:

$$\min_{(c_0,c_1)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - c_0 - c_1 \beta(X_i, x))^2 \varrho_i K(h_K^{-1} \delta(x, X_i)), \tag{3}$$

where the locating functions $\beta(\cdot,\cdot)$ and $\delta(\cdot,\cdot)$ are defined on $(\mathfrak{F} \times \mathfrak{F})$ and map into $\mathbb{R}$, with $|\delta(\cdot,\cdot)| = d(\cdot,\cdot)$, where $d(\cdot,\cdot)$ is a distance metric. The function β refers to the local behavior of our model, while $K(\cdot)$ is a kernel function, and $h_K = h_{K,n}$ is the bandwidth or smoothing parameter of the kernel $K(\cdot)$, controlling the size of the local neighborhood and the degree of smoothing. Let $\mathbf{Q}_\beta^\top$ denote the matrix defined by

$$\begin{aligned} \mathbf{Q}_\beta^\top &:= \begin{pmatrix} 1 & \cdots & 1 \\ \beta(X_1, x) & \cdots & \beta(X_n, x) \end{pmatrix}, \\ \mathbf{Y} &= {}^\top(Y_1, \ldots, Y_n), \\ \mathbf{W} &:= \operatorname{diag}\Big(\varrho_i K\Big(h_K^{-1} \delta(X_i, x)\Big)\Big). \end{aligned}$$

where $^\top$ designates the transpose operator. Accordingly, the solution of (3) may be represented in the form

$$\widehat{c}(x) = \left(\mathbf{Q}_\beta^\top \mathbf{W} \mathbf{Q}_\beta\right)^{-1} \mathbf{Q}_\beta^\top \mathbf{W}\, \mathbf{Y}. \tag{4}$$

Now, we define the following auxiliary quantities:

$$\begin{aligned} u_{n,j} &= \frac{1}{n\mathbb{E}(K(h_K^{-1}(\delta(X_1,x))))}\sum_{i=1}^{n}\left(\frac{\beta(X_i,x)}{h_K}\right)^j \varrho_i K(h_K^{-1}(\delta(X_i,x))), \\ v_{n,j} &= \frac{1}{n\mathbb{E}(K(h_K^{-1}(\delta(X_1,x))))}\sum_{i=1}^{n}\left(\frac{\beta(X_i,x)}{h_K}\right)^j K(h_K^{-1}(\delta(X_i,x))\varrho_i Y_i, \\ \mathbf{A}_n &= \begin{pmatrix} u_{n,0} & \cdots & u_{n,1} \\ u_{n,1} & \cdots & u_{n,2} \end{pmatrix} \text{ and } \quad L_n =^{\top} (v_{n,0},\ldots,v_{n,p}). \end{aligned}$$

With this notation in place, the estimator can be written as

$$\widehat{c}(x) = \mathrm{diag}(1, h_K^{-1})\ \mathbf{A}_n^{-1}\ L_n.$$

By some algebraic manipulation, the quantity $\widehat{c}_0(x)$ is explicitly defined by

$$\widehat{c}_0(x) = \frac{\sum\limits_{i,i\neq j,j=1}^{n} W_{ij}Y_j\varrho_j}{\sum\limits_{i,i\neq j,j=1}^{n} W_{ij}\varrho_j} = \frac{\sum\limits_{j=1}^{n}\Delta_j K_j Y_j \varrho_j}{\sum\limits_{i,i\neq j,j=1}^{n} W_{ij}\varrho_j}, \tag{5}$$

where

$$W_{ij} = \beta_i(\beta_i - \beta_j)\varrho_i K_i K_j$$

and

$$\Delta_j = K_j^{-1}\left(\sum_{i=1}^{n} W_{ij}\right) = \sum_{i=1}^{n}\beta_i^2 K_i\varrho_i - \left(\sum_{i=1}^{n}\beta_i K_i\varrho_i\right)\beta_j,$$

with

$$\beta_i = \beta(X_i,x), \text{ and } K_i = K\Big(h_K^{-1}(\delta(x,X_i))\Big).$$

Moreover, the estimator $\widehat{c}_0(x)$ can be equivalently expressed as follows:

$$\widehat{c}_0(x) = \frac{\hat{r}_{n,2}(x)}{\hat{r}_{n,1}(x)},$$

where

$$\hat{r}_{n,t} = \frac{1}{n\mathbb{E}(\Delta_1 K_1)}\sum_{j=1}^{n}\Delta_j K_j Y_j^{t-1}\varrho_j, \quad \text{for } \ t\in\{1,2\}.$$

The performance of the estimate depends crucially on the two locating functions, $\beta(\cdot,\cdot)$ and $\delta(\cdot,\cdot)$, which are defined on $(\mathfrak{F}\times\mathfrak{F})$ and map into $\mathbb{R}$. These functions satisfy $|\delta(\cdot,\cdot)| = d(\cdot,\cdot)$, where $K(\cdot)$ is a kernel function, and $h_K = h_{K,n}$ is the smoothing parameter of the kernel $K(\cdot)$. To clarify, specific forms for $\delta(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ can be given. For example, if the functional data are 'smooth' curves, one might use the following family of locating functions:

$$\mathrm{loc}_a^{(q)}(x_1,x_2) = \int\theta(t)\Big(x_1^{(q)}(t) - x_2^{(q)}(t)\Big)\mathrm{d}t = \langle\theta,\, x_1^{(q)} - x_2^{(q)}\rangle_{\mathcal{F}},$$

where θ is a function that can be adapted to the data, x^q denotes the qth derivative of the function x, $\mathcal{F}$ is a Hilbert space, and $\langle\cdot,\cdot\rangle$ denotes the inner product. Choosing $\beta(\cdot,\cdot)$ from such a family is motivated by its connection to the following minimization problem:

$$\min_{c(x)\in\mathbb{R}^2}\sum_{i=1}^{n}\left(Y_i - c_0 - c_1\Big\langle\theta, X_i^{(q)} - x^{(q)}\Big\rangle_{\mathcal{F}}^{l}\right)^2 K\Big(h^{-1}|\delta(X_i,x)|\Big),$$

which can be viewed as a type of *local linear* regression approach when considering a functional explanatory variable. Metrics, or more generally semi-metrics, based on derivatives could also be suitable for locating one curve relative to another. For example, one might define

$$\mathrm{loc}_b^{(q)}(x_1, x_2) = \left(\int \left(x_1^{(q)}(t) - x_2^{(q)}(t) \right)^2 \mathrm{d}t \right)^{1/2},$$

which is a semi-metric. This second family of location functions is particularly well suited for $\delta(\cdot, \cdot)$, as it measures the proximity between two elements of $\mathcal{F}$.

3. Main Results for the Regression Operator

3.1. Uniform Almost Complete Convergence

In this section, we aim to establish the rate of uniform almost complete convergence of the estimator over a subset of $\mathcal{F}$. The following definition is needed.

Definition 1. *Let $\epsilon > 0$ be given number. A finite set of points $x_1, x_2, \ldots, x_N$ in $\mathcal{F}$ is called an $\epsilon-$ network for S if $S \subset \cup_{k=1}^{N} B(x_k, \epsilon)$. The quantity $\Psi_s(\epsilon) = \log(N_\epsilon(S))$, where $N_s(\epsilon)$ is the minimal number of open balls in $\mathcal{F}$ of radius ϵ that is necessary to cover S, is called the Kolmogorov $\epsilon-$ entropy of the set S.*

The notion of entropy, originally introduced by Kolmogorov in the mid-1950s (see [36]), provides a quantitative measure of the complexity of a set. Intuitively, a high entropy indicates that a substantial amount of information is required to accurately describe an element of the set within a given tolerance ε. For a more comprehensive account of this concept and its developments, we refer the reader to [37]. In this framework, the choice of topology—and, in particular, the specification of the semi-metric—plays a decisive role in shaping the uniform convergence properties of asymptotic results considered over a subset $S \subset \mathcal{F}$. A judiciously chosen semi-metric not only enhances the concentration of the probability measure associated with the functional variable X, but also contributes to reducing the ε-entropy of the underlying subset S.

This interplay between concentration phenomena and entropy reduction has been highlighted by [3], who investigated the concentration of probability measures for functional variables by evaluating small ball probabilities under various standard scenarios (see also [38]). For readers seeking deeper insight into these concepts, namely entropy, small ball probabilities, and the application of Kolmogorov's ε-entropy in dimensionality reduction, we recommend the references [39,40].

In the present work, we establish the uniform almost complete convergence of our estimator over a subset $S_{\mathcal{F}}$ satisfying the covering property

$$S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n} B(x_k, \iota_n), \tag{6}$$

where $x_k \in \mathcal{F}$, and ι_n (respectively, d_n) denotes a sequence of positive real numbers.

Throughout the paper, for a fixed point $x \in \mathfrak{F}$ (respectively, $y \in \mathbb{R}$), we denote by N_x (respectively, N_y) a fixed neighborhood of x (respectively, y). Moreover, for $x \in \mathfrak{F}$ and $h > 0$, the closed ball in $\mathfrak{F}$ centered at x with radius h will be denoted by

$$B(x, h) := \left\{ x' \in \mathfrak{F} : |\delta(x, x')| \leq h \right\}.$$

3.2. Assumptions and Notations

To present our results, we introduce the following notation. For each integer i, let $\mathcal{F}_i$ be the σ-field generated by the pairs $\big((X_1, Y_1), \ldots, (X_i, Y_i)\big)$, and let $\mathcal{G}_i$ be the σ-field generated by $\big((X_1, Y_1), \ldots, (X_i, Y_i), X_{i+1}\big)$. For a given point $x \in \mathcal{E}$, we define the non-negative random variable $D_i := d\big(x, X_i\big)$, Working in the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let

$$F_x(u) = \mathbb{P}\big(D_i \leq u\big) = \mathbb{P}\big(X_i \in B(x,u)\big),$$

be the distribution function of D_i, and

$$F_x^{\mathcal{F}_{i-1}}(u) = \mathbb{P}\big(D_i \leq u \mid \mathcal{F}_{i-1}\big) = \mathbb{P}\big(X_i \in B(x,u) \mid \mathcal{F}_{i-1}\big),$$

is the conditional distribution function given $\mathcal{F}_{i-1}$.

Finally, by $o_{\text{a.s.}}(u)$ we refer to a real-valued random function ℓ such that $\ell(u)/u \to 0$ almost surely as $u \to 0$. Likewise, we write $O_{\text{a.s.}}(u)$ for a real-valued random function ℓ such that $\ell(u)/u$ is almost surely bounded as $u \to 0$.

We define

$$\phi_x^{\mathcal{F}_{i-1}}(l_1, l_2) = \mathbb{P}(l_2 \leq \delta(X_i, x) \leq l_1 \mid \mathcal{F}_{i-1}),$$

where l_1 and l_2 are two real numbers. It is easy to see that for any $u > 0$,

$$\phi_x^{\mathcal{F}_{i-1}}(-u, u) = F_x^{\mathcal{F}_{i-1}}(u).$$

When the function $|\delta(\cdot,\cdot)|$ satisfies the conditions of a metric or semi-metric, the quantity $F_x(u)$ represents the probability that the observation X falls within a ball of radius u centered at x . When u approaches zero, the term "small ball probabilities" is frequently employed in probability theory, which has been extensively studied (see [41] for a comprehensive summary in the context of Gaussian processes). However, in the context of ergodic functional data and finite-dimensional spaces ($\mathcal{F} = \mathbb{R}^d$ equipped with Euclidean norm, the expression $F_x^{\mathcal{F}_{i-1}}(u)$ takes the classical form:

$$\mathbb{P}(X_i \in B(x,u) \mid \mathcal{F}_{i-1}) = c_0 f_{i,d}(x) u^d + o(u^d) \quad \text{a.s. as } u \to 0,$$

where $f_{i,d}(x)$ is the conditional density at x (for more details, see [42]). The function $\phi(\cdot)$, an extension of the concept of small ball probability, serves a similar purpose in the functional setting as the density function in the finite-dimensional case.

Throughout this paper, we denote by C and C' generic positive constants that may take different values in different contexts.

To establish the asymptotic behavior of the local linear estimator, we require the following assumptions.

(H.1) There exists a differentiable function $\varphi(\cdot)$ such that, for all $x \in S_{\mathcal{F}}$, the following conditions hold:
$0 < C\varphi(h) \leq \mathbb{P}\big(X \in B(x, h_K)\big) \leq C'\varphi(h) < \infty$, and there exists $\eta_0 > 0$ such that for all $\eta < \eta_0$, $\varphi'(\eta) < C$, where φ' denotes the first derivative of φ.

(H.2) The stochastic process $(X_i, Y_i)_{i \in \mathbb{N}^*}$ satisfies:

(i) For each $i = 1, \ldots, n$, there exists a deterministic function $\varphi_i(\cdot)$ such that

$$C'\varphi_i(u) < F_x^{\mathcal{F}_{i-1}}(u) \leq C\varphi_i(u), \quad \text{for all } u > 0.$$

(ii) For all $u > 0$,

$$\frac{1}{n\varphi(u)} \sum_{i=1}^{n} \varphi_i(u) \to 1 \quad \text{a.co.}$$

(H.3) The conditional expectation of Y_i given $\mathcal{G}_{i-1}$ depends solely on X_i; that is, for any $i \geq 1$, almost surely,

$$\mathbb{E}(Y_i \mid \mathcal{G}_{i-1}) = \mathbb{E}(Y_i \mid X_i).$$

and there exists $C > 0$ such that, for some γ, almost surely,

$$\mathbb{E}(|Y_i|^{\gamma} \mid X_i = x) < m_{\gamma}(x) < C < \infty,$$

with $m_{\gamma}(x)$ continuous on $S_{\mathcal{F}}$

(H.4) The locating operators $\beta(\cdot,\cdot)$ and $\delta(\cdot,\cdot)$ satisfy:

(i) For all $z \in \mathfrak{F}$,

$$C_1|\delta(x,z)| \leq |\beta(x,z)| \leq C_2|\delta(x,z)|.$$

(ii) For all $(x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$,

$$|\beta(x,x_1) - \beta(x,x_2)| \leq Cd(x_1,x_2).$$

(iii)

$$\sup_{v \in B(x,r)} \mid \beta(v,x) - \delta(x,v) \mid = o(r).$$

(H.5) (i) The regression operator $r(\cdot)$ satisfies that there exist $C > 0$ and $s > 0$ such that for all $x \in \mathcal{F}$ and $z \in B(x, h_K)$,

$$|r(x) - r(z)| \leq Cd^s(x,z).$$

(ii) The function $p(\cdot)$ is continuous on $S_{\mathcal{F}}$ and there exist positive constants C and C' such that

$$0 < C' < \inf_{x \in S_{\mathcal{F}}} p(x) < \sup_{x \in S_{\mathcal{F}}} p(x) < C < \infty.$$

(H.6) The kernel $K(\cdot)$ is a bounded, positive Lipschitz function on its support $[-1,1]$. Moreover, The first derivative $K'(t)$ satisfies

$$K(1) > 0 \quad \text{and} \quad K'(t) < 0, \ \forall\, t \in [-1,1].$$

(H.7) Assuming $\iota_n = O(\log n/n)$, the sequence d_n satisfies:

$$\frac{(\log n)^2}{n\phi(h_K)} < \log d_n < \frac{n\phi(h_K)}{\log n},$$

and

$$\sum_{n=1}^{\infty} d_n^{1-\rho} < \infty \quad \text{for some } \rho > 1.$$

(H.8) (i) There exists $n_0 > 0$ such that for $n > n_0$,

$$-\frac{1}{\phi(h_K)} \int_{-1}^{1} \phi(zh_K, h_K) \frac{d}{dz}\Big(z^2 K(z)\Big) dz > C_3 > 0,$$

and

$$h_K \int_{B(x,h_K)} \beta(u,x) d\mathbb{P}_X^{\mathcal{F}_{i-1}}(u) = o\left(\int_{B(x,h_K)} \beta^2(u,x) d\mathbb{P}_X^{\mathcal{F}_{i-1}}(u)\right),$$

where $d\mathbb{P}_X$ is conditional cumulative distribution function given $\mathbb{P}_{i-1}$ of X.

(ii) The bandwidths h_K satisfy

$$\lim_{n\to\infty} h_K = 0, \quad \lim_{n\to\infty} \frac{\log n}{n\phi(h_K)} = 0, \quad \lim_{n\to\infty} \frac{\mathfrak{D}_{l,j}(h_K)}{n^2\varphi^2(h_K)} = 0,$$

where

$$\mathfrak{D}_{l,j}(h_K) = \sum_{\substack{j=1 \\ l\neq j}}^{n} \sum_{l=1}^{n} \varphi_l(h_K)\varphi_j(h_K)$$

and there exists $\beta > 0$ such that

$$\lim_{n\to\infty} nh_K^{\beta} = 0.$$

Before stating our result, we introduce an additional notation. Specifically, we define certain quantities that will be used in the computation of $E(K_j^r \mid \mathcal{F}_{j-1})$ for $r = 1, 2$. Let

$$M_r = K^r(1) - \int_{-1}^{1} (K^r(u))' \chi_x(u) du, \tag{7}$$

For all $a > 0$, and $b = 2, 4$, let

$$Z(a,b) = K^a(1) - \int_{-1}^{1} (u^b K^a(u))' \chi_x(u) du, \tag{8}$$

where $\chi_x(\cdot)$ is a a function such that

$$\lim_{h_K\to 0} \frac{\phi_x^{\mathcal{F}_{j-1}}(-h_K, th_K)}{F_x^{\mathcal{F}_{j-1}}(h_K)} = \chi_x(t) \text{ for any } t \in [-1, 1].$$

3.3. Comments on the Conditions

This part presents the theoretical analysis that supports assumptions (H.1)–(H.8), which are necessary to prove the consistency and asymptotic normality of the proposed local linear estimator for functional ergodic data with missing at random (MAR) responses.

To take into account the uniformity aspect in the asymptotic analysis, we introduce a strictly positive and increasing function $\varphi(\cdot)$, which does not depend on x and plays a similar role to $F_x(\cdot)$. Assumption (H.1) is based on small ball probability techniques. This condition is not considered restrictive as it is commonly used in the framework of Functional Data Analysis (FDA). This control is crucial for establishing uniform convergence rates of local estimators (for more details, see [3]).

The assumption (H.2) shows the links between the ergodic behavior of the functional data and the small ball probability techniques. Unlike the classical strong-mixing framework, the dependence structure considered here is more general and accommodates a broader class of functional processes. Such a structure has notably been explored by [42] in the specific context of functional autoregressive models. Assumption (H.3) is a Markov-type condition and characterizes the conditional moments. The assumption (H.4-i) is required to control the shape of the local functional mapping $\beta(\cdot,\cdot)$. Assumption (H.4-iii) is not restrictive and holds, for example, when $\delta(\cdot,\cdot) = \beta(\cdot,\cdot)$ or, more generally, when the locating operators δ and β satisfy the condition

$$\lim_{\delta(z,x)\to 0} \left| \frac{\beta(z,x)}{\delta(z,x)} - 1 \right| = 0.$$

Indeed, for all $z \in B(x, h_K)$, we have

$$\left|\frac{\beta(z,x) - \delta(z,x)}{h_K}\right| \leq \left|\frac{\beta(z,x)}{\delta(z,x)} - 1\right| \to 0 \quad \text{as } \delta(z,x) \to 0.$$

This ensures that both operators are asymptotically equivalent in the neighborhood of x. Assumption (H.5-i) is a mild regularity condition that allows us to characterize the functional space involved in our nonparametric model. It is typically used to evaluate the bias term in the asymptotic analysis of the estimator. Moreover, the second part of this assumption stipulates that the conditional probability of observing the response variable, given the infinite-dimensional covariate, is uniformly bounded below by a strictly positive constant. This condition is essential to avoid the presence of regions in the functional covariate space where the response is never observed, which would hinder the validity of statistical inference and impair the estimation of the regression function. Imposing this assumption ensures the availability of sufficient information across the covariate domain and guarantees the identifiability of the model.

Assumption (H.6) is a standard requirement in nonparametric regression. It ensures that the kernel function K is bounded, Lipschitz continuous, and has compact support. The assumptions (H.7) and (H.8) constitute technical requirements, similar to those used by [43].

The following theorem provides the first result on the almost sure uniform convergence, including the convergence rate.

Theorem 1. *Assume that assumptions (H.1)–(H.8) are satisfied. Then, as $n \to \infty$,*

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{c}_0(x) - c(x)| = O(h_K^b) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{n\phi(h_K)}}\right),$$

where $O_{a.co.}$ denotes almost complete convergence. (Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of real random variables. We say (u_n) converges almost completely (a.co.) to zero if, for all $\epsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}(|u_n| > \epsilon) < \infty$. The rate of convergence is of order v_n (with $v_n \to 0$), written $u_n = O_{a.co.}(v_n)$, if there exists $\epsilon > 0$ such that $\sum_{n=1}^{\infty} \mathbb{P}(|u_n| > \epsilon v_n) < \infty$. This implies almost sure convergence and convergence in probability.)

Before proving the theorem, we define the conditional bias of $\widehat{c}_0(x)$ as

$$\overline{B}_n(x) = \frac{\overline{r}_{n,2}(x)}{\overline{r}_{n,1}(x)} - c_0(x), \tag{9}$$

where

$$\overline{r}_{n,t}(x) = \frac{1}{n\mathbb{E}[\Delta_1 K_1]} \sum_{j=1}^{n} \mathbb{E}\Big(\Delta_j K_j Y_j^{t-1} \varrho_j \mid \mathcal{F}_{j-1}\Big), \quad t \in \{1,2\}.$$

where $\mathbb{E}(X \mid \mathcal{F})$ denotes the conditional expectation of the random variable X given the σ-field $\mathcal{F}$. We decompose the error as

$$\widehat{c}_0(x) - c_0(x) = \overline{B}_n(x) + \frac{[\hat{r}_{n,2}(x) - \overline{r}_{n,2}(x)] - [(\hat{r}_{n,1}(x) - \overline{r}_{n,1}(x))(c_0(x) + \overline{B}_n(x)]}{\hat{r}_{n,1}(x)}. \tag{10}$$

The proof relies on the following lemmas.

Lemma 1. *Under the assumptions (H.1)–(H.6) and (H.8), the conditional bias satisfies*

$$\sup_{x \in S_{\mathcal{F}}} |\overline{B}_n(x)| = O(h_K^b).$$

Lemma 2. *Under the assumptions (H.1)–(H.8), for $t = 1, 2$, we have*

$$\sup_{x \in S_{\mathcal{F}}} |\hat{r}_{n,t}(x) - \bar{r}_{n,t}(x)| = O_{a.co.}\left(\sqrt{\frac{\log d_n}{n\varphi(h_K)}}\right).$$

Lemma 3. *Under the assumptions (H.1)–(H.8), we have*

$$\exists\, C > 0 \text{ such that } \sum_{n=1}^{\infty} \mathbb{P}(|\bar{r}_{n,1}(x)| \leq C) < \infty.$$

The proof of Theorem 1 is deferred to Appendix A.

Now, we establish the second result, which describes the limiting distribution of the proposed estimator.

In order to establish the asymptotic convergence of $\widehat{c}_0(x)$, we set the following hypotheses that will be needed to announce our result.

(B.1) The kernel K satisfies assumption (H.4-i), and its first derivative K' satisfies

$$K^2(1) - \int_{-1}^{1} (K^2(u))' \chi_x(u) du > 0.$$

(B.2) (i) The conditional variance of Y_i given $\mathcal{G}_{i-1}$ also depends only on X_i; that is, for any $i \geq 1$, almost surely,

$$\mathbb{E}\Big[(Y_i - r(X_i))^2 \mid \mathcal{G}_{i-1}\Big] = \mathbb{E}\Big[(Y_i - r(X_i))^2 \mid X_i\Big] = \sigma^2(X_i).$$

(ii) $r(\cdot)$ and $\sigma(\cdot)$ are continuous in a neighborhood of x; that is, they belong to the set

$$\left\{ r : \mathcal{F} \to \mathbb{R}, \lim_{|\delta(z,x)| \to 0} r(z) = r(x) \right\},$$

3.4. Asymptotic Normality

Theorem 2. *Consider assumptions (H.1)–(H.8), (B.1) and (B.2) as $n \to \infty$ and that*

$$n\varphi(h) \to \infty \quad as \quad n \to \infty.$$

Then, for any $x \in \mathcal{F}$, we have:

(i) Asymptotic normality of the centered estimator:

$$\sqrt{n\varphi(h)}(\widehat{r}_n(x) - U_n(x)) \xrightarrow{\mathcal{D}} N(0, V(x)),$$

where

$$U_n(x) = \overline{B}_n(x) + c_0(x) \quad and \quad V(x) = \frac{p(x)M_2}{M_1^2}\,\sigma^2(x).$$

(ii) In addition, if the bandwidth h_K satisfies

$$(n\varphi(h_k))^{1/2} h_k^{\beta} \to 0 \quad as \quad n \to \infty,$$

where β is specified in assumption (H.8), then

$$\sqrt{n\varphi(h)}(\widehat{r}_n(x) - c_0(x)) \xrightarrow{\mathcal{D}} N(0, V(x)).$$

The relationship (10) plays a crucial role in establishing the asymptotic normality of the regression function estimator $\hat{c_0}(x)$. Furthermore, under Assumption (H.8-ii) and Lemma 1 we ensure the asymptotic negligibility of the term $\sqrt{n\varphi(h)}B_n(x,y)$. Moreover, according to Lemma 2, we have $\hat{r}_{n,1}(x) - \overline{r}_{n,1}(x) = o_P(1)$ when $n \to \infty$. Consequently, under the condition that $\hat{r}_{n1}(x)$, converges in probability to 1 as $n \to \infty$, we obtain the following decomposition:

$$\sqrt{n\varphi(h)}\big(\widehat{c}_0(x) - c_0(x) - \overline{B}_n(x)\big) = \frac{Q_n(x) + R_n(x)}{\hat{r}_{n1}(x)}(1 + o_p(1)), \tag{11}$$

where

$$Q_n(x) = \sqrt{n\varphi(h)}(\hat{r}_{n,2}(x) - \overline{r}_{n,2}(x) + c_0(x)(\hat{r}_{n,1}(x) - \overline{r}_{n,1}(x))).$$

and

$$R_n(x) = \sqrt{n\varphi(h)}\overline{B}_n(x)(\hat{r}_{n,1}(x) - \overline{r}_{n,1}(x)).$$

The proof of Theorem 2 follows from the following lemmas (see Appendix A for proofs).

Lemma 4. *Assuming that conditions (H.1), (H.2), (H.3), (H.4), (H.5-ii), (H.6), and (H.8) are satisfied, the following assertion holds:*

$$\hat{r}_{n,1}(x) - p(x) = o_P(1), \quad as\ n \to \infty$$

Lemma 5. *Under the set of assumptions stipulated in Theorem 2, we have the following result:*

$$\sqrt{n\varphi(h)}Q_n(x) \xrightarrow{\mathcal{D}} N(0, V(x)).$$

3.5. Application to the Confidence Regions

The asymptotic variance presented in Theorem 2 depends on several unknown components, including the functional term $p(\cdot)$ and the parameters M_j for $j = 1, 2$. As these elements are typically unobservable, consistent estimation is essential for the practical application of theoretical results, such as the construction of asymptotic confidence intervals for $c_0(x)$. For this aim, we propose the following estimation strategy. Specifically, we adopt a plug-in approach to define an estimator of $\dfrac{V(x)}{n\varphi(h_K)}$, which is given by

$$\frac{p(x)M_2}{M_1^2 n\varphi(h_K)}\sigma^2(x). \tag{12}$$

The quantity (12) can be estimated by the empirical version:

$$\frac{\widehat{\sigma^2}(x)\,\widehat{p}(x)\,\sum_{i=1}^n K_i^2(x)}{\left(\sum_{i=1}^n K_i(x)\right)^2},$$

where

$$\widehat{p}(x) = \frac{\sum_{i=1}^n \delta_i K_i(x)}{\sum_{i=1}^n K_i(x)},$$

and the conditional variance $\sigma^2(x)$ can be defined as

$$\widehat{\sigma^2}(x) = \frac{\sum_{i=1}^n \delta_i Y_i^2 K_i(x)}{\sum_{i=1}^n \delta_i K_i(x)} - (\widehat{c}_0(x))^2.$$

After choosing the kernel function k that satisfies assumption (H.6), while selecting the bandwidth h_K using a cross-validation procedure and discussing the choice of the locating functions $\beta(\cdot,\cdot)$ and $\delta(\cdot,\cdot)$ in Section 2, the confidence intervals for $\widehat{c}_0(x)$ at the asymptotic level $1-\xi$ can be constructed as follows:

$$\left[\widehat{c}_0(x) - u_{1-\xi/2}\,\widehat{\sigma}(x),\ \widehat{c}_0(x) + u_{1-\xi/2}\,\widehat{\sigma}(x)\right], \tag{13}$$

where $u_{1-\xi/2}$ denotes the $1-\xi/2$ quantile of the standard normal distribution.

Remark 1. *For the reader's convenience, we introduce some details defining the ergodic property of processes and its link with the mixing one. Let $\{X_n, n \in \mathbb{Z}\}$ be a stationary sequence. Consider the backward field $\mathscr{A}_n = \sigma(X_k : k \le n)$ and the forward field $\mathscr{B}_m = \sigma(X_k : k \ge m)$. The sequence is strongly mixing if*

$$\sup_{A\in\mathscr{A}_0, B\in\mathscr{B}_n} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0, \ \text{as}\ n\to\infty.$$

The sequence is ergodic if, for any two measurable sets A, B,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\mathbb{P}\left(A\cap\tau^{-k}B\right) - \mathbb{P}(A)\mathbb{P}(B)\right| = 0,$$

where τ is the time evolution or shift transformation. The naming of strong mixing in the above definition is more stringent than what is ordinarily referred to (when using the vocabulary of measure-preserving dynamical systems) as strong mixing, namely that $\lim_{n\to\infty}\mathbb{P}(A\cap\tau^{-n}B) = \mathbb{P}(A)\mathbb{P}(B)$ for any two measurable sets A, B, see, for instance, [44] and more recent references [45,46]. Hence, strong mixing implies ergodicity, whereas the inverse is not always true (see, e.g., Remark 2.6 in page 50 in connection with Proposition 2.8 in page 51 in [47]). An example of an ergodic and non-mixing process was considered in Sect. 5.3 of [48]. Indeed, assume that the process $\{(T_i,\lambda_i) : i\in\mathbb{Z}\}$ is strictly stationary with $T_i \mid \mathcal{T}_{i-1} \sim$ Poisson(λ_i); let $\mathcal{T}_i$ be the σ-field generated by $(T_i,\lambda_i,T_{i-1},\lambda_{i-1},\ldots)$. We assume that $\lambda_i = \kappa(\lambda_{i-1},T_{i-1})$, where $\kappa : [0,\infty)\times\mathbb{N}\to(0,\infty)$. However, this process is not mixing in general; see Remark 3 of [49] for a counterexample. We refer to [48] for further details and motivations for the use of the ergodicity assumption. One of their arguments is that for certain classes of processes, it can be much easier to prove ergodicity rather than the mixing assumption.

Remark 2. *To illustrate that assumptions (H.3) and (B.2) are as in [50], we provide several examples that demonstrate how these conditions are satisfied in various contexts.*

(i) Long-memory discrete-time processes: Let $(\epsilon_t)_{t\in\mathbb{Z}}$ denote a white noise process with variance σ^2, and let I and B represent the identity operator and the backshift operator, respectively. Ref. [51] proved (see Theorem 1, p. 55) that the k-factor Gegenbauer process

$$\prod_{i\le i\le k}\left(I - 2v_iB + B^2\right)^{d_i} X_t = \epsilon_t,$$

where $0 < d_i < 1/2$ if $|v_i| < 1$, or $0 < d_i < 1/4$ if $|v_i| = 1$, for $i = 1,\ldots,k$, is a long-memory, stationary, causal, and invertible process. This process has a moving average representation, expressed as $X_t = \sum_{j\ge0}\psi_j(d,v)\epsilon_{t-j}$, where

$$\sum_{j=0}^{\infty}\psi_j^2(d,v) < \infty,$$

ensuring that the process is well-behaved in the asymptotic sense.

On the other hand, ref. [52] demonstrated that if $(\epsilon_t)_{t\in\mathbb{Z}}$ is a Gaussian process, the above process is not strongly mixing. Nevertheless, the moving average representation confirms that the process is stationary, Gaussian, and ergodic. This example highlights the subtlety of mixing conditions and the importance of the moving average representation in understanding the long-term behavior of the process.

(ii) *The stationary solution of the linear Markov $AR(1)$ process: Consider the process $X_i = \frac{1}{2}X_{i-1} + \epsilon_i$, where (ϵ_i) are independent symmetric Bernoulli random variables taking values -1 and 1. This process is not α-mixing, as shown by [53], because of the inherent dependence structure between consecutive values of X_i. However, despite the lack of strong mixing, the process (X_i) is Markovian, stationary, and ergodic. This example demonstrates how a Markov process, even when it does not exhibit strong mixing, can still exhibit ergodicity, which is crucial for many statistical analyses in time series and functional data analysis.*

(iii) *A stationary process with an $AR(1)$ representation: Let (u_i) be an i.i.d. sequence uniformly distributed on $\{1,\ldots,9\}$, and define the process*

$$X_t := \sum_{i=0}^{\infty} 10^{-i-1} u_{t-i},$$

where $u_t, u_{t-1}, \ldots$ represent the decimal expansion of X_t. This process is stationary and can be expressed in the form of an $AR(1)$ process:

$$X_t = \frac{1}{10}X_{t-1} + \frac{1}{10}u_t = \frac{1}{10}X_{t-1} + \frac{1}{2} + \epsilon_t,$$

where $\epsilon_t = \frac{1}{10}u_t - \frac{1}{2}$ is a strong white noise. While this process is not α-mixing (see [54], Example A.3, p. 349), it is ergodic. This example illustrates a situation where a process that is not strongly mixing can still exhibit the necessary statistical properties of ergodicity, making it suitable for applications in nonparametric FDA.

Remark 3. *The paper already includes several well-known examples from the literature of processes that are ergodic yet not mixing. Broadening this to include more general linear Markov models such as $ARMA(p,q)$ or $ARIMA(p,d,q)$ with $q \neq 0$ would certainly be of considerable interest. However, pursuing this extension would necessitate a meticulous and comprehensive study to rigorously delineate the circumstances under which such processes remain ergodic while failing to satisfy mixing conditions.*

4. Simulation Study

This section presents an empirical evaluation of the proposed regression operator for prediction. We consider a setting with stationary and ergodic functional data, where observations are subject to a Missing at Random (MAR) mechanism. The performance and accuracy of our proposed estimator are assessed using the Mean Squared Error (MSE) criterion. To contextualize its efficacy, we compare its performance against the standard local constant estimator, defined below:

$$\widehat{r}(x) = \frac{\sum_{j=1}^{n} K_j Y_j \varrho_j}{\sum_{j=1}^{n} K_j \varrho_j}. \tag{14}$$

To this end, we generate autoregressive AR(1) processes according to

$$b_i = \tfrac{1}{3}b_{i-1} + Z_i, \quad i = 1,\ldots,200,$$

where the innovations Z_i are independent and identically distributed as $N(0,1)$ and independent of the process $(b_i)_{1\le i\le 200}$. The initial state is drawn from $b_0 \sim N(0,1)$. Subsequently, we simulate 200 independent variables $w_i \sim U[0,1]$, independent of the Z_i. For each pair (w_i, b_i), we define the functional covariates $X_i = (X_i(t_j))_{1\le j\le 200}$ by

$$X_i(t) = w_i t^2 + \sin(\pi b_i t), \quad t \in [0,1], \quad i = 1, \ldots, 200.$$

The functional variables $X_i(t)$ are displayed in Figure 1.

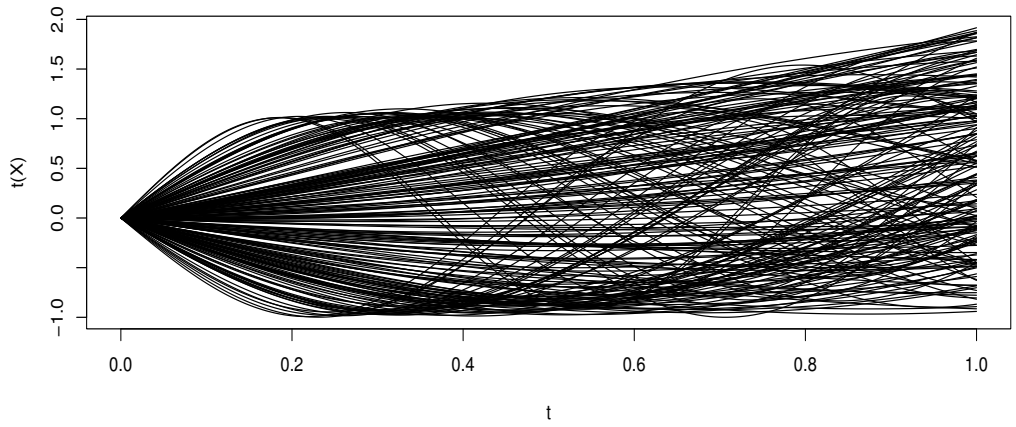

Figure 1. Sample trajectories $X_i(t)$, for $i = 1, \ldots, 200$ and $t \in [0,1]$.

The response variables are generated from the nonparametric functional regression model

$$Y_i = r(X_i) + \varepsilon_i,$$

where the regression operator is specified as

$$r(X_i) = \left(\int_0^1 X_i'(t)\, dt \right)^2,$$

with $X_i'(t)$ denoting the first derivative of $X_i(t)$, and where the errors satisfy $\varepsilon_i \sim N(0, 0.075)$ independently. Following [29], the MAR mechanism is modeled through the conditional probability of observation:

$$\mathbb{P}(\delta = 1 \mid X = x) = \text{expit}\left(\frac{2\alpha}{\pi} \int_0^1 x^2(t)\, dt \right),$$

where the logistic (expit) function is given by

$$\text{expit}(u) = \frac{e^u}{1 + e^u},$$

and the scalar parameter α controls the rate of missingness. The dependence between the functional covariate X and the missingness indicator δ thus varies with α. For each simulated dataset, the empirical proportion of missing values is estimated by

$$\bar{\delta} = 1 - \frac{1}{170} \sum_{i=1}^{170} \delta_i.$$

In the simulation, and under assumption (H.6), we employ the Epanechnikov kernel:

$$K(u) = \frac{3}{4}(1 - u^2)\, \mathbb{1}_{[-1,1[}(u).$$

As previously discussed, the location functions $\beta(\cdot;\cdot)$ and $\delta(\cdot;\cdot)$ are chosen to reflect both the data structure and the smoothness of the functional trajectories. Specifically, we set

$$\begin{aligned}\beta(x;x_0) &= \int_0^1 \theta(t)\big(x^{(1)}(t) - x_0^{(1)}(t)\big)dt,\\ \delta(x;x_0) &= \left(\int_0^1 \big(x^{(1)}(t) - x_0^{(1)}(t)\big)^2 dt\right)^{1/2},\end{aligned}$$

where θ is selected among the eigenfunctions of the empirical covariance operator

$$\frac{1}{n}\sum_{i=1}^{n}\big(X_i^{(1)}(t) - \bar{X}^{(1)}(t)\big)^T\big(X_i^{(1)}(t) - \bar{X}^{(1)}(t)\big),$$

corresponding to the leading eigenvalues, with $\bar{X}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)$ denoting the empirical mean function, see Figure 2.

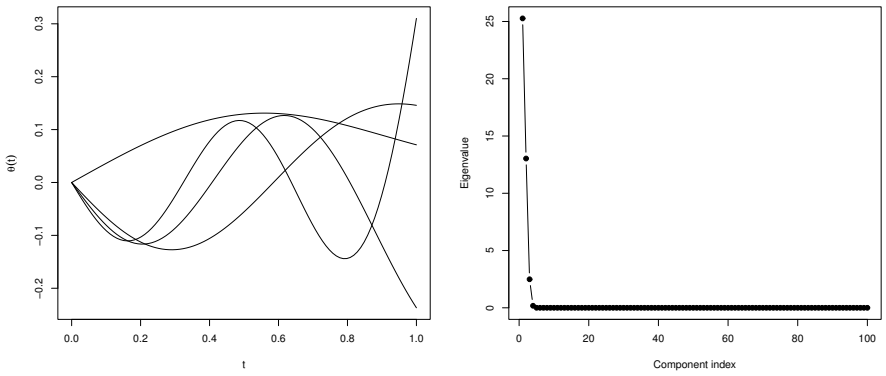

Figure 2. (**Left**): Eigenfunctions $\theta_i(t_j)$, with $t_j \in [0,1]$, for $i = 1,\ldots,4$. (**Right**): Scree plot of the corresponding empirical eigenvalues.

The choice of the bandwidth parameter h_K is of central importance in smoothing-based estimation. In line with the classical kernel regression framework (see [3]), the optimal bandwidth $h_{\text{opt}} := h_{n,K}$ is determined via cross-validation:

$$h_{\text{opt}} = \arg\min_h \text{CV}(h), \quad \text{CV}(h) = \sum_{i=1}^{n}\big(Y_i - \hat{c}_0^{(-i)}(X_i)\big)^2, \tag{15}$$

where $\hat{c}_0^{(-i)}$ denotes the leave-one-out estimator of the regression operator. To evaluate the influence of the parameter α on predictive performance, we consider samples of size $n = 200$, partitioned into a training set of 170 observations and a test set of 30 observations. This splitting procedure is repeated 40 times for each value of α. At each replication, the mean squared error (MSE) is computed for both estimators. The predictive accuracy is further assessed by plotting true versus predicted values and by computing the empirical MSE, defined as

$$\text{MSE}_{\text{LLE}} = \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{c}_0(X_i) - Y_i\big)^2, \quad \text{MSE}_{\text{LCE}} = \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{r}(X_i) - Y_i\big)^2.$$

Representative results are presented in Figures 3–5.

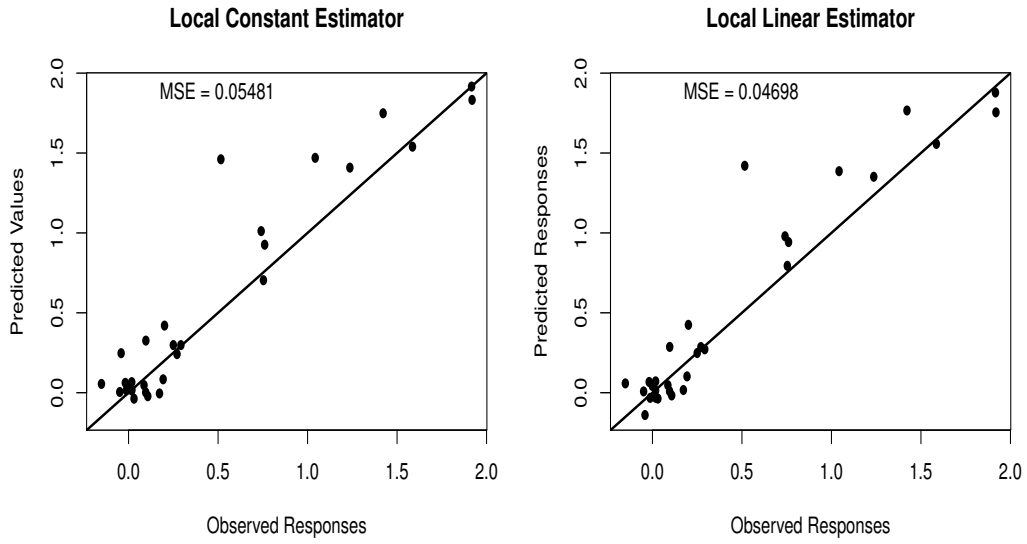

Figure 3. MSE of both estimators under MAR with $\alpha = 1$.

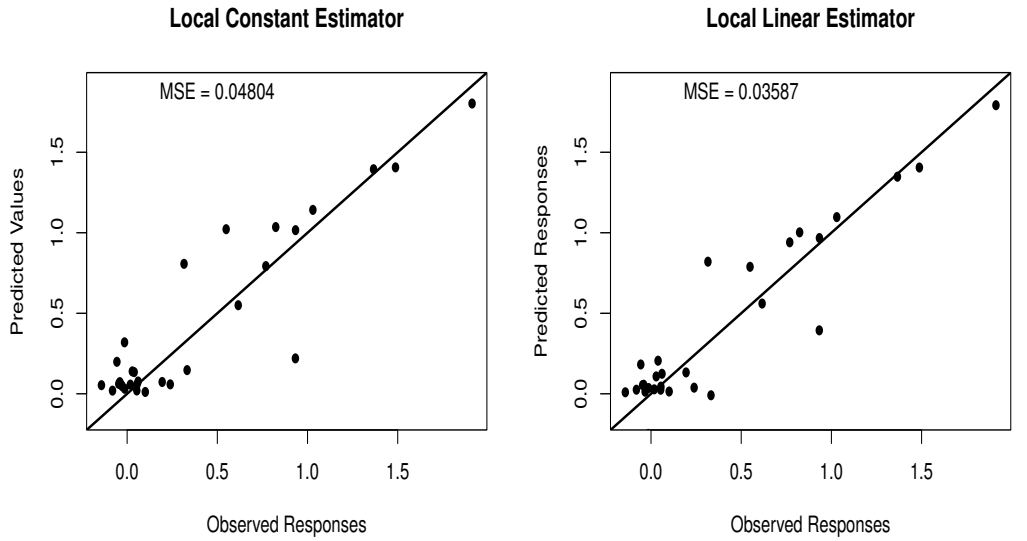

Figure 4. MSE of both estimators under MAR with $\alpha = 1.5$.

From the results reported in Figures 3–5, it follows that the local linear estimator (LLE) uniformly outperforms the local constant estimator (LCE) in terms of estimation accuracy. The performance of the LCE is markedly sensitive to the proportion of missing observations, thereby rendering it less reliable in the presence of incomplete data. In contrast, the LLE achieves a more favorable bias–variance trade-off (MSE), particularly under weak MAR mechanisms (i.e., for large values of α), where the attenuation of bias dominates the effect of the missingness. Consequently, the proposed procedure exhibits enhanced flexibility and improved fidelity to the underlying regression function. These findings are formally summarized in Table 1, which demonstrates a monotone increase in predictive accuracy as the fraction of missing data decreases. Furthermore, the LLE consistently exhibits superior finite-sample behavior relative to the LCE, with the performance differential being most striking for smaller values of α, corresponding to higher levels of missingness.

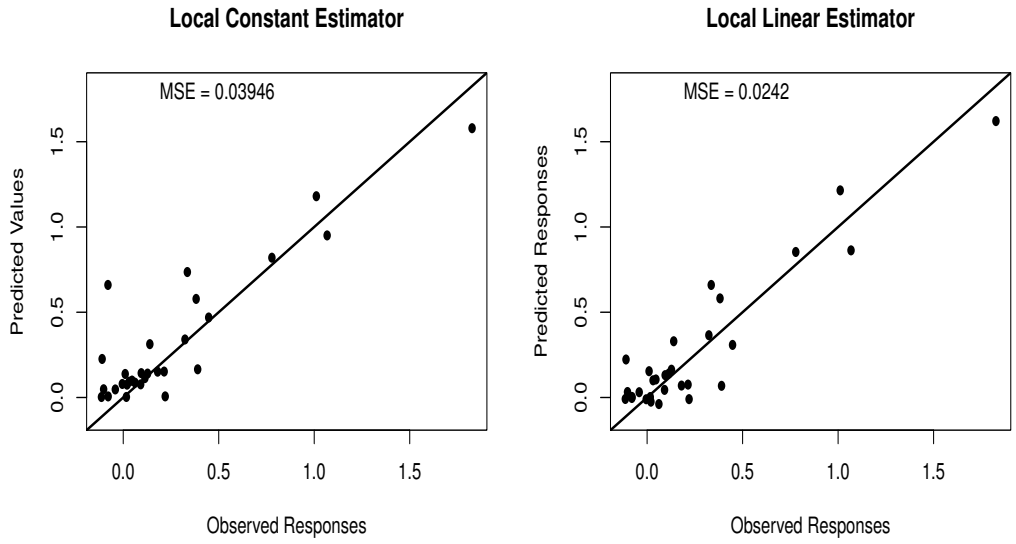

Figure 5. MSE of both estimators under MAR with $\alpha = 2$.

Table 1. Effect of α on missing data proportion and prediction error (MSE) for the Local Constant Estimator (LCE) and the Local Linear Estimator (LLE).

Figure	α	Missing Proportion	MSE (LCE)	MSE (LLE)
Figure 3	1	0.29	0.05481	0.04696
Figure 4	1.5	0.23	0.04804	0.03587
Figure 5	2	0.18	0.03946	0.02420

To enhance the reliability and robustness of the proposed Locally Linear Estimator (LLE), we fixed the sample size at $n = 100$ and examined several levels of missingness, expressed as different percentages of missing data. For each proportion α, we performed 100 independent replications of the experiment. The performance of the proposed methodology was then systematically assessed across a range of complex regression models, using two widely adopted error metrics: the Mean Squared Error (MSE) and the Mean Absolute Percentage Error (MAPE).

The regression models considered below were carefully chosen to reflect distinct levels of structural complexity and nonlinearity in the regression relationship. Specifically, we investigated the following designs:

M1.

$$Y_i = \frac{2}{3}\log\left(2 + 4\left|\int_0^1 X_i'(t)\,dt\right|\right) + \varepsilon_i,$$

where the signal is expressed through the integrated derivative of the covariate trajectory.

M2.

$$Y_i = \exp\left(-0.5\left|\int_0^1 X_i'(t)\,dt\right|\right) + \log\left(1 + \left|\int_0^1 X_i'(t)\,dt\right|\right) + 0.3\left(\int_0^1 X_i'(t)\,dt\right)^3 + \varepsilon_i,$$

with independent noise terms $\varepsilon_i \sim \mathcal{N}(0, 0.075)$.

The predictive accuracy of the estimators was further quantified through the Mean Absolute Percentage Error (MAPE), defined, respectively, for the LLE and the Local Constant Estimator (LCE) as

$$\mathrm{MAPE}_{\mathrm{LLE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{c}_0(X_i) - Y_i}{Y_i}\right|, \qquad \mathrm{MAPE}_{\mathrm{LCE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{r}(X_i) - Y_i}{Y_i}\right|.$$

The choice of both MSE and MAPE as performance measures is motivated by their complementary interpretive value. While the MSE penalizes large deviations more heavily and is therefore sensitive to outliers, the MAPE provides a scale-free measure that facilitates comparison across models with different magnitudes of response. Taken together, these criteria allow for a more comprehensive evaluation of the finite-sample behavior of the estimators.

In the following, we present the simulation results presenting the impact of the missing data mechanism on the performance of the proposed estimator in comparison with the local constant alternative. The overall performance, summarized by the mean values of the MSE and MAPE across 100 replications (denoted MMSE and MMAPE, respectively), is provided in Table 2.

Table 2. Mean Squared Error (MMSE) and Mean Absolute Percentage Error (MMAPE) of the two estimators over 100 replications.

	α	Missing Prop.	LCE		LLE	
Scenario	**Value**		**MMSE**	**MMAPE(%)**	**MMSE**	**MMAPE(%)**
M1	0.05	49%	0.03844	17.46	0.02961	15.98
	0.50	39%	0.03200	16.27	0.02765	15.48
	1.00	29%	0.02732	15.13	0.02453	14.47
	1.50	23%	0.02466	14.36	0.02219	13.81
	2.00	18%	0.02246	13.42	0.02114	13.06
	3.00	13%	0.02246	13.42	0.02114	13.06
M2	0.05	49%	0.14720	12.80	0.10688	11.30
	0.50	39%	0.11018	11.50	0.08653	10.90
	1.00	29%	0.09328	10.69	0.07662	10.10
	1.50	23%	0.08192	10.40	0.06882	10.00
	2.00	18%	0.08341	10.07	0.06750	9.80
	3.00	13%	0.08190	10.03	0.06728	9.60

Figures 6–9 display boxplot representations of the sampling distributions of the predictive performance measures associated with the estimator $\widehat{c}_0(x)$, computed over 100 independent replications. The comparison is carried out with respect to the response variable generated under models M1 and M2, and the evaluation is performed using two conventional accuracy criteria:

- The sampling distributions of the Mean Squared Error (MSE) and the Mean Absolute Percentage Error (MAPE) for $\widehat{c}_0(x)$, when estimated via the Local Linear Estimator (LLE), are reported in the panels on the left.
- The corresponding sampling distributions of the MSE and MAPE for $\widehat{r}(x)$, obtained through the Local Constant Estimator (LCE), are displayed in the panels on the right.

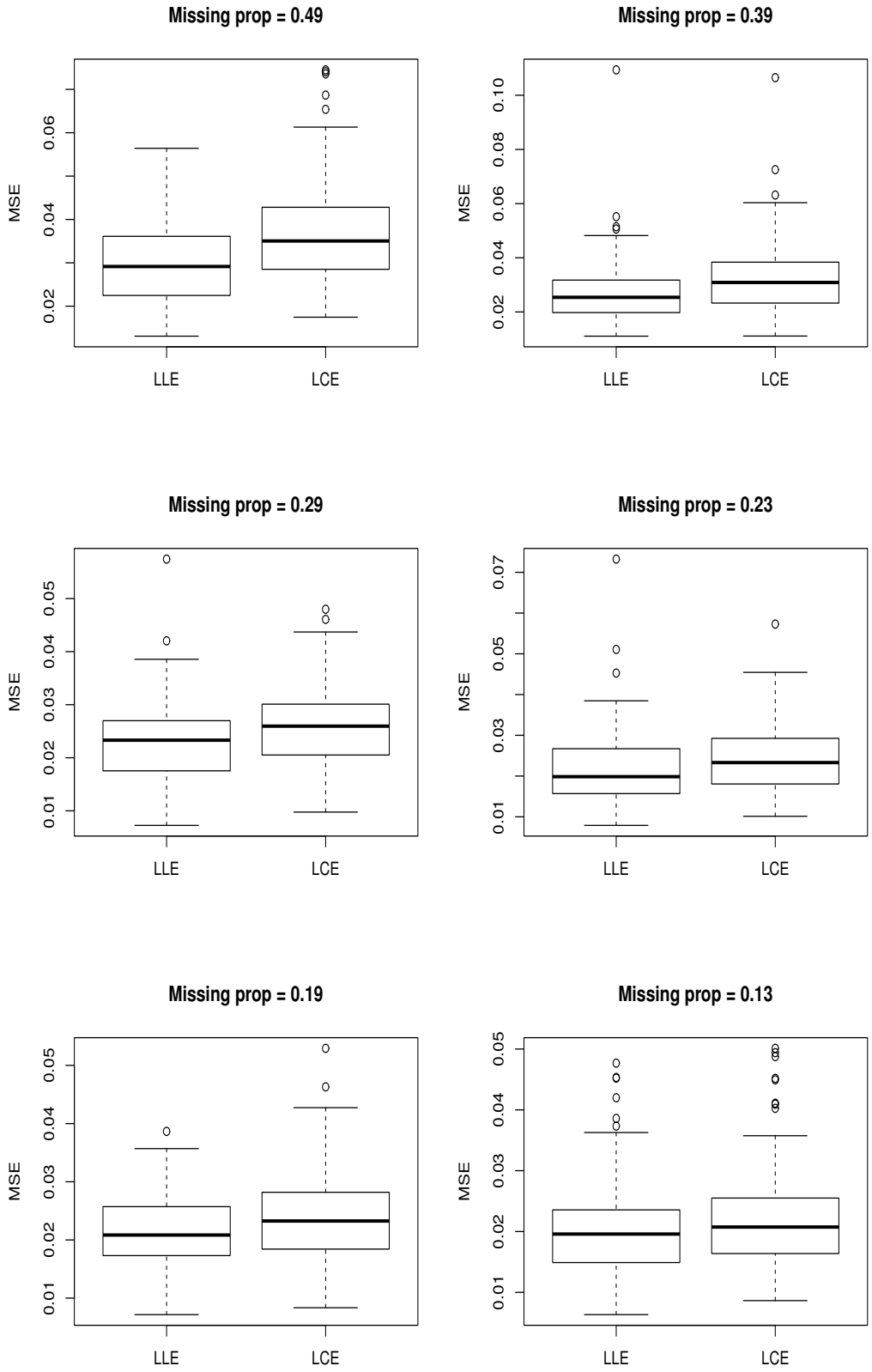

Figure 6. Boxplots of the Mean Squared Error (MSE) Distributions for Both Estimators under the MAR Mechanism, Based on 100 Replications of Model M1.

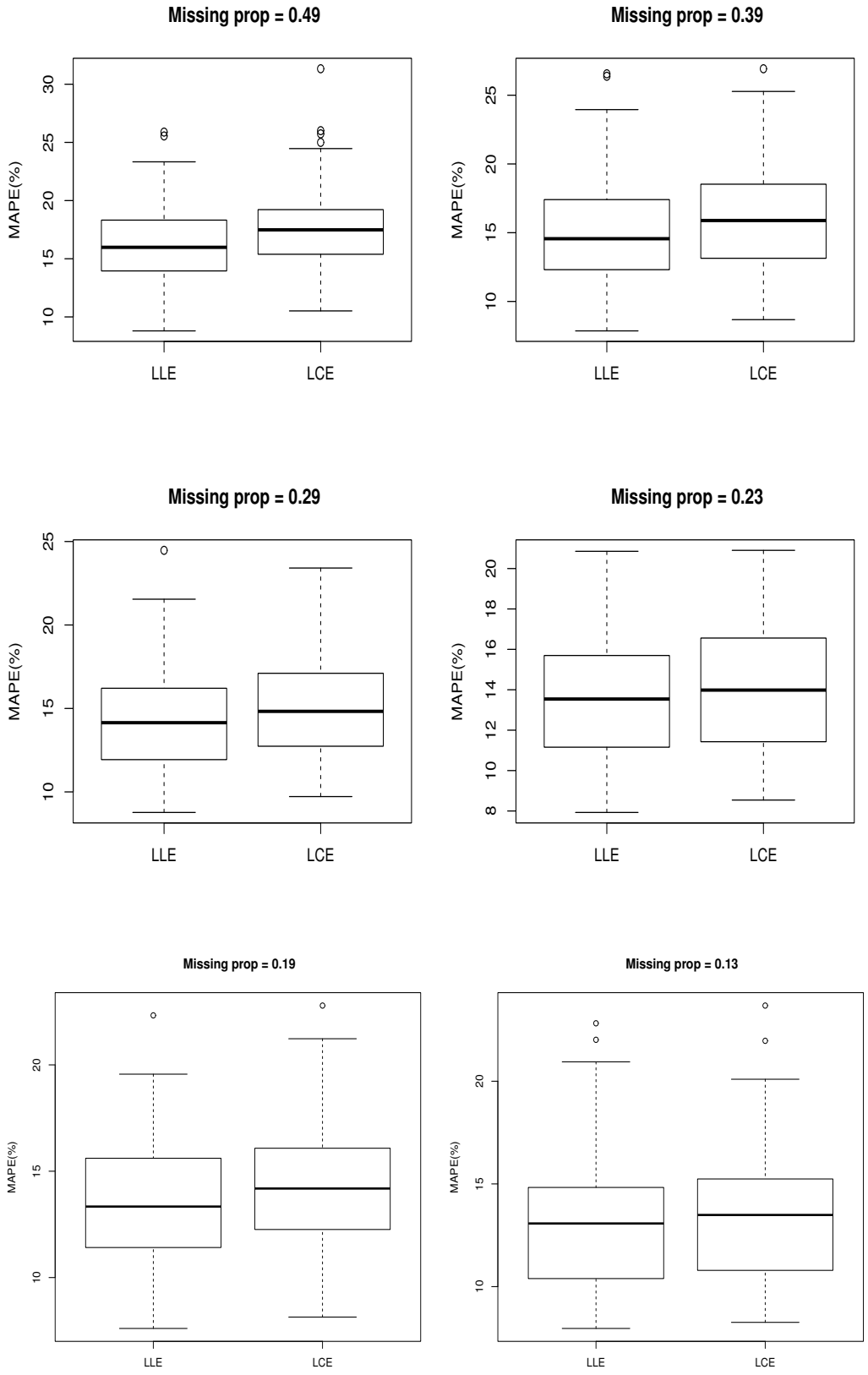

Figure 7. Boxplots of the empirical distribution of the Mean Absolute Percentage Error (MAPE) for both estimators under the Missing At Random (MAR) mechanism, based on 100 Monte Carlo replications using Model M1.

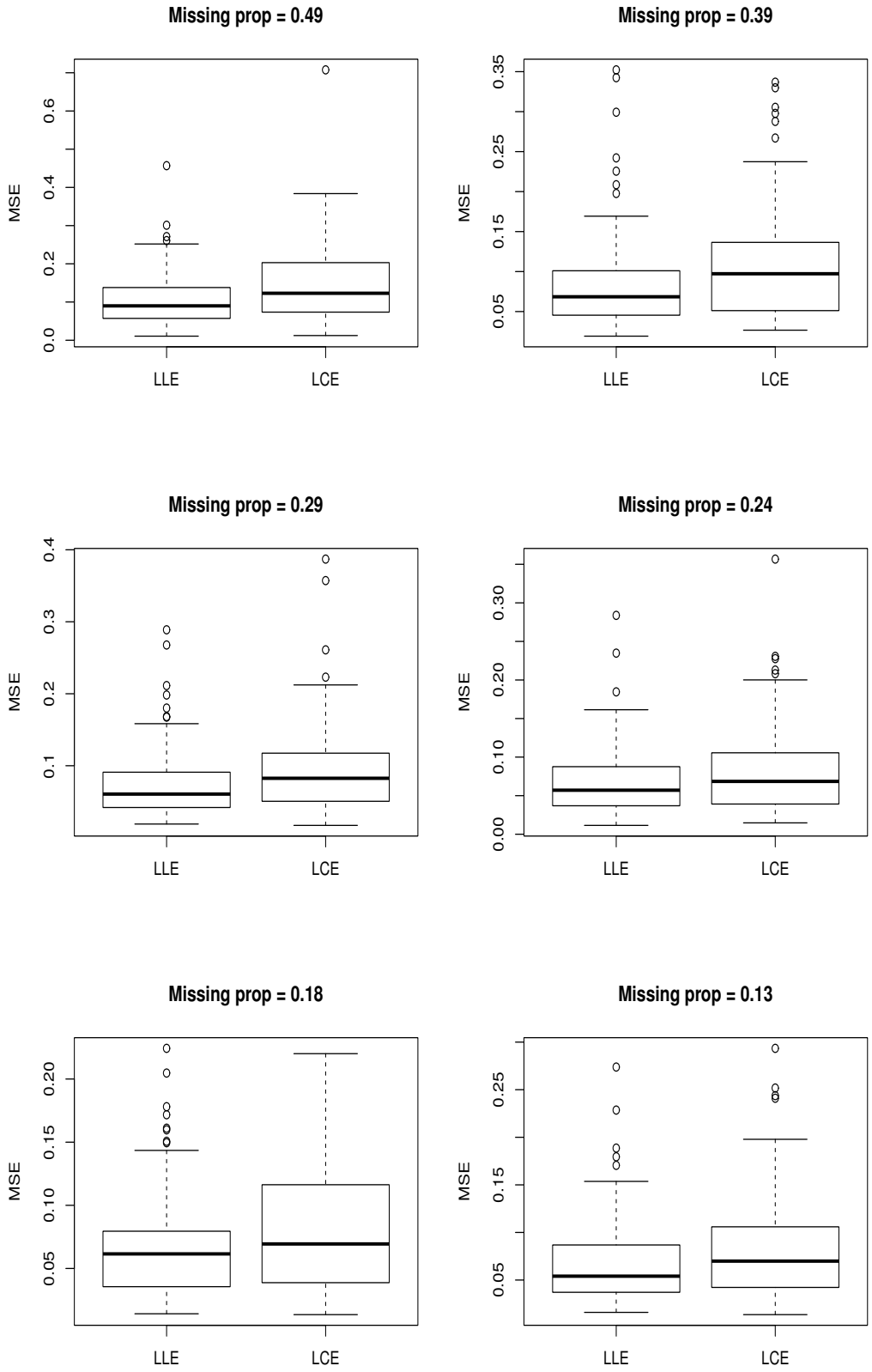

Figure 8. Boxplots of the Empirical Distribution of the MSE for Both Estimators under the MAR Mechanism, Based on 100 Monte Carlo Replications Using Model M2.

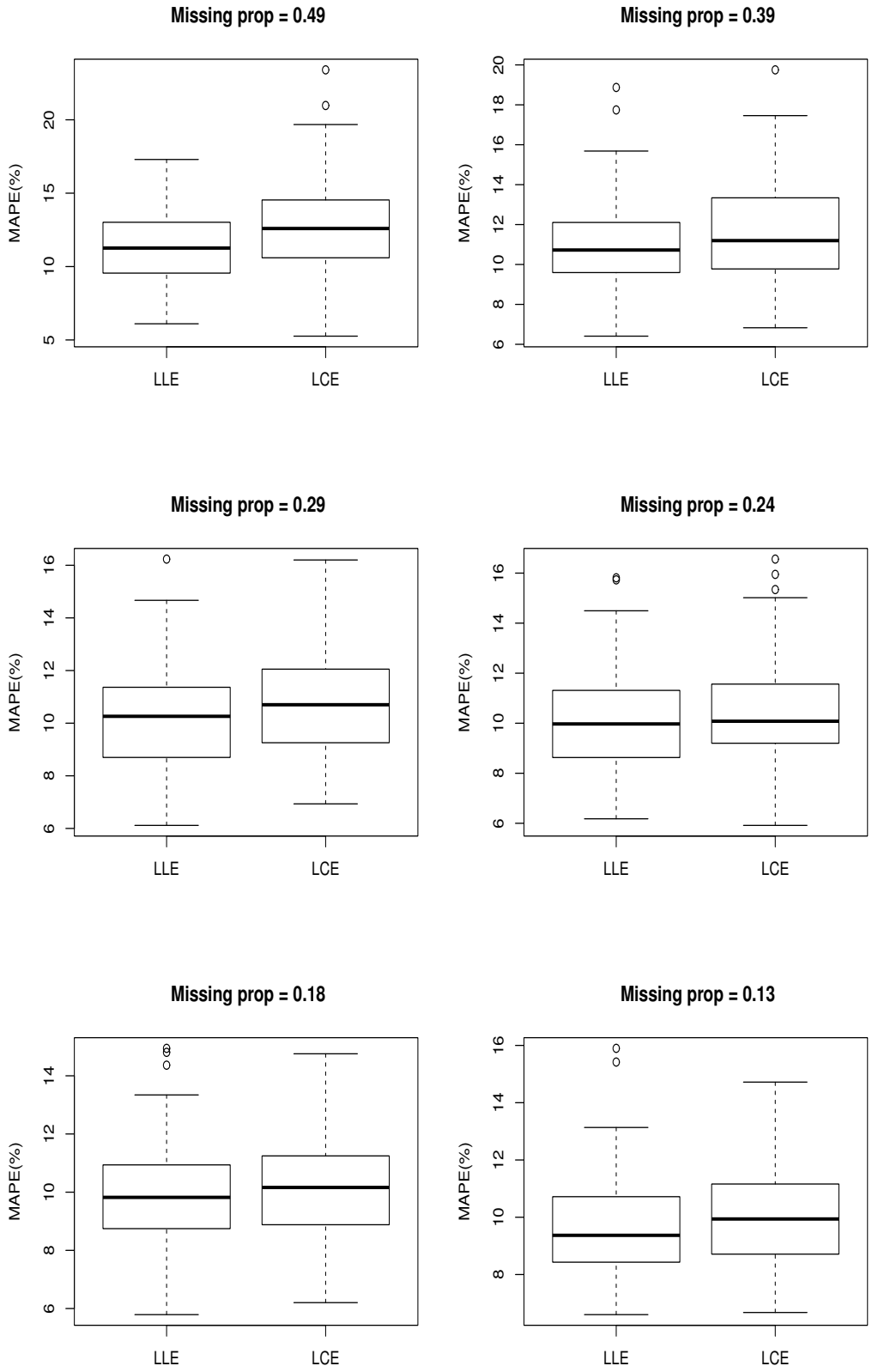

Figure 9. Boxplots of the empirical distribution of the Mean Absolute Percentage Error (MAPE) for the two estimators under the MAR mechanism, based on 100 replications from Model M2.

5. Concluding Remarks

In this article, we have developed a kernel-based estimation procedure for functional regression in the presence of missing responses, with particular attention to the MAR mechanism. Within the framework of stationary and ergodic functional processes, we introduced a local linear estimator of the regression operator and established its principal asymptotic properties. Specifically, we derived uniform almost sure convergence rates, established asymptotic normality, and provided explicit closed-form expressions for both the asymptotic bias and variance. Taken together, these results yield a rigorous and comprehensive asymptotic characterization of the proposed estimator under the MAR paradigm.

Several important theoretical and methodological directions remain open. A central unresolved issue concerns the optimal selection of the smoothing parameters. In kernel-based nonparametric estimation, bandwidth selection plays a decisive role in governing the bias–variance trade-off, thereby directly influencing the estimator's risk profile. While the present study refrains from addressing this aspect, future work should focus on developing fully data-driven selection rules, such as cross-validation, plug-in approaches, penalized criteria, or multiscale adaptive methods, accompanied by a precise theoretical analysis of their asymptotic properties. Establishing minimax-optimal bandwidth selection procedures under MAR represents a particularly challenging, yet essential, research avenue.

A second limitation of the current framework lies in the assumption of strict stationarity. Although this assumption simplifies the asymptotic analysis, it may be unduly restrictive in empirical contexts characterized by temporal heterogeneity or structural evolution. Extending the methodology to locally stationary functional ergodic processes would permit the regression operator to evolve smoothly over time, thereby capturing a richer class of nonstationary dynamics. From a theoretical perspective, this extension would require the derivation of uniform laws of large numbers, central limit theorems, and strong approximation results in the locally stationary functional setting, where classical ergodic tools are no longer directly applicable.

Beyond these aspects, several additional perspectives merit attention. First, it is natural to inquire into the minimax efficiency of the proposed estimator. Establishing lower bounds for estimation under MAR mechanisms and determining whether the convergence rates achieved herein are sharp would provide a deeper understanding of the optimality properties of the procedure. Second, one may consider adaptive estimation strategies capable of accommodating heterogeneity in smoothness across the domain of the regression operator, for example via anisotropic kernels or locally adaptive bandwidths.

Another fertile line of investigation concerns the relaxation of the MAR assumption itself. Extending the framework to allow for Missing Not At Random (MNAR) mechanisms, where the probability of missingness depends explicitly on the unobserved responses, would represent a substantial generalization. Addressing this setting would likely require combining kernel-based functional regression techniques with tools from semiparametric efficiency theory, inverse-probability weighting, and possibly instrumental variable methods.

Finally, the advent of high-dimensional and large-scale functional data calls for the development of computationally efficient variants of the proposed methodology. Embedding the kernel estimators in reproducing kernel Hilbert space (RKHS) structures, employing random projection techniques, or leveraging sparsity and dimension reduction methods could provide avenues for scalability without sacrificing statistical rigor.

Author Contributions: Conceptualization, Y.B., O.B. and S.B.; methodology, Y.B., O.B. and S.B.; validation, Y.B., O.B. and S.B.; formal analysis, Y.B., O.B. and S.B.; investigation, Y.B., O.B. and S.B.; resources, Y.B., O.B. and S.B.; writing—original draft preparation, Y.B., O.B. and S.B.; writing—review and editing, Y.B., O.B. and S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors extend their sincere gratitude to the Editor-in-Chief, the Associate Editor, and the three referees for their invaluable feedback and for pointing out a number of oversights in the version initially submitted. Their insightful comments have greatly refined and focused the original work, resulting in a markedly improved presentation.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

We begin by presenting several preliminary technical lemmas that will be instrumental in obtaining our asymptotic results.

Lemma A1. *Let assumptions (H.4), (H.6), and (H.8) hold. Then, for $h_K \to 0$ and $n \to \infty$, the following conditional moment expansions are valid:*

$$\mathbb{E}\Big(K_j^r \varrho_j \,\Big|\, \mathcal{F}_{j-1}\Big) = p(x)\, M_r\, F_x^{\mathcal{F}_{j-1}}(h_K) + o\Big(F_x^{\mathcal{F}_{j-1}}(h_K)\Big), \qquad r \geq 1, \tag{A1}$$

$$\mathbb{E}\Big(K_j^{\alpha} \beta_j \varrho_j \,\Big|\, \mathcal{F}_{j-1}\Big) = o\Big(p(x)\, h_K\, F_x^{\mathcal{F}_{j-1}}(h_K)\Big), \qquad \alpha > 0, \tag{A2}$$

$$\mathbb{E}\big(\Delta_j K_j \varrho_j \,\big|\, \mathcal{F}_{j-1}\big) = O\Big(p(x)\, n\, h_K^2\, F_x^{\mathcal{F}_{j-1}}(h_K)\Big), \tag{A3}$$

$$\mathbb{E}\Big(K_j^a \beta_j^b \,\Big|\, \mathcal{F}_{j-1}\Big) = N(a,b)\, h_K^b\, F_x^{\mathcal{F}_{j-1}}(h_K) + o\Big(h_K^b\, F_x^{\mathcal{F}_{j-1}}(h_K)\Big), \quad a > 0, \;\; b \in \{2,4\}, \tag{A4}$$

$$\mathbb{E}(\Delta_1 K_1) = (n-1)\, N(1,2)\, M_1\, h_K^2\, F_x^2(h_K)\, \big(1 + o(1)\big). \tag{A5}$$

Remark A1. *Provided that conditions (H.1) and (H.2) are satisfied, the assertion of Lemma A1 is valid uniformly for all $x \in S_{\mathcal{F}}$.*

Lemma A2. *Suppose that assumptions (H.1), (H.2), (H.5-ii), (H.6), and (H.8) hold. Then, in conjunction with Lemma A1, the following assertions are valid:*

$$\text{(i)} \quad \sup_{x \in S_{\mathcal{F}}} \bar{r}_{n,1}(x) = O(1),$$

$$\text{(ii)} \quad \inf_{x \in S_{\mathcal{F}}} \bar{r}_{n,1}(x) = O(1).$$

Proof of Lemma A1. (a) By combining the property of conditional expectation with the condition characterizing missing data at random (MAR), for $r \geq 1$, we can write

$$\begin{aligned}
\mathbb{E}\big(K_j^r \varrho_j \mid \mathcal{F}_{j-1}\big) &= \mathbb{E}\Big(K_j^r \mathbb{E}\big(\varrho_j \mid \mathcal{G}_{j-1}\big) \mid \mathcal{F}_{j-1}\Big) \\
&= p(x) \int_{-1}^{1} K^r(v)\, dP^{\mathbb{F}_{j-1}}_{\frac{\delta(X_j,x)}{h_K}}(v) \\
&= p(x) \int_{-1}^{1} \left[K^r(1) - \int_v^1 (K^r)'(u)\, du\right] dP^{\mathcal{F}_{j-1}}_{\frac{\delta(X_j,x)}{h_K}}(v) \\
&= p(x) K^r(1) F_x^{\mathcal{F}_{j-1}}(h_K) - \int_{-1}^{1} \left[\int_{-1}^{u} dP^{\mathcal{F}_{j-1}}_{\frac{\delta(X_j,x)}{h_K}}(v)\right] (K^r)'(u)\, du \\
&= p(x) K_j^r(1) F_x^{\mathcal{F}_{j-1}}(h_K) - p(x) \int_{-1}^{1} (K^r)'(u) \phi_x^{\mathcal{F}_{j-1}}(-h_K; u h_K)\, du \\
&= p(x) \left[K^r(1) - \int_{-1}^{1} (K^r)'(u) \chi_x(u)\, du\right] F_x^{\mathcal{F}_{j-1}}(h_K) + o\big(F_x^{\mathcal{F}_{j-1}}(h_K)\big) \\
&= p(x) M_r F_x^{\mathcal{F}_{j-1}}(h_K) + o\big(F_x^{\mathcal{F}_{j-1}}(h_K)\big).
\end{aligned}$$

(b) For all $\alpha > 0$, the following holds under assumption (H.6):

$$\mathbb{E}(K_j^\alpha \beta_j \varrho_j \mid \mathcal{F}_{j-1}) \leq C p(x) \int_{B(x,h_K)} \beta(u,x)\, dP_{X_j}^{\mathcal{F}_{j-1}}(h_K).$$

Therefore, applying Assumption (H.8-ii) yields

$$\begin{aligned}
h_K \mathbb{E}(K_j^\alpha \beta_j \varrho_j \mid \mathcal{F}_{j-1}) &= p(x) o\left(\int_{B(\chi,h)} \beta_2(u,\chi)\, dP_{X_j}^{\mathcal{F}_{j-1}}(h_K)\right) \\
&= o(p(x) h_K^2 F_x^{\mathcal{F}_{j-1}}(h_K),
\end{aligned}$$

which immediately implies

$$\mathbb{E}(K_j^\alpha \beta_j \mid \mathcal{F}_{j-1}) = o(p(x) h_K F_x^{\mathcal{F}_{j-1}}(h_K).$$

(c) Since the variables ϱ_j and Y are independent given X_j, we have

$$\mathbb{E}\big(K_j \Delta_j \mathbb{E}\big(\varrho_j \mid \mathcal{G}_{j-1}\big) \mid \mathcal{F}_{j-1}\big) = p(x) \mathbb{E}\big(K_j \Delta_j \mid \mathcal{F}_{j-1}\big).$$

By applying the conditional Cauchy–Schwarz inequality, we obtain

$$\mathbb{E}(\mid K_j \Delta_j \mid\mid \mathcal{F}_{j-1}) \leq \Big[\mathbb{E}(K_j^2 \mid \mathcal{F}_{j-1})\Big]^{1/2} \Big[\mathbb{E}(\Delta_j^2 \mid \mathcal{F}_{j-1})\Big]^{1/2}.$$

Furthermore, we have

$$\Delta_j^2 = \left(\sum_{i=1}^{n} \Big(\beta_i^2 \varrho_i K_i - \beta_j \beta_i \varrho_i K_i\Big)\right)^2 \leq 2n \sum_{i=1}^{n} \Big(\beta_i^2 \varrho_i K_i\Big)^2 + 2n \sum_{i=1}^{n} (\beta_j \beta_i \varrho_i K_i)^2. \tag{A6}$$

Alternatively, under Assumptions (H.4) and (H.6), we establish the following result:

$$\Big[\mathbb{E}(\Delta_j^2 \mid \mathcal{F}_{j-1})\Big]^{1/2} \leq C n h_K^2,$$

Using Lemma A1, we deduce that

$$\mathbb{E}(|K_j\Delta_j| \mid \mathcal{F}_{j-1}) = O\Big(nh_K^2 F_x^{\mathcal{F}_{j-1}}(h_K)\Big) \tag{A7}$$

(d)

Let $a > 0$ and $b = 2, 4$; we can write

$$\mathbb{E}(K_j^a\beta_j^b \mid \mathcal{F}_{j-1}) = \mathbb{E}(K_j^a\delta_j^b) \mid \mathcal{F}_{j-1}) + \mathbb{E}\Big[K_j^a\Big(\beta_j^b - \delta_j^b\Big) \mid \mathcal{F}_{j-1}\Big].$$

where

$$\begin{aligned}
\mathbb{E}(K_j^a\delta_j^b \mid \mathcal{F}_{j-1}) &= \mathbb{E}(K_j^a|\delta_j|^b \mid \mathcal{F}_{j-1}) = h^b\int_{-1}^{1} t^b K^a(t)\, dP^{\mathcal{F}_{j-1}}_{\frac{\delta(X_1,x)}{h_K}}(t),\\
&= h_K^b\int_{-1}^{1}\Big[K^a(1) - \int_t^1 (u^bK^a(u))'\, du\Big] dP^{\mathcal{F}_{j-1}}_{\frac{\delta(X_1,x)}{h_K}}(t)\\
&= h_K^b\Big[K^a(1)\, F_x^{\mathcal{F}_{j-1}}(h_K) - \int_{-1}^{1}(u^bK^a(u))'\, \phi_x^{\mathcal{F}_{i-1}}(-h_K; uh_K)\, du\Big]\\
&= h_K^b F_x^{\mathcal{F}_{j-1}}(h_K)\Big[K^a(1) - \int_{-1}^{1}(u^bK^a(u))'\, \chi_x(u)\, du\Big]\\
&\quad + o(h_K^b F_x^{\mathcal{F}_{j-1}}(h_K))\\
&= N(a,b)h_K^b F_x^{\mathcal{F}_{j-1}}(h_K) + o(h_K^b F_x^{\mathcal{F}_{j-1}}(h_K)).
\end{aligned}$$

Moreover, under assumption (H4-ii), we have

$$\begin{aligned}
&\mathbb{E}\Big[K_j^a\Big(\beta_j^b - \delta_j^b\Big) \mid \mathcal{F}_{j-1}\Big]\\
&\leq \sum_{l=0}^{b-1}\mathbb{E}\left(\sup_{u\in B(x,h_K)}|\beta(u,x) - \delta(u,x)|\Big[K_j^a\mathbf{1}_{B(x,h_k)}|\beta_j|^{b-1-l}|\delta_1|^l \mid \mathcal{F}_{j-1}\Big]\right)\\
&\leq o(h_k)h^{b-1}\mathbb{E}[K_j \mid \mathcal{F}_{j-1}] = o(h_K^b F_x^{\mathcal{F}_{j-1}}(h_K)).
\end{aligned}$$

Hence the proof is complete. □

Proof of Lemma A2. By invoking part (c) of Lemma A1, with $\mathcal{F}_0$ specified as the trivial σ-algebra, and subsequently applying Lemma A.1 of [35], under assumption (H.1), we deduce that

$$nC'h_K^2\varphi(h_K) \leq \mathbb{E}(\Delta_1K_1) \leq nCh_K^2\varphi(h_K). \tag{A8}$$

Applying Lemma A1, under the standing assumptions (H.2) and (H.5), we obtain the following result:

$$\begin{aligned}
\sup_{x\in S_{\mathcal{F}}}|\overline{r}_{n,1}(x)| &\leq \sum_{j=1}^{n}\sup_{x\in S_{\mathcal{F}}}\left|\frac{1}{n\mathbb{E}(\Delta_1K_1)}\mathbb{E}\big(K_j\Delta_j\mathbb{E}(\varrho_j \mid \mathcal{G}_{j-1}) \mid \mathcal{F}_{j-1}\big)\right|\\
&\leq \sup_{x\in S_{\mathcal{F}}} p(x)\sum_{j=1}^{n}\frac{1}{n\varphi(h_K)}\varphi_j(h_K) \to 1 \quad \text{a.co.}
\end{aligned}$$

Consequently, we arrive at the following uniform bound:

$$\sup_{x\in S_{\mathcal{F}}}|\overline{r}_{n,1}(x)| = O(1).$$

Moreover, recalling that the inequality $\big||a|-|b|\big| \leq |a-b|$ holds for all $a, b \in \mathbb{R}$, it follows that

$$\begin{aligned}\inf_{x\in S_{\mathcal{F}}}|\overline{r}_{n,1}(x)| &= \inf_{x\in S_{\mathcal{F}}}|\overline{r}_{n,1}(x)+p(x)-p(x)| \\ &\geq \inf_{x\in S_{\mathcal{F}}}|p(x)| - \underbrace{\sup_{x\in S_{\mathcal{F}}}|\overline{r}_{n,1}(x)-p(x)|}_{I_1}.\end{aligned} \tag{A9}$$

Since $I_1 \xrightarrow[n\to+\infty]{} 0$ almost completely as a consequence of assumption (H.2-ii), we deduce that there exists a strictly positive constant C such that

$$\inf_{x\in S_{\mathcal{F}}}\big|\overline{r}_{n,1}(x)\big| \geq C > 0.$$

This inequality ensures the validity of the required bound, thereby establishing the claim and completing the proof. □

Proof of Lemma 1. With reference to Equation (9), we write

$$\sup_{x\in S_{\mathcal{F}}} \mid \bar{B}_n(x) \mid \leq \frac{\sup_{x\in S_{\mathcal{F}}} \overbrace{\mid \overline{r}_{n,2}(x) - c_0(x)\overline{r}_{n,1}(x) \mid}^{B_n^*(x)}}{\inf_{x\in S_{\mathcal{F}}} \mid \overline{r}_{n,1}(x) \mid}.$$

Accordingly, we derive

$$\begin{aligned}\Big|B_n^*(x)\Big| &= \Big|\frac{p(x)}{n\mathbb{E}(\Delta_1 K_1)}\sum_{j=1}^{n}\big[\mathbb{E}\big(K_j\Delta_j\mathbb{E}\big(Y_j \mid \mathcal{G}_{j-1}\big) \mid \mathcal{F}_{j-1}\big) - c_0(x)\mathbb{E}\big(K_j\Delta_j \mid \mathcal{F}_{j-1}\big)\big]\Big| \\ &\leq \mid p(x) \mid \sum_{j=1}^{n}\Big|\frac{1}{n\mathbb{E}(\Delta_1 K_1)}\big[\mathbb{E}\big(K_j\Delta_j[r(x_j)-c_0(x)] \mid \mathcal{F}_{j-1}\big)\big]\Big|.\end{aligned}$$

Under assumption (H.5), we obtain

$$\begin{aligned}\Big|B_n^*(x)\Big| &\leq C \sup_{u\in B(x,h_K)}|r(x)-r(z)|\sum_{j=1}^{n}\Big|\frac{1}{n\mathbb{E}(\Delta_1 K_1)}\mathbb{E}\big(K_j\Delta_j \mid \mathcal{F}_{j-1}\big)\Big| \\ &\leq C\, h_K^s \sum_{j=1}^{n}\Big|\frac{1}{n\mathbb{E}(\Delta_1 K_1)}\mathbb{E}\big(K_j\Delta_j \mid \mathcal{F}_{j-1}\big)\Big|.\end{aligned}$$

From assumptions (H.1)–(H.2), in conjunction with Lemma (A1) and Equation (A7), we obtain

$$\sup_{x\in S_{\mathcal{F}}}\big|B_n^*(x)\big| = O(h_K^s)\sup_{x\in S_{\mathcal{F}}}\big|\overline{r}_{n1}(x)\big|.$$

This establishes the desired result and thereby completes the proof. □

Proof of Lemma 2. Building upon (6), we rigorously establish the uniform convergence of our estimator over a subset $S \subset \mathcal{F}$. For this purpose, we define

$$z(x) = \arg\min_{z\in\{1,\ldots,d_n\}}|\delta(x,x_z)|.$$

For $t \in \{1,2\}$, we have

$$\begin{aligned}\sup_{x\in S_{\mathcal{F}}} |\hat{r}_{n,t}(x) - \bar{r}_{n,t}(x)| &\leq \sup_{x\in S_{\mathcal{F}}} |\hat{r}_{n,t}(x) - \hat{r}_{n,t}(x_{z(x)}))| \\ &\quad + \sup_{x\in S_{\mathcal{F}}} |\hat{r}_{n,t}(x_{z(x)}) - \bar{r}_{n,t}(x_{z(x)}))| \\ &\quad + \sup_{x\in S_{\mathcal{F}}} |\bar{r}_{n,t}(x_{z(x)})) - -\bar{r}_{n,t}(x))| \\ &= \sum_{s=1}^{3} L_s^t.\end{aligned}$$

We now turn to the analysis of the term L_1^t. Exploiting the boundedness property of K, this term can be expressed as L_1^t as follows:

$$L_1^t \leq \sup_{x\in S_F} \frac{1}{n}\sum_{j=1}^{n}\left|\frac{\Delta_j(x)K_j(x)\varrho_j Y_j^t}{\mathbb{E}[\Delta_1(x)K_1(x)]} - \frac{\Delta_j(x_{z(x)})K_j(x_{k(x)})\varrho_j Y_j^t K_i(x_{z(x)})}{\mathbb{E}[\Delta(x_{z(x)})(K_1(x_{z(x)})]}\right| \cdot \mathbb{1}_{B(x,h_K)\cup B(x_{z(x)},h_K)}(X_j).$$

By applying inequality (A8), we obtain

$$\begin{aligned}L_1^t &\leq \frac{C}{n^2h_K^2\varphi(h_K)} \sup_{x\in S_F}\sum_{j=1}^{n}\left|Y_j^t\right|\left|\Delta_j(x)K_j(x) - \Delta_j(x_{z(x)})K_j(x_{k(x)})\right| \cdot \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_j) \\ &\leq \frac{C}{n^2h_K^2\varphi(h_K)} \sup_{x\in S_F}\sum_{j=1}^{n}\left|Y_j^t\right|\left|\Delta_j(x)\right|\left|K_j(x) - K_j(x_{k(x)})\right| \cdot \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_j) \\ &\quad + \frac{C}{n^2h_K^2\varphi(h_K)} \sup_{x\in S_F}\sum_{j=1}^{n}\left|Y_j^t\right|\left|K_j(x_{k(x)})\right|\left|\Delta_j(x) - \Delta_j(x_{z(x)})\right| \cdot \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_j) \\ &\leq B_1^t + B_2^t.\end{aligned}$$

We begin by examining the term B_1^t. Under assumption (H.6), and invoking the bound

$$\left|\Delta_j(x)\right| \leq nCh_K^2,$$

the quantity B_1^t can be decomposed into the sum of the following three components:

$$B_1^t \leq \sup_{x\in C_{\mathcal{F}}}(B_{11} + B_{12} + B_{13}),$$

with

$$\begin{aligned}B_{11} &= \frac{1}{n\varphi(h_K)}\sum_{j=1}^{n}\left|Y_j^t\right|\left|K_j(x) - K_j(x_{z(x)})\right|\mathbb{1}_{B(x,h_K)\cap B(x_{z(x)},h_K)}(X_j), \\ B_{12} &= \frac{1}{n\varphi(h_K)}\sum_{j=1}^{n}\left|Y_j^t\right|K_j(x)\mathbb{1}_{B(x,h_K)\cap\overline{B(x_{z(x)},h_K)}}(X_j), \\ B_{13} &= \frac{1}{n\varphi(h_K)}\sum_{j=1}^{n}\left|Y_j^t\right|K_j(x_{z(x)})\mathbb{1}_{\overline{B(x,h_K)}\cap B(x_{z(x)},h_K)}(X_j).\end{aligned}$$

Since the kernel K is Lipschitz continuous on the interval $[-1,1]$ in this setting, it immediately follows that

$$B_{11} \leq \frac{C}{n}\sum_{j=1}^{n} Z_j \quad \text{with} \quad Z_j = \frac{\iota_n|Y_j^t|}{h_K\varphi(h_K)} \sup_{x\in C_{\mathbf{F}}} \mathbb{1}_{B(x,h_K)\cap B(x_{z(x)},h_K)}(X_j).$$

Under assumption (H.3), we have

$$\mathbb{E}[|Y|^{\gamma}] = \mathbb{E}[\,\mathbb{E}[|Y|^{\gamma} \mid X]\,] = \int_{X \in B(x,h_K)} m_{\gamma}(x)\, d\mathbb{P}_X(x) < C,$$

which implies that

$$\mathbb{E}(|Z_j|^{\gamma}) \leq \frac{C\iota^{\gamma}}{h_K^{\gamma}\varphi(h_K)^{\gamma-1}}.$$

By invoking the exponential inequality for unbounded random variables (cf. [3], Corollary A.8), and selecting $a^2 = \frac{\varepsilon}{h_K\,\varphi(h_K)}$, we deduce that

$$B_{11} = O_{\text{a.co.}}\left(\frac{\iota_n}{h_K}\sqrt{\frac{\log n}{n\varphi(h_K)}}\right).$$

It suffices to combine the conditions (H.1) and (H.7) to obtain

$$B_{11} = O_{\text{a.co.}}\left(\sqrt{\frac{\log d_n}{n\varphi(h_K)}}\right).$$

Following the same approach used for B_{11}, one can write

$$B_{1l} \leq \frac{C}{n}\sum_{i=1}^{n} Z_j, \text{ for } l \in \{2,3\},$$

where

$$Z_j = \begin{cases} \frac{\left|Y_j^t\right|}{\phi(h_K)} \sup\limits_{x\in S_{\mathcal{F}}}\left[\mathbb{1}_{B(x;h_K)\cap\overline{B(x_{z(x)};h_K)}}(X_j)\right], & \text{for } B_{12}, \\ \frac{\left|Y_j^t\right|}{\phi(h_K)} \sup\limits_{x\in S_{\mathcal{F}}}\left[\mathbb{1}_{B(x_{z(x)},h_K)\cap\overline{B(x;h_K)}}(X_i)\right], & \text{for } B_{13}. \end{cases}$$

Under assumptions (H.1) and (H.3), we have

$$\mathbb{E}(|Z_j|^{\gamma}) \leq \frac{C}{\varphi(h_K)^{\gamma-1}}.$$

By invoking the exponential inequality for unbounded random variables and selecting the parameter $a^2 = \frac{1}{\varphi(h_K)}$, we obtain the following result:

$$B_{1l} = O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n\varphi(h_K)}}\right).$$

We now analyze the term B_2^t. Using the same decomposition as used for B_1^t, we have

$$B_2^t \leq \sup_{x\in S_{\mathcal{F}}}(B_{21} + B_{13}),$$

where

$$B_{21} = \frac{C}{n^2h_K^2\varphi(h_K)} \sup_{x\in S_F}\sum_{j=1}^{n}\left|Y_j^t\right| K_j(x_{k(x)}) \underbrace{\left|\Delta_j(x) - \Delta_j(x_{z(x)})\right| \cdot \mathbb{1}_{B(x,h_K)\cap B(x_{z(x)},h_K)}(X_j)}_{S_1}.$$

Concerning S_1, we have

$$\begin{aligned} S_1 &= \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j)\Big|\Big(\sum_{i=1}^{n}\beta_i^2(x)K_i(x)\varrho_i - \beta_i^2(x_{k(x)})K_i(x_{k(x)}\varrho_i)\Big) \\ &\quad -\Big(\big(\sum_{i=1}^{n}\beta_i(x)K_i(x)\varrho_i)\beta_j(x)\Big) - \Big(\sum_{i=1}^{n}\beta_i(x_{z(x)})K_i(x_{k(x)}\varrho_i)\beta_j(x_{k(x)})\Big)\Big|. \end{aligned}$$

Therefore, we obtain

$$S_1 \leq S_{11} + S_{12},$$

where

$$\begin{aligned} S_{11} &= \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j)\left|\sum_{i=1}^{n}\Big(\beta_i^2(x)K_i(x)\varrho_i - \beta_i^2(x_{z(x)})K_i(x_{z(x)})\varrho_i\Big)\right|, \\ S_{12} &= \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j)\left|\left(\sum_{i=1}^{n}\beta_i(x)K_i(x)\varrho_i\right)\beta_j - \left(\sum_{i=1}^{n}\beta_i(x_{z(x)})K_i(x_{z(x)})\varrho_i\right)\beta_j(x_{z(x)})\right|. \end{aligned} \tag{A10}$$

We put

$$\begin{aligned} A^{(\rho,\beta)} &= \mathbb{1}_{B(x_{k(x)},h_K)\cap B(x,h_K)}(X_j)\left|\left(\sum_{i=1}^{n}\beta_i^{\rho}(x)K_i(x)\varrho_i\right)\beta_j^{\alpha}(x)\right. \\ &\quad \left.-\left(\sum_{i=1}^{n}\beta_i^{\rho}(x_{z(x)})K(x_{z(x)})\varrho_i)\right)\beta_j^{\alpha}(x_{z(x)})\right| \quad \text{with } \rho = 1,2 \quad \text{and} \quad \alpha = 0,1. \end{aligned}$$

We consider the following decomposition:

$$A^{(\rho,\alpha)} \leq A_1^{(\rho,\alpha)} + A_2^{(\rho,\alpha)}, \tag{A11}$$

where

$$A_1^{(\rho,\alpha)} = \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j)\sum_{i=1}^{n}\Big(|\beta_i^{\rho}(x)|\varrho_i K_i(x)\Big|\big(\beta_j^{\alpha}(x) - \beta_j^{\alpha}(x_{z(x)})\big)\Big|\Big),$$

and

$$\begin{aligned} &A_2^{(\rho,\alpha)} \\ &= \mathbb{1}_{B(x_{z(x)},h_k)\cap B(x,h_K)}(X_j)\Big(\sum_{i=1}^{n}|\beta_j^{\rho}(x_{z(x)})|K_i(x)\varrho_i \times |\beta_i^{\alpha}(x) - \beta_i^{\alpha}(x_{z(x)})|\Big) \\ &\quad +\mathbb{1}_{B(x_{z(x)},h_k)\cap B(x,h_K)}(X_j)\Big(\sum_{i=1}^{n}|\beta_j^{\rho}(x_{z(x)})||\beta_i^{\alpha}(x_{z(x)})|\varrho_i \times |K_i(x) - K_i(x_{z(x)})|\Big). \end{aligned}$$

Under assumption (H.4) for all $\alpha \geq 1$, we have

$$\begin{aligned} &\mathbb{1}_{B(x_{z(x)},h_k)\cap B(x,h_K)}(X_j)\Big|\beta_i^{\alpha}(x) - \beta_i^{\alpha}(x_{k(x)})\Big| \\ &\leq \mathbb{1}_{B(x_{z(x)},h_k)\cap B(x,h_K)}(X_j) \mid \beta_i(x) - \beta_i(x_{z(x)}) \mid \sum_{l=0}^{\alpha-1} \mid \beta_i^{\alpha-1-l}(x)\beta_i(x_{z(x)}) \mid \\ &\leq C\iota_n h_k^{\alpha-1}\mathbb{1}_{B(x_{z(x)},h_k)\cap B(x,h_K)}(X_j). \end{aligned}$$

This implies that, for $\rho = \alpha = 1$, we obtain

$$A_1^{(1,1)} \leq C\, n\, \iota_n\, h_K\, \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j),$$

and

$$A_2^{(1,1)} \leq C\, n\, \iota_n\, h_K\, \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j).$$

However, in the case $\rho = 2$ and $\alpha = 0$, we obtain

$$A_2^{(2,0)} \leq C\, n\, \iota_n\, h_K\, \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j), \qquad A_1^{(2,0)} = 0.$$

Combining (A10) and (A11), we find that

$$\begin{aligned} S_1 = S_{11} + S_{12} &\leq A^{(2,0)} + A^{(1,1)} \\ &\leq C\, n\, \iota_n\, h_K\, \mathbb{1}_{B(x_{z(x)},h_K)\cap B(x,h_K)}(X_j). \end{aligned}$$

Then, by assumption (H.6), we obtain

$$B_2^t \leq \frac{C\iota_n}{nh_K\varphi(h_K)} \sum_{j=1}^{n} |Y_j^t| \sup_{x\in S_{\mathbf{F}}} \mathbb{1}_{B(x,h_K)\cap B(x_{z(x)},h_K)}(X_j).$$

Following the same idea used in the proof of B_1^t, we obtain

$$B_2^t = O_{\text{a.co.}}\left(\sqrt{\frac{\log d_n}{n\varphi(h_K)}}\right).$$

Concerning the term L_2^t, for all $\epsilon > 0$, we have

$$\begin{aligned} \mathbb{P}(L_2^t > \epsilon) &= \mathbb{P}\left(\max_{z\in\{1,\ldots,d_n\}} |\hat{r}_{nt}(x_{z(x)}) - \bar{r}_{nt}(x_{z(x)}))| > \epsilon\right) \\ &\leq d_n \max_{z\in\{1,\ldots,d_n\}} \mathbb{P}\Big(|\hat{r}_{nt}(x_{z(x)}) - \bar{r}_{nt}(x_{z(x)}))| > \epsilon\Big). \end{aligned}$$

However, we can write

$$\hat{r}_{nt}(x_{z(x)}) - \bar{r}_{nt}(x_{z(x)})) = \frac{1}{n\mathbb{E}(\Delta_1 k_1)} \sum_{j=1}^{n} \Gamma_{jt}(x), \tag{A12}$$

where $\Gamma_{jt}(x) = \Delta_j k_j(x) Y_j^t \varrho_j - \mathbb{E}\Big(\Delta_j k_j(x) Y_j^t \varrho_j \mid \mathcal{F}_{j-1}\Big)$ is a martingale difference. Consequently, Lemma A3 can be invoked to establish an exponential upper bound for the quantity L_2^t. To this end, we verify the conditions under which such an exponential bound can be obtained. Thereafter, applying Jensen's inequality yields

$$\begin{aligned} \mathbb{E}((\Gamma_{jt}(x))^2 \mid \mathcal{F}_{j-1}) &\leq \mathbb{E}(K_j^2\Delta_j^2\varrho_j \mid \mathcal{F}_{j-1}) \\ &\leq \mathbb{E}\Big(K_j^2\Delta_j^2\mathbb{E}(\varrho_j \mid \mathcal{G}_{j-1}) \mid \mathcal{F}_{j-1}\Big) \\ &\leq P(x)\mathbb{E}\Big(K_j^2\Delta_j^2 \mid \mathcal{F}_{j-1}\Big). \end{aligned} \tag{A13}$$

From Lemma A1 and inequality (A6), it follows that

$$\mathbb{E}((\Gamma_{jt}(x))^2 \mid \mathcal{F}_{j-1}) \leq 2Cn^2h_K^4\varphi_j(h_K).$$

We take

$$D_n = \sum_{j=1}^{n} d_j^2, \text{ with } d_j^2 = n^2h_k^4\varphi_j(h_K).$$

Thus, for any $\epsilon > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\left|L_2^t(x_{z(x)})\right| \geq \epsilon_0\sqrt{\frac{\log(d_n)}{n\varphi(h_K)}}\right) \\
&= \mathbb{P}\left(\left|\sum_{j=1}^{n}\Gamma_{jt}(x)\right| \geq n\mathbb{E}(\Delta_1 K_1)\epsilon_0\sqrt{\frac{\log(d_n)}{n\varphi(h_K)}}\right) \\
&\leq 2\exp\left(-\frac{(n\mathbb{E}(\Delta_1 K_1)\epsilon_0)^2\frac{\log(d_n)}{n\varphi(h_K)}}{\left(2D_n\left[1+\frac{Cn\mathbb{E}(\Delta_1 K_1)}{D_n}\epsilon_0\sqrt{\frac{\log(d_n}{n\varphi(h_K))}}\right]\right)}\right).
\end{aligned}$$

Making use of the statement (A8), we have

$$\frac{\varphi(h_K)}{nh_K^2\sum_{j=1}^{n}\varphi_j(h_K)} \leq \frac{\mathbb{E}(\Delta_1 K_1)}{D_n} \leq \frac{\varphi(h_K)}{nh_K^2\sum_{j=1}^{n}\varphi_j(h_K)}. \tag{A14}$$

Using assumptions (H.2-ii) and (H.8), we deduce that

$$\lim_{n\to\infty}\frac{\mathbb{E}(\Delta_1 K_1)}{D_n} = 0.$$

Then, we derive

$$\mathbb{P}\left(|L_2^t| > \epsilon_0\sqrt{\frac{\log d_n}{n\phi(h_K)}}\right) \leq 2\exp\left(-C\epsilon_0^2\log d_n\right).$$

Thus, by choosing ϵ_0 such that $C\epsilon_0^2 = \rho$, we get

$$d_n\max_{k\in\{1,\ldots,d_n\}}\mathbb{P}\left(|L_2^t| > \epsilon_0\sqrt{\frac{\log d_n}{n\phi(h_K)}}\right) \leq Cd_n^{1-\rho}.$$

Since

$$\sum_{n=1}^{\infty} d_n^{1-\rho} < \infty,$$

we obtain that

$$L_2^t = O_{a.co.}\left(\sqrt{\frac{\log d_n}{n\phi(h_K)}}\right).$$

Concerning the term of L_3^t, it is obvious that

$$L_3^t \leq \mathbb{E}\left(\sup_{x\in S_{\mathcal{F}}} \mid \hat{r}_{nt}(z(x)) - \bar{r}_{nt}(z(x)) \mid \left|\mathcal{F}_{j-1}\right.\right),$$

and we infer that

$$L_3^t = O_{a.co.}\left(\sqrt{\frac{\log\, d_n}{n\phi(h_K)}}\right).$$

Hence, the proof is complete. □

Proof of Lemma 3. According to the inequality of (A9), we have

$$\sup_{x\in S_{\mathcal{F}}}|\bar{r}_{n,1}(x) - p(x)| \leq \frac{1}{2}\inf_{x\in S_{\mathcal{F}}}|p(x)| \Longrightarrow \inf_{x\in S_{\mathcal{F}}}|\bar{r}_{n,1}(x)| \geq \frac{1}{2}\inf_{x\in\mathcal{C}_F} p(x)$$

meaning that there exists $x \in S_{\mathcal{F}}$ such that

$$\mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}}|\bar{r}_{n,1}(x)| \leq \frac{1}{2}\inf_{x\in S_{\mathcal{F}}} p(x)\right) \leq \mathbb{P}\left(\sup_{x\in C_F}|\bar{r}_{n,1}(x) - P(x)| \geq \frac{1}{2}\inf_{x\in S_{\mathcal{F}}}|p(x)|\right).$$

Taking $\mathcal{F}_0$ as a trivial algebra $\sigma-$ in Lemma 2, we deduce that there exists

$$\epsilon = \frac{1}{2}\inf_{x\in S_{\mathcal{F}}} p(x)$$

such that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}}|\bar{r}_{n,1}(x)| \leq C\right) < \infty.$$

Hence, the proof is concluded. □

Proof of Lemma 4. We have the following decomposition:

$$\hat{r}_{n1}(x) = A_{n,1}(x) + \bar{r}_{n,1}(x),$$

where

$$A_{n,1}(\chi) = \frac{1}{n\mathbb{E}[\Delta_1 K_1]}\sum_{j=1}^{n}(\Delta_j K_j \varrho_j) - \mathbb{E}[\Delta_j K_j \varrho_j \mid \mathcal{F}_{j-1}]).$$

The primary objective is to establish that

$$A_{n,1}(x) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty.$$

We denote

$$\eta_{n,i}(\chi) := \Delta_i(x)K_i(x)\varrho_i - \mathbb{E}[\Delta_i(x)K_i(x)\varrho_i \mid \mathcal{F}_{i-1}], \quad i = 1, 2, \ldots, n.$$

Thus, the sequence $\eta_{n,i} : 1 \leq i \leq n$ naturally arises as a triangular array of martingale difference terms adapted to the filtration $\mathscr{F}_{i-1}$. Note that

$$\hat{r}_{n,1}(x) - \bar{r}_{n,1}(x) = \frac{1}{n\mathbb{E}[\Delta_1 K_1]}\sum_{i=1}^{n}\eta_{n,i}(\chi).$$

On the other hand, by appealing to Burkholder's inequality for martingale difference sequences [55], it follows that

$$\begin{aligned}\mathbb{P}(\hat{r}_{n,1}(x) - \bar{r}_{n,1}(x)| > \varepsilon) &= \mathbb{P}\left(\left|\sum_{i=1}^{n}\eta_{n,i}(\chi)\right| > \varepsilon n\mathbb{E}[\Delta_1 K_1]\right)\\ &\leq \frac{C\mathbb{E}[\eta_{n,1}^2(\chi)]}{\varepsilon^2 n^2(\mathbb{E}[\Delta_1 K_1])^2}\\ &\leq \frac{CP(x)\mathbb{E}[\Delta_1^2 K_1^2]}{\varepsilon^2 n^2\mathbb{E}^2[\Delta_1 K_1]}.\end{aligned}$$

By invoking inequality (A8) and considering $\mathcal{F}_0$ as the trivial σ-algebra in (A13), we consequently obtain that

$$\mathbb{P}(\hat{r}_{n,1}(x) - \bar{r}_{n,1}(x)| > \varepsilon) \leq \frac{Cp(x)}{\varepsilon^2 n^2\varphi(h_K)} \to 0, \quad \text{as} \quad n \to \infty.$$

Secondly, we will prove that

$$\bar{r}_{n,1}(\chi) \xrightarrow{\mathbb{P}} p(x) \quad \text{as } n \to \infty.$$

By invoking the sufficient condition for convergence in probability, it remains only to establish the following results:

$$\lim_{n\to+\infty} \mathbb{E}(\bar{r}_{n1}(x)) = P(x) \quad \text{and} \quad \lim_{n\to+\infty} \mathrm{Var}(\bar{r}_{n1}(x)) = 0.$$

We have

$$\begin{aligned}\mathbb{E}(\bar{r}_{n1}(x)) &= \frac{1}{n\mathbb{E}(\Delta_1 K_1)} \sum_{j=1}^{n} \mathbb{E}\big[\mathbb{E}\big(K_j\Delta_j \mathbb{E}\big(\varrho_j \mid \mathbb{G}_{j-1}\big) \mid \mathbb{F}_{j-1}\big)\big] \\ &= p(x) \underbrace{\frac{1}{n\mathbb{E}(\Delta_1 K_1)} \sum_{j=1}^{n} \mathbb{E}(K_j\Delta_j)}_{=1}.\end{aligned}$$

Note that

$$\begin{aligned}&\mathrm{Var}(\bar{r}_{n1}(x)) \le \mathbb{E}((\bar{r}_{n1}(x))^2) \\ &\le \frac{1}{n^2\mathbb{E}^2(\Delta_1 K_1)} \left[\sum_{j=1}^{n} \mathbb{E}(K_j^2\Delta_j^2\varrho_j) + \sum_{\substack{j=1 \\ l\neq j}}^{n}\sum_{l=1}^{n} \big|\mathbb{E}\big(\mathbb{E}[\Delta_l K_l \varrho_l \mid \mathcal{F}_{l-1}]\mathbb{E}\big[\Delta_j K_j \varrho_j \mid \mathcal{F}_{j-1}\big]\big)\big|\right].\end{aligned}$$

By applying Lemma A1, we obtain

$$\mathrm{Var}(\bar{r}_{n1}(x)) \le \frac{1}{n^2\varphi^2(h_K)} \Big(p(x) + p^2(x)\mathfrak{D}_{l,j}(h_K)\Big).$$

Under assumption (H.8), we deduce that

$$\lim_{n\to+\infty} \mathrm{Var}(\bar{r}_{n1}(x)) = 0.$$

Hence, the proof is complete. □

Proof of Lemma 5. We denote

$$\mathrm{Y}_{n,j} = \frac{\sqrt{n\varphi(h_K)}}{n\mathbb{E}(\Delta_1 K_1)} \Delta_j K_j \varrho_j (Y_j - c_0(x))).$$

We define $\zeta_{n,j} = \mathrm{Y}_{n,j} - \mathbb{E}(\mathrm{Y}_{n,j} \mid \mathcal{F}_{j-1})$. It is clear that

$$\sqrt{n\varphi(h)} Q_n(x) = \sum_{j=1}^{n} \zeta_{n,j}.$$

Moreover, to prove Lemma 5, it suffices to establish the following two statements:

$$\sum_{j=1}^{n} \mathbb{E}(\zeta_{n,j} \mid \mathcal{F}_{j-1}) \xrightarrow{\mathbb{P}} V(x), \tag{A15}$$

and

$$\forall \varepsilon > 0, \quad n\mathbb{E}\Big(\zeta_{n,j} \mathbf{1}_{\{|\zeta_{n,j}|>\varepsilon\}}\Big) = o(1). \tag{A16}$$

Concerning the proof of (A15), we have

$$\mathbb{E}(\zeta_{n,j}^2 \mid \mathcal{F}_{j-1}) = \mathbb{E}(\mathrm{Y}_{n,j}^2 \mid \mathcal{F}_{j-1}) - \big(\mathbb{E}(\mathrm{Y}_{n,j} \mid \mathcal{F}_{j-1})\big)^2. \tag{A17}$$

For the second term on the right-hand side of Equation (A17), we use Lemma 1 and the statement (A8) to conclude that

$$\big|\mathbb{E}[\mathrm{Y}_{n,i} \mid \mathcal{F}_{j-1}]\big| \leq \frac{p(x)\sqrt{n\varphi(h_K)}}{nh_K^2\varphi(h_K)}\big|\mathbb{E}\big[(r(X_j) - c_0(x))\Delta_j K_j \mid \mathcal{F}_{j-1}\big]\big|.$$

Moreover,

$$\begin{aligned}\big|\mathbb{E}\big[(r(X_j) - c_0(x))\Delta_j K_j \mid \mathcal{F}_{j-1}\big]\big| &\leq \sup_{u\in B(x,h_K)} |r(u) - r(x)|\,\mathbb{E}[\Delta_j K_j \mid \mathcal{F}_{j-1}] \\ &\leq n\,h_K^{b+2}\,F_x^{\mathcal{F}_{j-1}}(h_K).\end{aligned}$$

Then

$$\big|\mathbb{E}[\mathrm{Y}_{n,i} \mid \mathcal{F}_{j-1}]\big| = O\left(\sqrt{n\varphi(h_K)}h_K^b\frac{F_x^{\mathcal{F}_{j-1}}(h_K)}{n\varphi(h_K)}\right).$$

Under assumption (H.2), we obtain

$$\sum_{j=1}^{n}\mathbb{E}[(\mathrm{Y}_{n,i} \mid \mathcal{F}_{j-1}])^2 = O(n\varphi(h_K)h_K^{2b}).$$

In order to establish the statement (A16), we consider the following decomposition:

$$\sum_{j=1}^{n}\mathbb{E}\Big[\mathrm{Y}_{n,i}^2 \mid \mathcal{F}_{j-1}\Big] = \frac{\varphi(h_K)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[(Y_j - c_0(x))^2\varrho_j\Delta_j^2K_j^2 \mid \mathcal{F}_{i-1}\Big] := J_{1n} + J_{2n},$$

where

$$J_{1n} = \frac{\varphi(h_K)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[(Y_j - r(X_j))^2\varrho_j\Delta_j^2K_j^2 \mid \mathcal{F}_{i-1}\Big],$$

and

$$J_{2n} = \frac{\varphi(h_K)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[(r(X_j) - c_0(x))^2\varrho_j\Delta_j^2K_j^2 \mid \mathcal{F}_{i-1}\Big].$$

By the properties of conditional expectations, we obtain

$$J_{1n} = \frac{\varphi(h_K)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[\mathbb{E}\Big[(Y_j - r(X_j))^2\varrho_j\Delta_j^2K_j^2 \mid \mathcal{G}_{i-1}\Big] \mid \mathcal{F}_{i-1}\Big].$$

Then, we have

$$= \frac{\phi(h)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[\Delta_j^2K_j^2\,\mathbb{E}\Big[(Y_j - r(X_j))^2\varrho_j \mid X_j\Big] \mid \mathcal{F}_{i-1}\Big].$$

Combining assumption B(2) and Lemma A1 with Lemma. A.2 (ii) of [56], we get

$$\begin{aligned} J_{1n} &= \sigma^2(x)p(x)\frac{\varphi(h_K)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[\Delta_j^2 K_j^2 \mid \mathcal{F}_{i-1}\Big] \\ &= \frac{\sigma^2(x)p(x)}{(n-1)^2\, N(1,2)^2\, M_1^2 h_K^4 \varphi^2(h_K)}\frac{\sum_{j=1}^{n}\mathbb{E}\Big[\Delta_j^2 K_j^2 \mid \mathcal{F}_{i-1}\Big]}{n\varphi(h_K)} \\ &\to \frac{M_2\,\sigma^2(x)p(x)}{M_1^2}, \quad \text{as } n\to\infty. \end{aligned}$$

By the assumptions (H.2) and (H.4), and using the statements (A13) and (A8), we obtain

$$\begin{aligned} |J_{2n}| &\leq p(x)h_K^{2b}\frac{\phi(h)}{n(\mathbb{E}[\Delta_1 K_1])^2}\sum_{j=1}^{n}\mathbb{E}\Big[\Delta_j^2 K_j^2\,\mathbb{E}\mid \mathcal{F}_{i-1}\Big] \\ &\leq h_K^{2b}\frac{\sum_{j=1}^{n}\varphi_j(h_K)}{n\varphi(h_K)} \to 0, \quad \text{as} \quad n\to\infty. \end{aligned}$$

Concerning Equation (A16), for all $\mu > 0$, we have

$$n\mathbb{E}\Big(\zeta_{n1}^2 \mathbb{1}_{(|(\zeta_{n1}|>\mu\sqrt{V(x))}}\Big) \leq 4n\mathbb{E}\Big(\zeta_{n1}^2 \mathbb{1}_{(|\zeta n1|>\mu/2\sqrt{V(x))}}\Big).$$

Let $\alpha > 1$ and $\beta > 1$ such that

$$\frac{1}{\alpha}+\frac{1}{\beta}=1.$$

By Hölder and Markov inequalities, one can write for all $\mu > 0$,

$$n\mathbb{E}\Big(\mathrm{Y}_{n1}^2 \mathbb{1}_{(|\mathrm{Y}_{n1}|>\mu/2\sqrt{V(x))}}\Big) \leq \frac{4n\mathbb{E}|\mathrm{Y}_{n1}|^{2\alpha}}{(\mu/2)^{\frac{2\alpha}{\beta}}\,(V(x))^{\frac{\alpha}{\beta}}}.$$

Let $a = 1+\frac{\delta}{2}$ for any $\delta > 0$ such that

$$\overline{M}_{2+\delta}(x) = \mathbb{E}\Big[|Y_1 - c_0(x)|^{2+\delta} \mid X_1 = x\Big],$$

is a continuous function. It follows that

$$\begin{aligned} 4n\,\mathbb{E}|\mathrm{Y}_{n,1}^2| &\leq C\left(\frac{\varphi(h_K)}{n}\right)^{\frac{21\delta}{2}} n\,(\mathbb{E}[\Delta_1 K_1])^{-2-\delta}\,\mathbb{E}\Big[|(Y_1 - c_0(x))\Delta_1 K_1|^{2+\delta}\Big] \\ &\leq C\left(\frac{\varphi(h_K)}{n}\right)^{\frac{2+\delta}{2}} n\,(\mathbb{E}[\Gamma_1 K_1])^{-2-\delta}\,\mathbb{E}\Big[|\Delta_1 K_1|^{2+\delta}\Big]\,\mathbb{E}\Big[|Y_1 - c_0(x)|^{2+\delta} \mid X\Big] \\ &\leq C\left(\frac{\varphi(h_K)}{n}\right)^{\frac{2+\delta}{2}} n\,(\mathbb{E}[\Delta_1 K_1])^{-2-\delta}\,\mathbb{E}\Big[|\Delta_1 K_1|^{2+\delta}\Big]\,\overline{M}_{2+\delta}(x) \\ &= O\Big((n\varphi(h_K))^{-\delta/2}\Big) \quad \text{as } n\to\infty. \end{aligned}$$

This completes the proof. □

Appendix B

This appendix presents supplementary material that is integral to developing a more thorough and nuanced understanding of the paper's contributions. In contexts involving dependent data—particularly under ergodicity—martingale difference sequences (MDSs) emerge as a central analytical framework. The derivation of general results for MDS is

essential for investigating the asymptotic properties of a wide class of statistics, stochastic processes, and estimators that can be represented in terms of these sequences.

Definition A1. *Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration (an increasing sequence of σ algebras), and let $\{D_n\}_{n\geq 1}$ be a sequence of integrable random variables with real values adapted to $\{\mathcal{F}_n\}$. The sequence $\{D_n\}$ is called a martingale difference sequence (MDS) with respect to $\{\mathcal{F}_n\}$ if, for all $n \geq 1$,*

$$\mathbb{E}[D_n \mid \mathcal{F}_{n-1}] = 0 \quad a.s.$$

In the asymptotic analysis of our estimators for functional ergodic data, rigorous control of the deviation probabilities associated with partial sums of unbounded martingale differences is of critical importance. The lemma below establishes an exponential deviation bound for these partial sums, under a suitable moment growth condition.

Lemma A3 (Exponential Bound for Martingale Partial Sums)**.** *Let $(Z_n, \mathcal{F}_n)_{n\geq 1}$ be a martingale difference sequence with respect to its natural filtration. Suppose there exist constants $C > 0$ and a sequence $\{d_n\}$ such that, for all $p \geq 2$,*

$$\mathbb{E}[|Z_n|^p \mid \mathcal{F}_{n-1}] \leq C^{p-2} p!\, d_n^2 \quad a.s.$$

Then, for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i\right| > \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right),$$

where

$$D_n = \sum_{i=1}^{n} d_i^2.$$

Theorem A1 (Lindeberg central limit theorem)**.** *Let $U_{n1}, \ldots, U_{nr_n}$ be a triangular array of independent random variables for each $n \geq 1$, satisfying*

$$\mathbb{E}(U_{nj}) = 0 \quad \text{and} \quad \mathrm{Var}(U_{nj}) < \infty \quad \text{for all } j = 1, \ldots, r_n.$$

Define the normalized variables

$$\tilde{U}_{nj} = \frac{U_{nj}}{\sqrt{\sum_{k=1}^{r_n} \mathrm{Var}(U_{nk})}}, \quad j = 1, \ldots, r_n.$$

Then the sum of the normalized variables converges in distribution to the standard normal distribution:

$$\sum_{j=1}^{r_n} \tilde{U}_{nj} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } n \to \infty,$$

provided the following Lindeberg condition holds: for every $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=1}^{r_n} \mathbb{E}\left[\tilde{U}_{nj}^2 \mathbb{1}_{\{|\tilde{U}_{nj}|>\epsilon\}}\right] = 0.$$

References

1. Bosq, D. *Linear Processes in Function Spaces*; Lecture Notes in Statistics; Theory and Applications; Springer: New York, NY, USA, 2000; Volume 149, p. xiv+283. [CrossRef]
2. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2005; p. xx+426. [CrossRef]

3. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Springer Series in Statistics; Theory and Practice; Springer: New York, NY, USA, 2006; p. xx+258. [CrossRef]
4. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer Series in Statistics; Springer: New York, NY, USA, 2012. [CrossRef]
5. Zhang, J. *Analysis of Variance for Functional Data*; Monographs on Statistics and Applied Probability; CRC Press: Boca Raton, FL, USA, 2014; Volume 127. [CrossRef]
6. Qing, S.; Choi, T. *Gaussian Process Regression Analysis for Functional Data*; CRC Press: Boca Raton, FL, USA, 2011. [CrossRef]
7. Goia, A.; Vieu, P. A partitioned single functional index model. *Comput. Statist.* **2015**, *30*, 673–692. [CrossRef]
8. Chen, D.; Hall, P.; Müller, H. Single and Multiple Index Functional Regression Models with Nonparametric Link. *Ann. Stat.* **2011**, *39*, 1720–1747. [CrossRef]
9. Aneiros, G.; Vieu, P. Partial linear modelling with multi-functional covariates. *Comput. Statist.* **2015**, *30*, 647–671. [CrossRef]
10. Ferré, L.; Yao, A.F. Smoothed functional inverse regression. *Statist. Sin.* **2005**, *15*, 665–683.
11. Horváth, L.; Rice, G. An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.* **2015**, *28*, 505–548. [CrossRef]
12. Müller, H. Peter Hall, functional data analysis and random objects. *Ann. Statist.* **2016**, *44*, 1867–1887. [CrossRef]
13. Ling, N.; Vieu, P. Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics* **2018**, *52*, 934–949. [CrossRef]
14. Goia, A.; Vieu, P. An introduction to recent advances in high/infinite dimensional statistics [Editorial]. *J. Multivar. Anal.* **2016**, *146*, 1–6. [CrossRef]
15. Bouzebda, S. Weak convergence of the conditional single index U-statistics for locally stationary functional time series. *AIMS Math.* **2024**, *9*, 14807–14898. [CrossRef]
16. Masry, E. Multivariate regression estimation—Local polynomial fitting for time series. *Stoch. Process. Appl.* **1996**, *65*, 81–101. [CrossRef]
17. Masry, E. Multivariate regression estimation of continuous-time processes from sampled data: Local polynomial fitting approach. *IEEE Trans. Inform. Theory* **1999**, *45*, 1939–1953. [CrossRef]
18. Masry, E. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal.* **1996**, *17*, 571–599. [CrossRef]
19. Masry, E.; Fan, J. Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* **1997**, *24*, 165–179. [CrossRef]
20. Masry, E.; Mielniczuk, J. Local linear regression estimation for time series with long-range dependence. *Stoch. Process. Appl.* **1999**, *82*, 173–193. [CrossRef]
21. Wu, J.; Peng, H.; Tu, W. Large-sample estimation and inference in multivariate single-index models. *J. Multivar. Anal.* **2019**, *171*, 382–396. [CrossRef]
22. Cheng, P.E. Nonparametric estimation of mean functionals with data missing at random. *J. Am. Stat. Assoc.* **1994**, *89*, 81–87. [CrossRef]
23. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley-Interscience [John Wiley & Sons]: Hoboken, NJ, USA, 2002; p. xviii+381. [CrossRef]
24. Nittner, T. Missing at random (MAR) in nonparametric regression—A simulation experiment. *Stat. Methods Appl.* **2003**, *12*, 195–210. [CrossRef]
25. Wang, Q.; Linton, O.; Härdle, W. Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.* **2004**, *99*, 334–345. [CrossRef]
26. Liang, H.; Wang, S.; Carroll, R.J. Partially linear models with missing response variables and error-prone covariates. *Biometrika* **2007**, *94*, 185–198. [CrossRef]
27. Efromovich, S. Nonparametric regression with predictors missing at random. *J. Amer. Statist. Assoc.* **2011**, *106*, 306–319. [CrossRef]
28. Josse, J.; Reiter, J.P. Introduction to the special section on missing data. *Statist. Sci.* **2018**, *33*, 139–141. [CrossRef]
29. Ferraty, F.; Sued, M.; Vieu, P. Mean estimation with data missing at random for functional covariables. *Statistics* **2013**, *47*, 688–706. [CrossRef]
30. Ling, N.; Liang, L.; Vieu, P. Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J. Statist. Plann. Inference* **2015**, *162*, 75–87. [CrossRef]
31. Febrero-Bande, M.; Galeano, P.; González-Manteiga, W. Estimation, imputation and prediction for the functional linear model with scalar response with responses missing at random. *Comput. Statist. Data Anal.* **2019**, *131*, 91–103. [CrossRef]
32. Kraus, D. Inferential procedures for partially observed functional data. *J. Multivar. Anal.* **2019**, *173*, 583–603. [CrossRef]
33. Bouanani, O.; Bouzebda, S. Limit theorems for local polynomial estimation of regression for functional dependent data. *AIMS Math.* **2024**, *9*, 23651–23691. [CrossRef]

34. Baíllo, A.; Grané, A. Local linear regression for functional predictor and scalar response. *J. Multivar. Anal.* **2009**, *100*, 102–111. [CrossRef]
35. Barrientos-Marin, J.; Ferraty, F.; Vieu, P. Locally modelled regression and functional data. *J. Nonparametr. Stat.* **2010**, *22*, 617–632. [CrossRef]
36. Kolmogorov, A.N.; Tihomirov, V.M. ε-entropy and ε-capacity of sets in function spaces. *Uspehi Mat. Nauk* **1959**, *14*, 3–86.
37. Bouzebda, S.; Nezzal, A. Uniform in number of neighbors consistency and weak convergence of kNN empirical conditional processes and kNN conditional U-processes involving functional mixing data. *AIMS Math.* **2024**, *9*, 4427–4550. [CrossRef]
38. Ferraty, F.; Laksaci, A.; Tadj, A.; Vieu, P. Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference* **2010**, *140*, 335–352. [CrossRef]
39. Kuelbs, J.; Li, W.V. Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.* **1993**, *116*, 133–157. [CrossRef]
40. Nicoleris, T.; Yatracos, Y.G. Rates of convergence of estimates, Kolmogorov's entropy and the dimensionality reduction principle in regression. *Ann. Statist.* **1997**, *25*, 2493–2511. [CrossRef]
41. Li, W.V.; Shao, Q.M. Gaussian processes: Inequalities, small ball probabilities and applications. In *Handbook of Statistics, Stochastic Processes: Theory and Methods*; Elsevier: Amsterdam, The Netherlands, 2001; Volume 19, pp. 533–597. [CrossRef]
42. Laib, N.; Louani, D. Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *J. Multivar. Anal.* **2010**, *101*, 2266–2281. [CrossRef]
43. Litimein, O.; Alshahrani, F.; Bouzebda, S.; Laksaci, A.; Mechab, B. Kolmogorov Entropy for Convergence Rate in Incomplete Functional Time Series: Application to Percentile and Cumulative Estimation in High Dimensional Data. *Entropy* **2023**, *25*, 1108. [CrossRef]
44. Rosenblatt, M. Uniform ergodicity and strong mixing. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1972**, *24*, 79–84. [CrossRef]
45. Bouzebda, S.; Didi, S. Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: Asymptotic results. *Comm. Statist. Theory Methods* **2017**, *46*, 1367–1406. [CrossRef]
46. Bouzebda, S.; Didi, S. Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes. *Rev. Mat. Complut.* **2021**, *34*, 811–852. [CrossRef]
47. Bradley, R.C. *Introduction to Strong Mixing Conditions, Vol. 3*; Kendrick Press: Heber City, UT, USA, 2007; p. xii+597.
48. Leucht, A.; Neumann, M.H. Degenerate U- and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.* **2013**, *65*, 349–386. [CrossRef]
49. Neumann, M.H. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **2011**, *17*, 1268–1284. [CrossRef]
50. Chaouch, M.; Laïb, N. Regression estimation for continuous-time functional data processes with missing at random response. *J. Nonparametr. Stat.* **2025**, *37*, 1–32. [CrossRef]
51. Giraitis, L.; Leipus, R. A generalized fractionally differencing approach in long-memory modeling. *Liet. Mat. Rink.* **1995**, *35*, 65–81. [CrossRef]
52. Guégan, D.; Ladoucette, S. Non-mixing properties of long memory processes. *C. R. Acad. Sci. Paris Sér. I Math.* **2001**, *333*, 373–376. [CrossRef]
53. Andrews, D.W.K. Non-strong mixing autoregressive processes. *J. Appl. Probab.* **1984**, *21*, 930–934. [CrossRef]
54. Francq, C.; Zakoïan, J.M. *GARCH Models*; Structure, Statistical Inference and Financial Applications; John Wiley & Sons, Ltd.: Chichester, UK, 2010; p. xiv+489. [CrossRef]
55. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Application*; Academic Press: Cambridge, MA, USA, 1980. [CrossRef]
56. Ayad, S.; Laksaci, A.; Rahmani, S.; Rouane, R. Local linear modelling of the conditional distribution function for functional ergodic data. *Math. Model. Anal.* **2022**, *27*, 360–382. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.