



---

*Research article*

## Limit theorems for local polynomial estimation of regression for functional dependent data

Oussama Bouanani<sup>1</sup> and Salim Bouzebda<sup>2,\*</sup>

<sup>1</sup> Faculty of Exact Sciences and Computer Science, University of Mostaganem, 27000, Algeria, LMSSA

<sup>2</sup> Laboratory of Applied Mathematics of Compiègne (LMAC), Université de technologie de Compiègne, Compiègne, France

\* **Correspondence:** Email: [salim.bouzebda@utc.fr](mailto:salim.bouzebda@utc.fr).

**Abstract:** Local polynomial fitting exhibits numerous compelling statistical properties, particularly within the intricate realm of multivariate analysis. However, as functional data analysis gains prominence as a dynamic and pertinent field in data science, the exigency arises for the formulation of a specialized theory tailored to local polynomial fitting. We explored the intricate task of estimating the regression function operator and its partial derivatives for stationary mixing random processes, denoted as  $(Y_i, X_i)$ , using local higher-order polynomial fitting. Our key contributions include establishing the joint asymptotic normality of the estimates for both the regression function and its partial derivatives, specifically in the context of strongly mixing processes. Additionally, we provide explicit expressions for the bias and the variance-covariance matrix of the asymptotic distribution. Demonstrating uniform strong consistency over compact subsets, along with delineating the rates of convergence, we substantiated these results for both the regression function and its partial derivatives. Importantly, these findings rooted in reasonably broad conditions that underpinned the underlying models. To demonstrate practical applicability, we leveraged our results to compute pointwise confidence regions. Finally, we extended our ideas to the nonparametric conditional distribution, and obtained its limiting distribution.

**Keywords:** functional data analysis; Kernel method; local linear estimation method; local polynomial estimation method; uniform almost complete convergence; normality

**Mathematics Subject Classification:** 60F05, 60F15, 60F17, 62G07

---

## 1. Introduction

In recent decades, there has been a growing interest in the statistical study of functional random variables, which are values in infinite-dimensional spaces. Simply put, functional data analysis (FDA) typically deals with statistical problems where the data consists of a sample of  $n$  functions  $x_1 = x_1(t), \dots, x_n = x_n(t)$ , defined on a compact interval of the real line  $[0, 1]$ . FDA focuses on statistical issues, often referred to as “inference in stochastic processes”, where sample information is derived from a partial trajectory  $(x(t), t \in [0, T])$  of a stochastic process  $(\{X(t), t \geq 0\})$ . In this context, the duration ( $T$ ) of the observation interval acts as the sample size ( $n$ ). This increasing interest has brought to light several statistical challenges associated with functional random variables. The drive to explore this field has been fueled by the increasing availability of high-resolution temporal and spatial data. This trend is particularly notable in fields such as meteorology, medicine, satellite imaging, and various other scientific disciplines. Consequently, the statistical modeling of this data, viewed as stochastic functions, has led to numerous complex theoretical and computational research questions. To gain a comprehensive understanding of both the theoretical and practical components of functional data analysis, it is recommended that the reader consult the monographs authored by [13] for linear models about random variables that assume values in a Hilbert space, and [74] for scalar-on-function and function-on-function linear models, as well as functional principal component analysis and parametric discriminant analysis. Ferraty and Vieu [38] primarily concentrated on nonparametric techniques, particularly kernel-based estimation, for scalar-on-function nonlinear regression models. These methods were further expanded to encompass classification and discrimination analysis. In their study, [48] examined the extension of various statistical concepts, including goodness-of-fit tests, portmanteau tests, and change detection, to the functional data framework. Zhang [91] conducted a study on the analysis of variance for functional data, while [71] primarily investigated regression analysis for Gaussian processes. Various semi-parametric models have been explored in the literature, such as functional single index models [44], projection pursuit models [27], partial linear models [5], and functional sliced inverse regression [43]. Additional studies on functional data modeling and analysis have been documented in the following sources: [45, 49, 57, 67] and for recent references [1, 2, 15, 17–22, 66, 79]. One of the most used estimators is the Nadaraya-Watson [68, 85] estimator. From a function approximation perspective, the Nadaraya-Watson estimator employs local constant approximations. As highlighted by the numerical analyst [80], “Through all of scientific computing runs this common theme: Increase the accuracy at least to second order. What this means is: Get the linear term right”. In other words, local constant approximations are often inadequate, and local linear fits are preferable. Local polynomial fitting emerges as an attractive method from both theoretical and practical standpoints. Local polynomial fitting offers several advantages, as noted in [37]. It adapts to various design types, including random and fixed designs, highly clustered, and nearly uniform designs. Moreover, boundary effects are absent: the bias at the boundary remains of the same order as in the interior, without requiring specific boundary kernels. This is a significant departure from other existing methods. With local polynomial fitting, no boundary modifications are necessary, which is particularly beneficial in multidimensional situations where the boundary can involve a substantial number of data points (see [28, 78]). Boundary modifications in higher dimensions pose a considerable challenge. The pivotal method proposed by [60] involved utilizing local higher-order polynomial fitting to estimate the function  $m$  and its partial derivatives. The author demonstrated mean-square

convergence and joint asymptotic normality (see also [61]). This research holds significant implications for estimating widely used ARCH time series. Masry [59] established the strong consistency of the regression estimator and its partial derivatives up to a fixed total order  $p$ , obtaining sharp rates of almost sure convergence that are uniform over compact sets. In [63], the local polynomial estimation of regression functions and their derivatives is examined, establishing the joint asymptotic normality of the relevant estimates for both strongly mixing and  $\rho$ -mixing processes. Masry and Mielniczuk [64] considered the nonparametric estimation of a multivariate regression function and its derivatives for a regression model with long-range dependent errors. The authors adopted a local linear fitting approach and established the joint asymptotic distributions for the estimators of the regression function and its derivatives.

The entire body of literature on functional data is primarily centered around the theme of local linear estimation. Rachdi et al. [73] innovatively amalgamated the  $k$ -Nearest Neighbors method with the local linear estimation approach to formulate a novel estimator for the regression operator. This innovative approach was tailored for situations where the regressor adopts a functional form, and the response variable, though scalar, is subject to random missing observations. On a parallel note, [7] undertook the task of estimating the conditional density function through the application of the local linear methodology. Moreover, Chikr-Elmezouar et al. [29] delved into the estimation of the conditional density and mode when confronted with functional covariates. In the realm of local linear regression, [8] presented a novel local linear regression estimator and conducted a comprehensive study of its asymptotic behavior. The work by [9] and [11] introduced a general framework elucidating the local behavior of the regression operator, refer to [3] for recent references. The foundational work of [92] established mean-squared convergence and asymptotic normality for the local linear estimator. Expanding the scope to instances where both the response and explanatory variables are functionally characterized, [32] devised a nonparametric local linear estimator for the regression function. The principal objective of the current investigation is to establish a comprehensive framework for polynomial fitting. The central aim of this paper is to furnish the inaugural comprehensive theoretical validation for the kernel polynomial estimator. This entails determining the uniform consistency rate and discerning the asymptotic distribution. Nonetheless, it becomes evident that addressing this issue 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.

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 3.1, and its application to the confidence regions is presented in Section 3.3. The establishment of uniform almost complete convergence is detailed in Section 3.4. Local linear estimators for the conditional distribution and their asymptotic normality are presented in Section 4. In Section 5, we summarize the findings and highlight remaining open issues. All proofs are deferred to Section 6, with a focus on the most crucial arguments due to the lengthiness of the proofs. Additionally, relevant technical results are provided in the appendix.

## 2. Local polynomial regression

We begin by introducing the necessary notation and definitions for the forthcoming sections. Specifically, we recall the definition of the strong mixing property. Let  $\mathcal{F}_i^k(Z)$  denote the  $\sigma$ -algebra generated by  $\{Z_j : i \leq j \leq k\}$ . The following definition of the strong mixing coefficient is attributed to [76]. For further details, refer to [24, 75, 82, 83].

**Definition 2.1.** Let  $Z = \{Z_i, i = 1, 2, \dots\}$  be a strictly stationary sequence of random variables. Given a positive integer  $n$ , define

$$\alpha(n) = \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^k(Z) \text{ and } B \in \mathcal{F}_{k+n}^\infty(Z), k \in \mathbb{N}^* \right\}.$$

The sequence  $Z$  is said to be  $\alpha$ -mixing (strong mixing) if the mixing coefficient  $\alpha(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

The  $\alpha$ -mixing condition is the weakest among the commonly used mixing conditions and has numerous practical applications, particularly in economics and finance. For example, several time series models, such as ARCH, ARMA, GARMA, and stochastic volatility models, satisfy the  $\alpha$ -mixing condition.

**Remark 2.2.** Asymptotic independence, or “mixing”, of the data-generating process is a common assumption in statistical learning for time series. For many typical processes, the general forms of various mixing rates are often assumed to be known, but specific coefficients are not. These mixing assumptions are rarely tested, and there are no established methods for estimating mixing coefficients from data. In [65], an estimator for beta-mixing coefficients based on a single stationary sample path is introduced. Khaleghi and Lugosi [51] propose strongly consistent estimators for the  $\ell_1$  norm of the sequence of  $\alpha$ -mixing and  $\beta$ -mixing coefficients of a stationary ergodic process. These estimators are subsequently used to develop strongly consistent goodness-of-fit hypothesis tests. Specifically, Khaleghi and Lugosi [51] develop hypothesis tests to determine whether, under the same summability assumption, the  $\alpha$ -mixing or  $\beta$ -mixing coefficients of a process are upper bounded by a given rate function.

We consider a sequence  $\{(X_i, Y_i) : i \geq 1\}$  of stationary\*  $\alpha$ -mixing random copies of the random vector [rv]  $(X, Y)$ , where  $X$  takes its values in some abstract space  $\mathfrak{X}$  and  $Y$  in  $\mathbb{R}$ . Suppose that  $\mathfrak{X}$  is endowed with a semi-metric  $d(\cdot, \cdot)^\dagger$  defining a topology to measure the proximity between two elements of  $\mathfrak{X}$  and which is disconnected from the definition of  $X$  to avoid measurability problems. We will consider especially the conditional expectation of  $\psi(Y)$  given  $X = x$ ,

$$r(x) = r_\psi(x) = \mathbb{E}(\psi(Y) | X = x), \quad (2.1)$$

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

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

\*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 (or 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  $\{X_{t_1}, \dots, X_{t_n}\}$  is the same as the joint distribution of  $\{X_{t_1+h}, \dots, X_{t_n+h}\}$  for all  $t_1, \dots, t_n \in \mathbb{Z}$ ,  $n \geq 1$ , and  $h \geq 1$ .

†A semi-metric (sometimes called pseudo-metric)  $d(\cdot, \cdot)$  is a metric which allows  $d(x_1, x_2) = 0$  for some  $x_1 \neq x_2$ .

In this work, we study the problem of nonparametric estimation of the regressor function  $r(\cdot)$  using, for the first time in the functional data, the local polynomial fitting such that the regressor belongs to an infinite dimensional set. From now on, for the ease of the 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 (see [9]) the linear approximation of  $r(\cdot)$  for any  $z$  in the neighborhood of  $x$  is given by

$$r(z) = r(x) + b\beta(z, x) + o(\beta(z, x)),$$

where the quantities  $a = r(x)$  and  $b$  are estimated by minimizing the following criterion

$$\min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n (Y_i - a - b\beta(X_i, x))^2 K(h_K^{-1} \delta(x, X_i)), \quad (2.2)$$

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  $\beta$  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. 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  $d(\cdot, \cdot)$  is a distance metric.  $K(\cdot)$  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:

$$\text{loc}_a^{(q)}(x_1, x_2) = \int \theta(t) (x_1^{(q)}(t) - x_2^{(q)}(t)) dt = \langle \theta, x_1^{(q)} - x_2^{(q)} \rangle_{\mathcal{X}},$$

where  $\theta$  is a function that can be adapted to the data,  $\mathcal{X}$  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:

$$\widehat{c}(x) = \arg \min_{c(x) \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left( Y_i - \sum_{l=0}^p c_l(x) \langle \theta, X_i^{(q)} - x^{(q)} \rangle_{\mathcal{X}}^l \right)^2 K(h^{-1} |\delta(X_i, x)|),$$

which can be viewed as a type of ‘local polynomial’ 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:

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

which is a semi-metric. This second family of locating functions is particularly well-suited for  $\delta(\cdot, \cdot)$ , as it measures the proximity between two elements of  $\mathcal{X}$ . Let  $U$  be an open subset of a real Banach space  $\mathfrak{F}$ . If  $f : U \rightarrow \mathbb{R}$  is differentiable  $p + 1$  times on  $U$ , it can be expanded using Taylor’s formula:

$$f(x) = f(a) + Df(a) \cdot h + \frac{1}{2!} D^2 f(a) \cdot h^2 + \cdots + \frac{1}{n!} D^n f(a) \cdot h^n + R_p(x),$$

with the following expressions for the remainder term  $R_p(x)$  :

$$\begin{aligned} R_p(x) &= \frac{1}{p!} D^{p+1} f(\eta) \cdot (x - \eta)^p h \quad \text{Cauchy form of remainder,} \\ R_p(x) &= \frac{1}{(p+1)!} D^{p+1} f(\xi) \cdot h^{p+1} \quad \text{Lagrange form of remainder,} \\ R_p(x) &= \frac{1}{p!} \int_0^1 D^{p+1} f(a + th) \cdot ((1-t)h)^p h dt \quad \text{integral form of remainder.} \end{aligned}$$

Here  $a$  and  $x$  must be points of  $U$  such that the line segment between  $a$  and  $x$  lie inside  $U$ ,  $h$  is  $x - a$ , and the points  $\xi$  and  $\eta$  lie on the same line segment, strictly between  $a$  and  $x$ . If we collect the equal mixed partials (assuming that they are continuous) then

$$\frac{1}{k!} D^k f(a) \cdot h^k = \sum_{|J|=k} \frac{1}{J!} \frac{\partial^J f}{\partial x^J} h^J,$$

where  $J$  is a multi-index of  $m$  components, and each component  $J_i$  indicates how many times the derivative with respect to the  $i$ th coordinate should be taken, and the exponent that the  $i$ th coordinate of  $h$  should be raised to in the monomial  $h^J$ . The multi-index  $J$  runs through all combinations such that  $J_1 + \dots + J_m = |J| = k$  in the sum. The notation  $J!$  means  $J_1! \dots J_m!$ . Let us specify our setting as in [11]. Let  $U$  be an open subset of  $\mathfrak{F}$ . If  $f : U \rightarrow \mathbb{R}$  is  $p + 1$  times differentiable on  $U$ , the Taylor expansion of  $f$  around  $x \in U$  is

$$f(y) = \sum_{j=0}^p \frac{1}{j!} D^j f(x) d^j(x, y) + R_p(y),$$

where  $D^j f(x)$  denote the  $j$ th Fréchet derivative of  $f$  at  $x$ , with  $x$  and  $y$  being points of  $U$  and  $d(\cdot, \cdot)$  the metric of  $\mathfrak{F}$ . The remainder is given by

$$R_p(y) = \frac{1}{p!} \int_0^1 D^{p+1} f\{x + td(x, y)\} d^{p+1}(x, y) (1-t)^p dt = o\{d^p(x, y)\},$$

as  $d(x, y) \rightarrow 0$ ; see [26] and [90] for more details. Then, in our setting, by following a similar reasoning as in [9] and [11], the local polynomial estimator of order  $p$  of the regression function denoted by  $\widehat{r}(x)$  of  $r(x)$  is defined as the first component of the estimate  $\widehat{c}(x) = (\widehat{c}_0(x), \dots, \widehat{c}_p(x))$  which is obtained by the following minimization problem

$$\widehat{c}(x) = \arg \min_{c(x) \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left( Y_i - \sum_{l=0}^p c_l(x) \beta^l(X_i, x) \right)^2 K(h_k^{-1} \delta(x, X_i)). \quad (2.3)$$

Furthermore, the estimator  $\widehat{r}^l(x) = l! \widehat{c}_l(x)$ , where  $0 \leq l \leq p$  can be expressed similarly as an estimator of the existing  $l$ th order derivatives of the regression function  $r(\cdot)$ . Then the estimator  $\widehat{r}^l(x)$  can be written as

$$\widehat{c}(x) = \left( {}^\top \mathbf{Q}_\beta \mathbf{W} \mathbf{Q}_\beta \right)^{-1} {}^\top \mathbf{Q}_\beta \mathbf{W} \mathbf{Y}, \quad (2.4)$$

where  ${}^\top \mathbf{Q}_\beta$  is the matrix defined by

$${}^{\top}\mathbf{Q}_{\beta} = \begin{pmatrix} 1 & \cdots & 1 \\ \beta(X_1, x) & \cdots & \beta(X_n, x) \\ \vdots & \cdots & \vdots \\ \beta^p(X_1, x) & \cdots & \beta^p(X_n, x) \end{pmatrix}, \quad Y = {}^{\top}(Y_1, \dots, Y_n), \quad \text{and} \quad \mathbf{W} = \text{diag}(K(h_K^{-1}\delta(X_i, x))),$$

where  ${}^{\top}$  is the transpose symbol. Let us introduce

$$u_{n,j} = \frac{1}{n\mathbb{E}(K)} \sum_{i=1}^n \left( \frac{\beta(X_i, x)}{h_K} \right)^j K(h_K^{-1}(\delta(X_i, x))), \quad v_{n,j} = \frac{1}{n\mathbb{E}(K)} \sum_{i=1}^n \left( \frac{\beta(X_i, x)}{h_K} \right)^j K(h_K^{-1}(\delta(X_i, x)))Y_i,$$

$$\mathbf{U}_n = \begin{pmatrix} u_{n,0} & \cdots & u_{n,p} \\ \vdots & \ddots & \vdots \\ u_{n,p} & \cdots & u_{n,2p} \end{pmatrix}, \quad \text{and} \quad \mathbf{V}_n = {}^{\top}(v_{n,0}, \dots, v_{n,p}).$$

Keeping in mind this notation, we have the representation

$$\widehat{c}(x) = \text{diag}(1, h_K^{-1}, h_K^{-2}, \dots, h_K^{-p}) \mathbf{U}_n^{-1} \mathbf{V}_n.$$

Interestingly, this class of estimators includes both the classical Nadaraya-Watson estimator, which minimizes (2.3) when  $p = 0$ , and the local linear kernel estimator, which corresponds to  $p = 1$ .

### 3. Main results for the regression operator

Let  $x$  (resp.  $y$ ) be a fixed point in  $\mathfrak{F}$  (resp.  $\in \mathbb{R}$ ), and  $N_x$  (resp.  $N_y$ ) denote a fixed neighborhood of  $x$  (resp.  $y$ ).

#### 3.1. Asymptotic distribution

In the remainder of the paper, we denote the closed ball in  $\mathfrak{F}$  of center  $x \in \mathfrak{F}$  and radius  $h$  by

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

We define

$$\phi_x(l_1, l_2) = \mathbb{P}(l_2 \leq \delta(X, x) \leq l_1),$$

where  $l_1$  and  $l_2$  are two real numbers. To establish the asymptotic behavior of the local polynomial of our estimator, we need some following assumptions.

(H1) For any  $u > 0$ ,  $\phi_x(u) := \phi_x(-u, u) > 0$  where  $\lim_{u \rightarrow 0} \phi_x(u) = 0$ .

It is easy to see that

$$\varphi_x(u) = \mathbb{P}(X \in B(x, u)).$$

When the function  $|\delta(\cdot, \cdot)|$  satisfies the conditions of a metric or semi-metric, the expression  $\varphi_x(u)$  can be understood as the probability of a ball in the set  $\mathfrak{F}$  that is centered at  $x$  and has a radius of  $u$ . When the value of  $u$  approaches zero, the term ‘‘small ball probabilities’’ is frequently employed in the field of probability theory, which has been extensively studied (refer to [54] for a comprehensive summary of this topic in the context of Gaussian processes). The function  $\varphi_x(\cdot)$ , which is an extension

of the small ball probability concept, serves a similar purpose in the functional situation as the density function does in the finite-dimensional context. In the context of multivariate nonparametric analysis, it is customary to estimate a specific quantity at a given position using a sufficiently large number of observations available in the vicinity. One commonly employed method for assuming this nature is to claim that the density function evaluated at this particular location possesses a strictly positive value. In the context of infinite dimensions, the absence of a reference measure, such as the Lebesgue measure in finite dimensions, necessitates the adoption of a similar assumption that does not rely on the concept of density. The objective of Hypothesis 1 (H1) is to incorporate the functional aspect, indicating that there are a sufficient number of observations surrounding  $x$ , hence justifying the estimation of the regression operator at the specific point  $x$ . The following conditions are needed in our analysis.

(H2)  $(X_i, Y_i)_{i \in \mathbb{N}}$  is an  $\alpha$ -mixing sequence and

- (i)  $\exists a > 0, \exists C > 0, \forall n \in \mathbb{N}, \alpha(n) \leq C n^{-a}$ ,
- (ii)  $\sup_{i \neq j} \mathbb{P}((X_i, X_j) \in B(x, h_K) \times B(x, h_K)) \leq \psi_x(h_K)$ , where  $\psi_x(h_K)$  is such that there exists  $\epsilon \in ]0, 1[$  for which  $0 < \psi_x(h_K) = O(\phi_x^{1+\epsilon}(h_K))$ ;

(H3) (i)  $r(\cdot)$  and  $\sigma(\cdot)$  are continuous in the neighbourhood of  $x$ , which means that  $r(\cdot)$  and  $\sigma(\cdot)$  are both in the set

$$\left\{ f : \mathfrak{F} \rightarrow \mathbb{R}, \lim_{|\delta(x', x)| \rightarrow 0} f(x') = f(x) \right\},$$

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

$$\sup_{i \neq j} \mathbb{E}(|Y_i Y_j| | (X_i, X_j)) \leq C < \infty,$$

- (ii) For any  $k \in \{1, 2, \dots, p, p + 1\}$ , the quantities  $\Psi^{(k)}(0)$  exist, where  $\Psi^{(k)}(\cdot)$  denotes the  $k$ th derivative of  $\Psi(\cdot)$  defined by  $\Psi(s) = \mathbb{E}(r(X) - r(x) | \beta(X, x) = s)$ ;

(H4) The locating operators  $\beta(\cdot, \cdot)$  and  $\delta(\cdot, \cdot)$  satisfy the following conditions:

- (i)  $\forall z \in \mathfrak{F}, C_1 |\delta(x, z)| \leq |\beta(x, z)| \leq C_2 |\delta(x, z)|$ ;
- (ii)  $\sup_{v \in B(x, r)} |\beta(v, x) - \delta(x, v)| = o(r)$ ;

(H5) The kernel  $K(\cdot)$  is a bounded and positive function which is supported within  $[-1, 1]$  and for which the first derivative  $K'(\cdot)$  satisfies:  $K(1) > 0, K'(t) < 0$ , for  $t \in [-1, 1]$ ;

(H6) There exists a positive integer  $n_0$  for which

$$-\frac{1}{\phi_x(h_K)} \int_{-1}^1 \phi_x(z h_K, h_K) \frac{d}{dz} (z^2 K(z)) dz > C_3 > 0 \text{ for } n > n_0,$$

and

$$h_K^2 \int_{B(x, h_K)} \int_{B(x, h_K)} \beta(u, x) \beta(t, x) d\mathbb{P}_{(X_1, X_2)}(u, t) = o\left( \int_{B(x, h_K)} \int_{B(x, h_K)} \beta^2(t, x) \beta^2(u, x) d\mathbb{P}_{(X_1, X_2)} \right),$$

where  $d\mathbb{P}_{(X_1, X_2)}$  is the cumulative distribution of  $(X_1, X_2)$ ;

(H7) (i) For  $a > 1$ , the function  $\phi_x(\cdot)$  satisfies

$$\exists \eta > 0, \quad C_2 n^{\frac{1}{1-a}} \leq \phi_x(h_K) \leq C_1 n^{\frac{1}{3-2a} - \eta};$$



(ii) The bandwidths  $h_K$  satisfy

$$\lim_{n \rightarrow \infty} h_K = 0, \quad \lim_{n \rightarrow \infty} \frac{\log n}{n\phi_x(h_K)} = 0 \text{ and } \exists \beta > 0 \text{ such that } \lim_{n \rightarrow \infty} n^\beta h_K = 0.$$

Before presenting our result, we need additional notation. In the sequel, we define certain quantities appearing in the dominant terms of the bias and of the variance in the asymptotic results. For  $j \in \{1, 2\}$ , let

$$M_j = K^j(1) - \int_{-1}^1 (K^j(u))' \chi_x(u) du, \quad (3.1)$$

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

$$\lim_{h_K \rightarrow 0} \frac{\phi_x(-h_K, th_K)}{\phi_x(h_K)} = \chi_x(t) \text{ for any } t \in [-1, 1].$$

For all  $a > 0$ , and  $b > 0$ , let

$$N(a, b) = K^a(1) - \int_{-1}^1 (u^b K^a(u))' \chi_x(u) du. \quad (3.2)$$

### 3.2. Comments on the conditions

Assumption (H2-i) requires that the mixing coefficients of the dependent case tend to zero at a suitably mild rate. Assumption (H2-ii) describes the behavior of the joint distribution of the pair  $(X_i, X_j)$  in relation to its marginal distributions, allowing us to present an explicit asymptotic variance term. Assumption (H3-i) imposes the usual moment condition on the responses and the covariance structure of the dependent sample, as detailed in [38]. Assumption (H3-ii) specifies the necessary smoothness condition for the current setting. Assumption (H4) is a standard condition in nonparametric estimation. Condition (H5) is very usual in nonparametric estimation literature devoted to the functional data context. Notice that [70] symmetric kernel is not adequate in this context since the random process  $d(x, X_i)$  is positive, therefore we consider  $K(\cdot)$  with support  $[0, 1]$ . This is a natural generalization of the assumption usually made on the kernel in the multivariate case where  $K(\cdot)$  is supposed to be a spherically symmetric density function. Assumption (H6) describes the behavior of the bandwidth  $h$  in relation to the small ball probabilities and the kernel function  $K(\cdot)$ . Assumption (H7-i) is a technical condition illustrating the relationship between the small ball probability and an arithmetically  $\alpha$ -mixing coefficient, as discussed in [42]. Assumption (H7-ii) is satisfied when  $h_K = n^{-\varrho}$  and  $\phi_x(h_K) = h_K^\beta$ , provided  $\vartheta > 0$  and  $\beta < \varrho < 1/\vartheta$ . In the following, we use the notation  $Z \stackrel{D}{=} \mathcal{N}(\mu, \sigma^2)$  to indicate that the random variable  $Z$  follows a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . The symbol  $\xrightarrow{D}$  denotes convergence in distribution, and  $\xrightarrow{\mathbb{P}}$  denotes convergence in probability.

**Remark 3.1.** According to [40], our methodology is heavily dependent on the distribution function  $\phi(\cdot)$ . This dependency is evident in our conditions and the convergence rates of our estimate (via the asymptotic behavior of the quantity  $n\phi(h)$ ). More precisely, the behavior of  $\phi(\cdot)$  around 0 is of paramount importance. Consequently, the small ball probabilities of the underlying functional variable  $X$  are crucial. In probability theory, the calculation of  $\mathbb{P}(\|X - x\| < s)$  for “small”  $s$  (i.e., for  $s$  tending toward zero) and for a fixed  $x$  is known as the “small ball problem”. Unfortunately, there are solutions

for very few random variables (or processes)  $X$ , even when  $x = 0$ . In several functional spaces, taking  $x \neq 0$  presents formidable obstacles that may be insurmountable. Typically, authors emphasize Gaussian random variables. We refer you to [54] for a summary of the key findings regarding the probability of small balls. If  $X$  is a Gaussian random element on the separable Banach space  $\mathcal{E}$  and  $x$  belongs to the reproducing kernel Hilbert space associated with  $X$ , then the following well-known result holds:

$$\mathbb{P}(\|X - x\| \leq s) \sim C_x \mathbb{P}(\|X\| \leq s), \quad \text{as } s \rightarrow 0.$$

As far as we know, the results available in the published literature are essentially all of the form

$$\mathbb{P}(\|X - x\| < s) \sim c_x s^{-\alpha} \exp(-C/s^\beta),$$

where  $\alpha, \beta, c_x$ , and  $C$  are positive constants, and  $\|\cdot\|$  may be a supremum norm, an  $L^p$  norm, or a Besov norm. The interested reader can refer to [12, 16, 20, 38–40, 79] for further discussion. Notably, the pioneering book by [38] extensively comments on the links between nonparametric functional statistics, small-ball probability theory, and the topological structure of the functional space  $\mathcal{E}$ .

In the following theorem, we present the limiting law.

**Theorem 3.2.** Assume that the assumptions (H1)–(H7) are satisfied. We have, as  $n \rightarrow \infty$ ,

$$\sqrt{n\phi_x(h_K)} \left\{ \text{diag}(1, h_K, \dots, h_K^p) \begin{pmatrix} \widehat{r}(x) - r(x) \\ \widehat{\Psi}^{(1)}(0) - \Psi^{(1)}(0) \\ \vdots \\ \widehat{\Psi}^{(p)}(0) - \Psi^{(p)}(0) \end{pmatrix} - \frac{h_K^{p+1}}{(p+1)!} \Psi^{(p+1)}(0) S^{-1} U \right\} \xrightarrow{D} \mathcal{N}(0, \sigma^2(x) S^{-1} V S^{-1}),$$

where

$$S = \begin{pmatrix} 1 & \dots & \frac{N(1, p)}{M_1} \\ \vdots & \ddots & \vdots \\ \frac{N(1, p)}{M_1} & \dots & \frac{N(1, 2p)}{M_1} \end{pmatrix},$$

$$V = \text{diag} \left( \frac{M_2}{M_1^2}, \frac{N(2, 2)}{M_1^2}, \dots, \frac{N(2, 2p)}{M_1^2} \right),$$

$$U = \begin{pmatrix} \frac{N(1, p+1)}{M_1} \\ \vdots \\ \frac{N(1, 2p+1)}{M_1} \end{pmatrix},$$

where  $\widehat{c}_p(x)$  is a consistent estimator of  $\Psi^{(p)}(0)$ . Henceforth, we will denote  $\widehat{c}_p(x)$  as  $\widehat{\Psi}^{(p)}(0)$ , and let  $\widehat{r}(x) = \widehat{c}_0(x)$ .

The proof of Theorem 3.2 is postponed until Section 6. To eliminate the bias term, we need the following additional assumption.

(H8) Assume that  $\lim_{n \rightarrow \infty} nh_K^{2p+2} \phi_x(h_K) = 0$ .

**Corollary 3.3.** Assume that the conditions (H1)–(H8) are satisfied. We have, as  $n \rightarrow \infty$ ,

$$\sqrt{n\phi_x(h_K)} \left\{ \text{diag}(1, h_K \cdots h_K^p) \begin{pmatrix} \widehat{r}(x) - r(x) \\ \widehat{\Psi}^{(1)}(0) - \Psi^{(1)}(0) \\ \vdots \\ \widehat{\Psi}^{(p)}(0) - \Psi^{(p)}(0) \end{pmatrix} \right\} \xrightarrow{D} \mathcal{N}(0, \sigma^2(x)S^{-1}VS^{-1}).$$

**Remark 3.4.** To cancel the bias term, we need  $nh_K^{2p+2} \phi_x(h_K) \rightarrow 0$ , as  $n \rightarrow \infty$ . Consequently, the last condition and condition  $n\phi_x(h_K) \rightarrow \infty$  are satisfied as soon as  $h_K = n^{-\xi}$ ,  $0 < \xi < 1$ , with and  $\phi_x(h_K) = h_K^c$ , for  $\frac{1}{\xi} - (2p - 2) < c < \frac{1}{\xi}$ .

### 3.3. Application to the confidence regions

A lower  $\alpha$  th quantile of the vector  $V_n$  is any quantity  $\tau_{n\alpha} \in \mathbb{R}^{p+1}$  satisfying  $\tau_{n\alpha} = \inf \{\varepsilon : \mathbb{P}(V_n \leq \varepsilon) \geq \alpha\}$ , where  $\varepsilon$  is an infimum over the given set only if there does not exist a  $\varepsilon_1 < \varepsilon$  in  $\mathbb{R}^{p+1}$  such that  $\mathbb{P}(V_n \leq \varepsilon_1) \geq \alpha$ . We can, without loss of generality, assume  $\mathbb{P}(V_n \leq \tau_{n\alpha}) = \alpha$ . Since variance matrix  $\Sigma(x) = \sigma^2(x)S^{-1}VS^{-1}$  is unknown and assumed non-singular, see Corollary 4 of Appendix 1 of [4], there exists a non-singular matrix  $C(x)$  such that

$$C(x)\Sigma(x)^{-1}C(x) = I_{p+1}.$$

Let us first give a consistent estimate of  $\Sigma(x)$ . Making use of the decomposition of  $\chi_x(t)$  in (3.1), one may estimate  $\chi_x(t)$  by

$$\widehat{\chi}_x(t) = \frac{F_{x,n,1}(-h_k, th_k)}{F_{x,n,2}(h_K)},$$

where

$$F_{x,n,1}(t, u) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{t \leq \delta(x, X_i) \leq u\}} \quad \text{and} \quad F_{x,n,2}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{|\delta(x, X_i)| \leq t\}}.$$

Hence we have the following estimates, for  $j = 1, 2$ ,

$$\widehat{M}_j = K^j(1) - \int_{-1}^1 (K^j(u))' \widehat{\chi}_x(u) du. \quad (3.3)$$

In a similar way, for all  $a > 0$ , and  $b > 0$ , we have

$$\widehat{N}(a, b) = K^a(1) - \int_{-1}^1 (u^b K^a(u))' \widehat{\chi}_x(u) du. \quad (3.4)$$

Hence one can consistently estimate  $\Sigma(x)$  by  $\widehat{\Sigma}(x) = \widehat{\sigma}^2(x) \widehat{S}^{-1} \widehat{V} \widehat{S}^{-1}$ , where

$$\widehat{S} = \begin{pmatrix} 1 & \cdots & \frac{\widehat{N}(1, p)}{\widehat{M}_1} \\ \vdots & \ddots & \vdots \\ \frac{\widehat{N}(1, p)}{\widehat{M}_1} & \cdots & \frac{\widehat{N}(1, 2p)}{\widehat{M}_1} \end{pmatrix},$$

$$V = \text{diag} \left( \frac{\widehat{M}_2}{\widehat{M}_1^2}, \frac{\widehat{N}(2, 2)}{\widehat{M}_1^2}, \dots, \frac{\widehat{N}(2, 2p)}{\widehat{M}_1^2} \right),$$

$$U = \begin{pmatrix} \frac{\widehat{N}(1, p+1)}{\widehat{M}_1} \\ \vdots \\ \frac{\widehat{N}(1, 2p+1)}{\widehat{M}_1} \end{pmatrix}.$$

For  $n$  large enough, we have

$$\widehat{C}(x)\widehat{\Sigma}(x)^{-1}\widehat{C}(x) = I_{p+1}.$$

An application of Slutsky with Corollary 3.3 gives the following result.

**Corollary 3.5.** *Assume that the conditions (H1)–(H8) are satisfied. We have, as  $n \rightarrow \infty$ ,*

$$\sqrt{n\phi_x(h_K)}\widehat{C}(x) \left\{ \text{diag}(1, h_K \cdots h_K^p) \begin{pmatrix} \widehat{r}(x) - r(x) \\ \widehat{\Psi}^{(1)}(0) - \Psi^{(1)}(0) \\ \vdots \\ \widehat{\Psi}^{(p)}(0) - \Psi^{(p)}(0) \end{pmatrix} \right\} \xrightarrow{D} \mathcal{N}(0, I_{p+1}).$$

This corollary can be applied directly to the construction of the confidence regions.

#### 3.4. Uniform almost-complete convergence

In this part, we will establish the rate of uniform almost complete convergence over some subset of  $\mathcal{F}$ . The following definition is needed.

**Definition 3.6.** *Let  $\epsilon > 0$  be given number. A finite set of points  $x_1, x_2, \dots, x_N$  in  $\mathcal{F}$  is called an  $\epsilon$ -net for  $S$  if  $S \subset \bigcup_{k=1}^N B(x_k, \epsilon)$ . The quantity  $\Psi_s(\epsilon) = \log(N_\epsilon(S))$ , where  $\Psi_s(\epsilon)$  is the minimal number of open balls in  $\mathcal{F}$  of radius  $\epsilon$  which is necessary to cover  $S$ , is called Kolmogorov's  $\epsilon$ -entropy of the set  $S$ .*

The concept of entropy, introduced by Kolmogorov in the mid-1950s (see [52]), serves as a measure of set complexity, indicating that high entropy implies a significant amount of information is required to accurately describe an element within a given tolerance  $\epsilon$ . Consequently, the selection of the topological structure, specifically the choice of the semi-metric, plays a crucial role when examining asymptotic results that are uniform over a subset  $S$  of  $\mathcal{F}$ . In particular, we subsequently observe that a well-chosen semi-metric can enhance the concentration of the probability measure for the functional variable  $X$ , while minimizing the  $\epsilon$ -entropy of the subset  $S$ . Ferraty and Vieu [38] emphasized this phenomenon of concentration of the probability measure for the functional variable by calculating small ball probabilities in different standard scenarios (refer to [41]). For readers interested in these concepts (entropy and small ball probabilities) and/or the utilization of Kolmogorov's  $\epsilon$ -entropy in dimensionality reduction problems, we recommend referring to [53] or/and [69], respectively. We establish the uniform almost-complete convergence of our estimator on some subset  $S_{\mathcal{F}}$  such that:

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

where  $x_k \in \mathcal{F}$  and  $\iota_n$  (resp.  $d_n$ ) is a sequence of positive real numbers. To establish our result, we need the following hypotheses.

(U1) The hypothesis (H1) satisfies the following condition: there exists a differentiable function  $\phi(\cdot)$ , such that

$$\forall x \in \mathcal{F}, \quad 0 < C_1 \phi(h_k) \leq \mathbb{P}(X \in B(x, h_k)) \leq C_2 \phi(h_k),$$

and there exists  $n_0$  such that for each  $n < n_0$

$$\phi'(n) < C,$$

where  $\phi'(\cdot)$  represents the first derivative of  $\phi(\cdot)$  and  $\phi(0) = 0$ ;

(U2) The hypothesis (H2) is satisfied uniformly on  $x \in S_{\mathcal{F}}$ ;

(U3) The function  $\beta(\cdot, \cdot)$  satisfies (H4) and the following Lipschitz condition, there exists a positive constant  $C$ , for all  $x_1, x_2$  in  $\mathcal{F}$ , we have

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

(U4) The regression operator  $r(\cdot)$  satisfies: there exists a positive constant  $C$  and  $s > 0$ , such for each  $x \in \mathcal{F}$  and  $z \in B(x, h_K)$ , we have

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

(U5) The kernel function  $K(\cdot)$  satisfies (H5) and is Lipschitzian on  $[-1, 1]$ ;

(U6) In addition to (H7), the Kolmogorov's  $\varepsilon$ -entropy of  $S_{\mathcal{F}}$  satisfies:

- (i)  $\exists n_0, \forall n > n_0, \frac{\log^2 n}{n\phi(h)} < \psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) < \frac{n\phi(h)}{\log n}$ .
- (ii)  $\exists \lambda > 1$ , such that

$$\sum_{n=1}^{\infty} \exp\left\{(1 - \lambda)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} = \sum_{n=1}^{\infty} d_n^{1-\lambda} < \infty.$$

### 3.5. Comments on the conditions

Conditions (U1) and (U2) correspondingly serve as the uniform counterparts of (H1) and (H2). Assumption (U3), initially introduced and discussed by [9], plays a pivotal role in our methodology, particularly when computing the leading dominant terms in our asymptotic results. Moreover, (U4) proves essential for evaluating the bias component of the uniform convergence rates, and (U5) is a classical requirement in functional estimation within finite or infinite-dimensional spaces, with examples of kernel functions satisfying this condition available in [38]. The final condition regarding entropy, (U6), implies that  $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) = o(n\phi(h_K))$  as  $n$  tends to infinity. However, in certain "usual" scenarios, such as a Hilbert space with a projection semi-metric, one can consider  $\Psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) \sim C \log n$ , and (H6 i) is fulfilled whenever  $(\log n)^2 = O(n\phi(h))$ . For further insights, one may refer to Example 4 in [41].

We are now equipped to state the main result of this section concerning the uniform almost complete convergence with rate.

**Theorem 3.7.** Assume that the assumptions (U1)–(U6) are satisfied. We have as  $n \rightarrow \infty$ ,

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} \|\widehat{c}(x) - c(x)\|_{\mathcal{F}} = O(h_K^{p+1}) + O_{a.co.} \left( \frac{\log d_n}{n\phi(h_K)} \right),$$

where

$$\widehat{c}(x) = \left( \widehat{r}(x), h_K \widehat{\Psi}^{(1)}(0), \dots, h_K^p \widehat{\Psi}^{(p)}(0) \right).$$

#### 4. The conditional distribution function

In this section, we will establish results analogous to Theorem 3.2 for the local polynomial estimator of order  $p$  of the Cumulative distribution function which we will denote by  $\widehat{F}^x(\cdot)$ . We notice that the distribution function is expressed in terms of regression depending on the choice of the function  $\psi(Y)$  (see Eq (2.1)). Then if  $\psi_y(Y) = \mathbf{1}_{Y \leq y}$ , the distribution function is defined by

$$\forall y \in \mathbb{R} : F^x(y) = \mathbb{P}(Y \leq y | X = x).$$

The functional local polynomial estimator of  $F^x(y)$  is based on the minimization, with respect to  $\widehat{a} = (\widehat{a}_0, \dots, \widehat{a}_p)$ , of

$$\widehat{a} = \arg \min_{(a_0, \dots, a_p) \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left( L(h_L^{-1}(y - Y_i)) - \sum_{l=0}^p a_l(x) \beta^l(X_i, x) \right)^2 K(h_K^{-1} \delta(x, X_i)), \quad (4.1)$$

where  $L(\cdot)$  is cumulative distribution function and  $(h_L = h_{L,n})$  is a sequence of positive real numbers. In the realm of functional statistics, the estimation of the conditional cumulative distribution function holds significant importance both theoretically and practically. This is evident in various applications, such as reliability and survival analysis. Additionally, within the domain of nonparametric statistics, numerous prediction tools, including conditional density, conditional quantiles, and conditional mode, play crucial roles. It is worth noting that these considerations apply particularly when  $p$ , is equal to 1. The almost-complete convergence of the estimator  $\widehat{F}^x(\cdot)$  in the case of  $p = 1$  has been investigated by [31]. Furthermore, [14] has explored the asymptotic normality of the local linear estimation of the conditional function in the scenario of independent observations. For the development of our estimator, we adopt a similar approach as used for the regression operator. Consequently, the local polynomial estimator of  $\widehat{F}^x(y)$  can be explicitly expressed as follows:

$$(\widehat{F}^x(y), \widehat{a}_1, \dots, \widehat{a}_p) = \text{diag}(1, h_K^{-1}, h_K^{-2}, \dots, h_K^{-p}) A_n^{-1} B_n,$$

where

$$\mathbf{A}_n = \begin{pmatrix} \Lambda_{n,0} & \cdots & \Lambda_{n,p} \\ \vdots & \ddots & \vdots \\ \Lambda_{n,p} & \cdots & \Lambda_{n,2p} \end{pmatrix}, \quad \text{and} \quad B_n = {}^T (\Upsilon_{n,0}, \dots, \Upsilon_{n,p}).$$

<sup>‡</sup>Let  $(u_n)$  for  $n \in \mathbb{N}$  be a sequence of real random variables. We say that  $(u_n)$  converges almost-completely (a.co.) toward zero if, and only if, for all  $\epsilon > 0$ ,  $\sum_{n=1}^{\infty} \mathbb{P}(|u_n| > \epsilon) < \infty$ . Moreover, we say that the rate of almost-complete convergence of  $(u_n)$  toward zero is of order  $v_n$  (with  $v_n \rightarrow 0$ ), and we write  $u_n = O_{a.co.}(v_n)$  if, and only if, there exists  $\epsilon > 0$  such that  $\sum_{n=1}^{\infty} \mathbb{P}(|u_n| > \epsilon v_n) < \infty$ . This type of convergence implies both almost sure convergence and convergence in probability.

We denote

$$A_{n,j} = \frac{1}{n\mathbb{E}(K)} \sum_{i=1}^n \left( \frac{\beta(X_i, x)}{h_K} \right)^j K(h_K^{-1}(\delta(X_i, x))),$$

$$\Upsilon_{n,j} = \frac{1}{n\mathbb{E}(K)} \sum_{i=1}^n \left( \frac{\beta(X_i, x)}{h_K} \right)^j K(h_K^{-1}(\delta(X_i, x)))L(h_L^{-1}(Y_i - y)).$$

#### 4.1. Main results

To establish the asymptotic convergence of  $\widehat{F}^x(y)$  we need some notation and assumptions. For any  $l \in \{0, 2, \dots, 2p\}$ , we set

$$\varphi_l(\cdot, y) = \frac{\partial^l F(\cdot, y)}{\partial y^l} \text{ and } \psi_l(s) = \mathbb{E}[\varphi_l(X, y) - \varphi_l(x, y) | \beta(X, x) = s].$$

(M1) For any  $k \in \{1, \dots, p+1\}$ , the quantities  $\psi_l^{(k)}(0)$  exist, where  $\psi_l^{(k)}(\cdot)$  denotes the  $k$ th order derivative of  $\psi_l(\cdot)$ .

(M2) The kernel  $K(\cdot)$  satisfies the assumption (H5) and its first derivative  $K^{(1)}(\cdot)$  satisfies:

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

(M3) The kernel function  $L(\cdot)$  is a positive function, symmetric, bounded, integrable, and such that and

$$\int_{\mathbb{R}} z^{2p} L(z) dz < \infty.$$

(M4) The coefficients of  $\alpha$ -mixing sequence  $(X_i, Y_i)_{i \in \mathbb{N}}$  satisfies (H2) and the following condition,

$$\sum_{l=1}^{+\infty} l^\delta (\alpha(l))^{1/p} < \infty \text{ for some } p > 0 \text{ and } \delta > \frac{1}{p}.$$

(M5) The function  $\phi_x(\cdot)$  satisfies the assumption (H1).

(M6) Let  $(w_n)$  and  $(q_n)$  be sequences of positive integers tending to infinity, such that  $(w_n + q_n) \leq n$ , and

$$(i) \ q_n = o\left((n\phi_x(h_K))^{1/2}\right) \text{ and } \lim_{n \rightarrow +\infty} \left(\frac{n}{\phi_x(h_K)}\right)^{1/2} \alpha(q_n) = 0,$$

$$(ii) \ r_n q_n = o\left((n\phi_x(h_K))^{1/2}\right) \text{ and } \lim_{n \rightarrow +\infty} r_n \left(\frac{n}{\phi_x(h_K)}\right)^{1/2} \alpha(q_n) = 0, \text{ where } r_n \text{ is the largest integer, such that } r_n(w_n + q_n) = O(n).$$

**Theorem 4.1.** Under assumptions (M1)–(M6), we have

$$\sqrt{n\phi_x(h_K)} \left\{ \text{diag}(1, h_K, \dots, h_K^p) \begin{pmatrix} \widehat{F}^x(y) - F^x(y) \\ \widehat{a}_1 - \Psi_0^{(1)}(0) \\ \vdots \\ \widehat{a}_p - \Psi_0^{(p)}(0) \end{pmatrix} - B_K^p(x, y) - B_L^p(x, y) \right\} \xrightarrow{D} \mathcal{N}(0, V_{LK}(x, y)),$$

where

$$\begin{aligned}
B_L^p(x, y) &= \sum_{j=0}^p \left[ \frac{\mathbb{E}(\beta_1^j K_1)}{h_K^j \mathbb{E}(K_1)} I + \sum_{k=1}^p \sum_{a=1}^p \frac{h_L^{2k}}{(2k)!} \int t^{2k} L'(t) dt \Psi_k^{(a)}(0) \frac{\mathbb{E}(\beta_1^{j+a} K_1)}{h_K^j \mathbb{E}(K_1)} \right], \\
I &= \left( \frac{h_L^2}{2} \varphi_2(x, y) \int t^2 L'(t) dt, \dots, \frac{h_L^{2p}}{(2p)!} \varphi_{2p}(x, y) \int t^{(2p)} L'(t) dt \right)^T, \\
B_K^p(x, y) &= \frac{h_K^{p+1}}{(p+1)!} \Psi_0^{(p+1)}(0) S^{-1} U, \\
V_{LK}(x, y) &= F^x(y) (1 - F^x(y)) S^{-1} V S^{-1}.
\end{aligned}$$

**Remark 4.2.** Let  $x \in \mathcal{F}$  be a fixed element, and  $y \in \mathbb{R}$ . If we define  $\psi_y(Y) = \mathbf{1}_{[-\infty, y]}(Y)$ , then the operator  $r_\psi(x) = \mathbb{E}(\psi_y(Y) | X = x)$  represents the conditional cumulative distribution function (CDF) of  $Y$  given  $X = x$ , denoted as  $F(y | x) = \mathbb{P}(Y \leq y | X = x)$ . This can be estimated as  $\tilde{F}(y | x) := \widehat{r}_{\psi, n}(x)$ . For a given  $\alpha \in (0, 1)$ , the  $\alpha$ -th order conditional quantile of the distribution of  $Y$  given  $X = x$  is defined as  $q_\alpha(x) = \inf\{y \in \mathbb{R} : F(y | x) \geq \alpha\}$ . If  $F(\cdot | x)$  is strictly increasing and continuous in a neighborhood of  $q_\alpha(x)$ , then  $F(\cdot | x)$  has a unique quantile of order  $\alpha$  at  $q_\alpha(x)$ , i.e.,  $F(q_\alpha(x) | x) = \alpha$ . In such cases:

$$q_\alpha(x) = F^{-1}(\alpha | x) = \inf\{y \in \mathbb{R} : F(y | x) \geq \alpha\},$$

which is uniquely estimated by  $\widehat{q}_{n, \alpha}(x) = \tilde{F}^{-1}(\alpha | x)$ . Conditional quantiles have been extensively studied when the predictor  $X$  is of finite dimension. Since  $F(q_\alpha(x) | x) = \alpha = \tilde{F}(\widehat{q}_{n, \alpha}(x) | x)$ , and  $\tilde{F}(\cdot | x)$  is continuous and strictly increasing, we have, for any  $\epsilon > 0$ , there exists  $\eta(\epsilon) > 0$  such that for all  $y$ :

$$\left| \tilde{F}(y | x) - \widehat{F}_T(q_\alpha(x) | x) \right| \leq \eta(\epsilon) \Rightarrow |y - q_\alpha(x)| \leq \epsilon.$$

This implies that, for any  $\epsilon > 0$ , there exists  $\eta(\epsilon) > 0$  such that:

$$\begin{aligned}
\mathbb{P}\left(\left|\widehat{q}_{n, \alpha}(x) - q_\alpha(x)\right| \geq \eta(\epsilon)\right) &\leq \mathbb{P}\left(\left|\tilde{F}(\widehat{q}_{n, \alpha}(x) | x) - \tilde{F}(q_\alpha(x) | x)\right| \geq \eta(\epsilon)\right) \\
&= \mathbb{P}\left(\left|F(q_\alpha(x) | x) - \tilde{F}(q_\alpha(x) | x)\right| \geq \eta(\epsilon)\right).
\end{aligned}$$

This result may be immediately used to establish the consistency of the quantile estimator. This topic will be investigated in future research.

**Remark 4.3.** Over time, it has become increasingly clear that the performance of kernel estimates deteriorates as the dimensionality of the problem increases. This decline is primarily due to the fact that, in high-dimensional spaces, local neighborhoods often lack sufficient sample observations unless the sample size is extraordinarily large. Consequently, in kernel estimation, computing local averages is not feasible unless the bandwidth is significantly wide. This overarching issue is known as the curse of dimensionality [10]. The paper by [77] provides a comprehensive analysis of the feasibility and challenges of high-dimensional estimation, complete with examples and computations. For recent references, see [6, 15, 16, 35]. To address the difficulties associated with high-dimensional kernel estimation and simplify the process, a wide range of dimension reduction techniques have been developed. These techniques major to explain the main features of a set of sample curves using a small set of uncorrelated variables. One effective solution is to generalize principal component



analysis (PCA) to the context of a continuous-time stochastic process [33]. The asymptotic properties of the estimators of Functional Principal Component Analysis (FPCA) have been extensively studied in the general context of functional variables [30]. Nonparametric methods have been developed to perform FPCA for cases involving a small number of irregularly spaced observations of each sample curve [50, 88]. As in the multivariate case, interpreting principal component scores and loadings is a valuable tool for uncovering relationships among the variables in a functional data set. To avoid misinterpretation of PCA results, a new type of plot, named Structural and Variance Information plots, was recently introduced by [25]. Numerical results suggest that regularization using a smoothness measure often yields more satisfactory outcomes than the FPCA approach. The FPCA method assumes that the top eigenfunctions of the predictors, which are unrelated to the coefficient function or the response, provide a good representation of the coefficient function. While it is generally believed that neither approach will consistently outperform the other, the Reproducing Kernel Hilbert Space (RKHS) method is an interesting alternative that merits attention [58, 89]. Spline smoothing is one of the most popular and powerful techniques in nonparametric regression [36, 46, 47, 84]. Penalized splines have gained widespread use in recent years due to their computational simplicity and their connection to mixed effects models [86]. These methods have also found significant success in Functional Data Analysis (FDA), as evidenced by tools like the R package *refund*. However, there is a consensus that the asymptotic theory of penalized splines is difficult to derive, and many theoretical gaps remain, even in the standard nonparametric regression setting. To our knowledge, there are very few theoretical studies on penalized splines for FDA [87]. Most approaches to functional regression are based on minimizing some  $L_2$ -norm and are therefore sensitive to outliers. Finally, we can mention methods based on the delta sequences [23] and the wavelet methods [34].

## 5. Concluding remarks

Local polynomial fitting exhibits various statistically significant characteristics, particularly in the intricate domain of multivariate analysis. As functional data analysis gains traction in the field of data science, there arises a need for a specialized theory focused on local polynomial fitting. This study employs local higher-order polynomial fitting to address the complex challenge of estimating the regression function operator and its partial derivatives for stationary mixing random processes, represented as  $(Y_i, X_i)$ . For the first time, we have achieved significant progress by demonstrating the joint asymptotic normality of the estimates for the regression function and its partial derivatives, particularly applicable to strongly mixed processes. Additionally, we derive precise formulas for the bias and the variance-covariance matrix of the asymptotic distribution. We establish the uniform strong consistency of the regression function and its partial derivatives across compact subsets, providing a detailed analysis of their convergence rates. Importantly, these conclusions are grounded in general conditions that form the foundation of the underlying models. To illustrate practical utility, we utilize our findings to compute confidence regions for individual points. Furthermore, we extend our concepts to encompass the nonparametric conditional distribution and acquire its limiting distribution. There are multiple avenues for further methodological development. As a prospective direction, relaxing the stationarity assumption and exploring similar uniform limit theorems for local stationary functional ergodic processes would be fruitful. Additionally, considering the functional  $k$ NN local polynomial approach and expectile regression estimators in future investigations could yield alternative

estimators benefiting from the advantages of both methods, as discussed in [20] and referenced in [19]. Investigating the extension of our framework to the censored data setting would also be of interest. In an upcoming investigation, obtaining a precise uniform-in-bandwidth limit law for the proposed estimators will be essential. This outcome would allow the bandwidth to vary across a comprehensive range, ensuring the estimator's consistency and serving as a valuable practical guideline for bandwidth selection in nonparametric functional data analysis. It is crucial to acknowledge that advancing research in this direction poses a more challenging task compared to previous efforts. The primary challenge lies in the need to develop new probabilistic results, such as inequalities and maximal moment inequalities specifically tailored for independent and identically distributed (i.i.d.) samples, as discussed in [81].

## 6. Proof of the main result

This section is devoted to the proof of our main result. The previously presented notation continues to be used in the following. Before studying the limit law of  $\widehat{c}(x)$ , we need to center the vector  $V_n$  as follows. Let

$$u_{n,j}^* = \frac{1}{n\mathbb{E}(K)} \sum_{i=1}^n \left( \frac{\beta_i}{h_K} \right)^j K_i(Y_i - r(x)), \quad (6.1)$$

where  $V_n^* = {}^\top (u_{n,0}^*, \dots, u_{n,p}^*)$ . Making use of assumption (H4) and by the Taylor expansion, we obtain

$$\begin{aligned} \mathbb{E}(r(X_i) \mid \beta_i) &= r(x) + \mathbb{E}(r(X_i) - r(x) \mid \beta_i) \\ &= r(x) + \Psi(\beta_i) \\ &= r(x) + \sum_{k=1}^p \frac{1}{k!} \Psi^{(k)}(0) \beta_i^k + o(\beta_i^p). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \Gamma &:= {}^\top ((r(X_1), \dots, r(X_n))) \\ &= \mathbf{Q}_\beta \begin{pmatrix} r(x) \\ \Psi^{(1)}(0) \\ \vdots \\ \Psi^{(p)}(0) \end{pmatrix} + \frac{1}{(p+1)!} \Psi^{(p+1)}(0) {}^\top (\beta_1^{p+1}, \dots, \beta_n^{p+1}) + {}^\top (o(\beta_1^{p+1}), \dots, o(\beta_n^{p+1})). \end{aligned}$$

Making used of Eq (6.1), we have:

$$\begin{aligned} \mathbf{U}_n^{-1} V_n^* &= \text{diag}(1, h_K, h_K^2, \dots, h_K^p) (\widehat{c}(x) - ({}^\top \mathbf{Q}_\beta \mathbf{W} \mathbf{Q}_\beta)^{-1} \mathbf{Q}_\beta \mathbf{W} \Gamma) \\ &= \text{diag}(1, h_K, h_K^2, \dots, h_K^p) \begin{pmatrix} \widehat{r}(x) - r(x) \\ \widehat{\Psi}^{(1)}(0) - \Psi^{(1)}(0) \\ \vdots \\ \widehat{\Psi}^{(p)}(0) - \Psi^{(p)}(0) \end{pmatrix} \\ &\quad - \frac{h_K^{p+1}}{(p+1)!} \Psi^{(p+1)}(0) U_n^{-1} \begin{pmatrix} u_{n,p+1} \\ \vdots \\ u_{n,2p+1} \end{pmatrix} - o(h_K^{p+1}) U_n^{-1} \begin{pmatrix} u_{n,p+1} \\ \vdots \\ u_{n,2p+1} \end{pmatrix}. \end{aligned} \quad (6.2)$$

The rest of the proof for this theorem relies on the following lemmas, the proofs of which are provided in the appendix.

**Lemma 6.1.** *Under the assumptions (H1)–(H7), we have, as  $n \rightarrow \infty$ ,*

$$\mathbf{U}_n \xrightarrow{\mathbb{P}} S \quad \text{and} \quad U_n^{-1} \begin{pmatrix} u_{n,p+1} \\ \vdots \\ u_{n,2p+1} \end{pmatrix} \xrightarrow{\mathbb{P}} S^{-1} U.$$

Hence, considering Lemma 6.1, Theorem 3.2 will hold if we can furnish the proof.

**Lemma 6.2.** *Under the assumptions of Theorem 3.2, we have, as  $n \rightarrow \infty$ ,*

$$\sqrt{n\phi_x(h_K)} V_n^* \xrightarrow{D} N(0, \sigma^2(x)V).$$

### Author contributions

Oussama Bouannani: Conceptualization, formal analysis, investigation, methodology, validation, writing – original draft, review, edition; Salim Bouzebda: Conceptualization, formal analysis, methodology, validation, writing – original draft, review, edition. All authors have read and approved the final version of the manuscript for publication.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

The authors extend their sincere gratitude to the Editor-in-Chief, the Associate Editor, and the three referees for their invaluable feedback. Their insightful comments have greatly refined and focused the original work, resulting in a markedly improved presentation.

### Conflict of interest

The authors declare no conflict of interest. Salim Bouzebda is the(a) Guest Editor of special issue “Advances in Statistical Inference and Stochastic Processes: Theory and Applications” for AIMS Mathematics. Salim Bouzebda was not involved in the editorial review and the decision to publish this article.

### References

1. I. M. Almanjahie, S. Bouzebda, Z. C. Elmezouar, A. Laksaci, The functional  $k$ NN estimator of the conditional expectile: Uniform consistency in number of neighbors, *Statist. Risk Model.*, **38** (2021), 47–63. <https://doi.org/10.1515/strm-2019-0029>

2. I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Nonparametric estimation of expectile regression in functional dependent data, *J. Nonparametr. Stat.*, **34** (2022), 250–281. <https://doi.org/10.1080/10485252.2022.2027412>
3. I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, The local linear functional  $k$ NN estimator of the conditional expectile: Uniform consistency in number of neighbors, *Metrika*, 2024. <https://doi.org/10.1007/s00184-023-00942-0>
4. T. W. Anderson, *An introduction to multivariate statistical analysis*, 3 Eds., John Wiley & Sons, Inc., 1958.
5. G. Aneiros, P. Vieu, Partial linear modeling with multi-functional covariates, *Comput. Stat.*, **30** (2015), 647–671. <https://doi.org/10.1007/s00180-015-0568-8>
6. S. Attaoui, B. Bentata, S. Bouzebda, A. Laksaci, The strong consistency and asymptotic normality of the kernel estimator type in functional single index model in presence of censored data, *AIMS Mathematics*, **9** (2024), 7340–7371. <https://doi.org/10.3934/math.2024356>
7. S. Ayad, A. Laksaci, S. Rahmani, R. Rouane, On the local linear modelization of the conditional density for functional and ergodic data, *Metron*, **78** (2020), 237–254. <https://doi.org/10.1007/s40300-020-00174-6>
8. A. Baíllo, A. Grané, Local linear regression for functional predictor and scalar response, *J. Multivariate Anal.*, **100** (2009), 102–111. <https://doi.org/10.1016/j.jmva.2008.03.008>
9. J. Barrientos-Marin, F. Ferraty, P. Vieu, Locally modelled regression and functional data, *J. Nonparametr. Stat.*, **22** (2010), 617–632. <https://doi.org/10.1080/10485250903089930>
10. R. E. Bellman, *Adaptive control processes: A guided tour*, Princeton University Press, 1961.
11. K. Benhenni, A. H. Hassan, Y. Su, Local polynomial estimation of regression operators from functional data with correlated errors, *J. Multivariate Anal.*, **170** (2019), 80–94. <https://doi.org/10.1016/j.jmva.2018.10.008>
12. N.-E. Berrahou, S. Bouzebda, L. Douge, Functional uniform-in-bandwidth moderate deviation principle for the local empirical processes involving functional data, *Math. Meth. Stat.*, **33** (2024), 26–69. <https://doi.org/10.3103/S1066530724700030>
13. D. Bosq, Linear processes in function spaces, In: *Lecture notes in statistics*, New York: Springer, 2000. <https://doi.org/10.1007/978-1-4612-1154-9>
14. O. Bouanani, A. Laksaci, M. Rachdi, S. Rahmani, Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics, *Behaviormetrika*, **46** (2019), 199–233. <https://doi.org/10.1007/s41237-018-0057-9>
15. S. Bouzebda, Limit theorems in the nonparametric conditional single-index  $u$ -processes for locally stationary functional random fields under stochastic sampling design, *Mathematics*, **12** (2024), 1996. <https://doi.org/10.3390/math12131996>
16. S. Bouzebda, Weak convergence of the conditional single index  $U$ -statistics for locally stationary functional time series, *AIMS Mathematics*, **9** (2024), 14807–14898. <https://doi.org/10.3934/math.2024720>

17. S. Bouzebda, B. Nemouchi, Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional  $U$ -statistics involving functional data, *J. Nonparametr. Stat.*, **32** (2020), 452–509. <https://doi.org/10.1080/10485252.2020.1759597>
18. S. Bouzebda, B. Nemouchi, Weak-convergence of empirical conditional processes and conditional  $U$ -processes involving functional mixing data, *Stat. Inference Stoch. Process.*, **26** (2022), 33–88. <https://doi.org/10.1007/s11203-022-09276-6>
19. S. Bouzebda, A. Nezzal, Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional  $U$ -statistics involving functional data, *Jpn. J. Stat. Data Sci.*, **5** (2022), 431–533. <https://doi.org/10.1007/s42081-022-00161-3>
20. S. Bouzebda, A. Nezzal, Uniform in number of neighbors consistency and weak convergence of  $k$ NN empirical conditional processes and  $k$ NN conditional  $U$ -processes involving functional mixing data, *AIMS Mathematics*, **9** (2024), 4427–4550. <https://doi.org/10.3934/math.2024218>
21. S. Bouzebda, A. Laksaci, M. Mohammedi, Single index regression model for functional quasi-associated time series data, *REVSTAT Stat. J.*, **20** (2022), 605–631. <http://dx.doi.org/10.57805/revstat.v20i5.391>
22. S. Bouzebda, A. Laksaci, M. Mohammedi, The  $k$ -nearest neighbors method in single index regression model for functional quasi-associated time series data, *Rev. Mat. Complut.*, **36** (2023), 361–391. <https://doi.org/10.1007/s13163-022-00436-z>
23. S. Bouzebda, S. Nezzal, T. Zari, Uniform consistency for functional conditional  $u$ -statistics using delta-sequences, *Mathematics*, **11** (2023), 161. <https://doi.org/10.3390/math11010161>
24. R. C. Bradley, *Introduction to strong mixing conditions*, Kendrick Press, 2007.
25. J. Camacho, J. Pico, A. Ferrer, Data understanding with PCA: Structural and variance information plots, *Chemometr. Intell. Lab. Syst.*, **100** (2010), 48–56. <https://doi.org/10.1016/j.chemolab.2009.10.005>
26. H. Cartan, *Calcul différentiel*, 1967.
27. D. Chen, P. Hall, H. Müller, Single and multiple index functional regression models with nonparametric link, *Ann. Statist.*, **39** (2011), 1720–1747. <http://dx.doi.org/10.1214/11-AOS882>
28. M. Y. Cheng, J. Fan, J. S. Marron, On automatic boundary corrections, *Ann. Statist.*, **25** (1997), 1691–1708. <https://doi.org/10.1214/aos/1031594737>
29. Z. Chikr-Elmezouar, I. M. Almanjahie, A. Laksaci, M. Rachdi, FDA: strong consistency of the  $k$ NN local linear estimation of the functional conditional density and mode, *J. Nonparametr. Stat.*, **31** (2019), 175–195. <https://doi.org/10.1080/10485252.2018.1538450>
30. J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference, *J. Multivariate Anal.*, **12** (1982), 136–154. [https://doi.org/10.1016/0047-259X\(82\)90088-4](https://doi.org/10.1016/0047-259X(82)90088-4)
31. J. Demongeot, A. Laksaci, M. Rachdi, S. Rahmani, On the local linear modelization of the conditional distribution for functional data, *Sankhya A*, **76** (2014), 328–355. <https://doi.org/10.1007/s13171-013-0050-z>

32. J. Demongeot, A. Naceri, A. Laksaci, M. Rachdi, Local linear regression modelization when all variables are curves, *Statist. Probab. Lett.*, **121** (2017), 37–44. <https://doi.org/10.1016/j.spl.2016.09.021>
33. J. C. Deville, Méthodes statistiques et numériques de l'analyse harmonique, *Annales de l'INSEE*, **15** (1974), 3–101. <https://doi.org/10.2307/20075177>
34. S. Didi, S. Bouzebda, Wavelet density and regression estimators for continuous time functional stationary and ergodic processes, *Mathematics*, **10** (2022), 4356. <https://doi.org/10.3390/math10224356>
35. Z. C. Elmezouar, F. Alshahrani, I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Strong consistency rate in functional single index expectile model for spatial data, *AIMS Mathematics*, **9** (2024), 5550–5581. <https://doi.org/10.3934/math.2024269>
36. R. L. Eubank, *Nonparametric regression and spline smoothing*, 2 Eds., CRC Press, 2014. <https://doi.org/10.1201/9781482273144>
37. J. Fan, I. Gijbels, *Local polynomial modeling and its applications*, Chapman & Hall, 1996.
38. F. Ferraty, P. Vieu, *Nonparametric functional data analysis*, New York: Springer, 2006. <https://doi.org/10.1007/0-387-36620-2>
39. F. Ferraty, A. Laksaci, P. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, *Stat. Infer. Stoch. Process.*, **9** (2006), 47–76. <https://doi.org/10.1007/s11203-004-3561-3>
40. F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: inference and practical aspects, *Aust. N. Z. J. Stat.*, **49** (2007), 267–286. <https://doi.org/10.1111/j.1467-842X.2007.00480.x>
41. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Rate of uniform consistency for nonparametric estimates with functional variables, *J. Statist. Plann. Inference*, **140** (2010), 335–352. <https://doi.org/10.1016/j.jspi.2009.07.019>
42. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes, *C. R. Acad. Sci. Paris, Ser. I*, **350** (2012), 717–720. <https://doi.org/10.1016/j.crma.2012.07.014>
43. L. Ferré, A. F. Yao, Smoothed functional inverse regression, *Statist. Sinica*, **15** (2005), 665–683.
44. A. Goia, P. Vieu, A partitioned single functional index model, *Comput. Stat.*, **30** (2015), 673–692. <https://doi.org/10.1007/s00180-014-0530-1>
45. A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.*, **146** (2016), 1–6. <https://doi.org/10.1016/j.jmva.2015.12.001>
46. P. J. Green, B. W. Silverman, *Nonparametric regression and generalized linear models*, New York: Chapman and Hall/CRC, 1993. <https://doi.org/10.1201/b15710>
47. C. Gu, *Smoothing spline ANOVA models*, New York: Springer, 2015. <https://doi.org/10.1007/978-1-4614-5369-7>
48. L. Horváth, P. Kokoszka, *Inference for functional data with applications*, New York: Springer, 2014. <https://doi.org/10.1007/978-1-4614-3655-3>

49. L. Horváth, G. Rice, An introduction to functional data analysis and a principal component approach for testing the equality of mean curves, *Rev. Mat. Complut.*, **28** (2015), 505–548. <https://doi.org/10.1007/s13163-015-0169-7>
50. G. M. James, T. J. Hastie, C. A. Sugar, Principal component models for sparse functional data, *Biometrika*, **87** (2000), 587–602. <https://doi.org/10.1093/biomet/87.3.587>
51. A. Khaleghi, G. Lugosi, Inferring the mixing properties of an ergodic process, *arXiv:2106.07054*, 2021. <https://doi.org/10.48550/arXiv.2106.07054>
52. A. N. Kolmogorov, V. M. Tihomirov,  $\varepsilon$ -entropy and  $\varepsilon$ -capacity of sets in function spaces, *Uspehi Mat. Nauk*, **14** (1959), 3–86.
53. J. Kuelbs, W. V. Li, Metric entropy and the small ball problem for Gaussian measures, *J. Funct. Anal.*, **116** (1993), 133–157. <https://doi.org/10.1006/jfan.1993.1107>
54. W. V. Li, Q. M. Shao, Gaussian processes: inequalities, small ball probabilities and applications, *Handbook of Statist.*, **19** (2001), 533–597. [https://doi.org/10.1016/S0169-7161\(01\)19019-X](https://doi.org/10.1016/S0169-7161(01)19019-X)
55. H. Y. Liang, J. I. Baek, Asymptotic normality of conditional density estimation with left-truncated and dependent data, *Stat. Papers*, **57** (2016), 1–20. <https://doi.org/10.1007/s00362-014-0635-1>
56. E. Liebscher, Estimation of the density and the regression function under mixing conditions, *Statist. Decisions*, **19** (2001), 9–26. <https://doi.org/10.1524/strm.2001.19.1.9>
57. N. Ling, P. Vieu, Nonparametric modeling for functional data: Selected survey and tracks for future, *Statistics*, **52** (2018), 934–949. <https://doi.org/10.1080/02331888.2018.1487120>
58. S. Lv, H. Lin, H. Lian, J. Huang, Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space, *Ann. Statist.*, **46** (2018), 781–813. <https://doi.org/10.1214/17-AOS1567>
59. E. Masry, Multivariate local polynomial regression for time series: uniform strong consistency and rates, *J. Time Ser. Anal.*, **17** (1996), 571–599. <https://doi.org/10.1111/j.1467-9892.1996.tb00294.x>
60. E. Masry, Multivariate regression estimation local polynomial fitting for time series, *Stochastic Process. Appl.*, **65** (1996), 81–101. [https://doi.org/10.1016/S0304-4149\(96\)00095-6](https://doi.org/10.1016/S0304-4149(96)00095-6)
61. E. Masry, Multivariate regression estimation of continuous-time processes from sampled data: Local polynomial fitting approach, *IEEE Trans. Inform. Theory*, **45** (1999), 1939–1953. <https://doi.org/10.1109/18.782116>
62. E. Masry, Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stochastic Process. Appl.*, **115** (2005), 155–177. <https://doi.org/10.1016/j.spa.2004.07.006>
63. E. Masry, Local polynomial estimation of regression functions for mixing processes, *Scand. J. Statist.*, **24** (1997), 165–179. <https://doi.org/10.1111/1467-9469.00056>
64. E. Masry, J. Mielniczuk, Local linear regression estimation for time series with long-range dependence, *Stochastic Process. Appl.*, **82** (1999), 173–193. [https://doi.org/10.1016/S0304-4149\(99\)00015-0](https://doi.org/10.1016/S0304-4149(99)00015-0)
65. D. J. McDonald, C. R. Shalizi, M. Schervish, Estimating beta-mixing coefficients via histograms, *Electron. J. Stat.*, **9** (2015), 2855–2883. <http://dx.doi.org/10.1214/15-EJS1094>

66. M. Mohammadi, S. Bouzebda, A. Laksaci, The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data, *J. Multivariate Anal.*, **181** (2021), 104673. <https://doi.org/10.1016/j.jmva.2020.104673>
67. H. Müller, Peter hall, functional data analysis and random objects, *Ann. Statist.*, **44** (2016), 1867–1887. <https://doi.org/10.1214/16-AOS1492>
68. E. A. Nadaraja, On estimating regression, *Theory Probab. Appl.*, **9** (1964), 141–142. <https://doi.org/10.1137/1109020>
69. T. Nicolieris, Y. G. Yatracos, Rates of convergence of estimates, Kolmogorov’s entropy and the dimensionality reduction principle in regression, *Ann. Statist.*, **25** (1997), 2493–2511. <https://doi.org/10.1214/aos/1030741082>
70. E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33** (1962), 1065–1076.
71. J. Q. Shi, T. Choi, *Gaussian process regression analysis for functional data*, New York: Chapman and Hall/CRC, 2011. <https://doi.org/10.1201/b11038>
72. M. Rachdi, A. Laksaci, J. Demongeot, A. Abdali, F. Madani, Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data, *Comput. Statist. Data Anal.*, **73** (2014), 53–68. <https://doi.org/10.1016/j.csda.2013.11.011>
73. M. Rachdi, A. Laksaci, Z. Kaid, A. Benchiha, F. A. Al-Awadhi,  $k$ -nearest neighbors local linear regression for functional and missing data at random, *Statist. Neerlandica*, **75** (2021), 42–65. <https://doi.org/10.1111/stan.12224>
74. J. O. Ramsay, B. W. Silverman, Functional data analysis, In: *Springer series in statistics*, New York: Springer, 2005. <https://doi.org/10.1007/b98888>
75. E. Rio, Théorie asymptotique des processus aléatoires faiblement dépendants, In: *Mathématiques et applications*, Heidelberg: Springer-Verlag Berlin, **31** (2000).
76. M. Rosenblatt, A central limit theorem and a strong mixing condition, *Proc. Natl. Acad. Sci.*, **42** (1956), 43–47. <https://doi.org/10.1073/pnas.42.1.43>
77. D. W. Scott, M. P. Wand, Feasibility of multivariate density estimates, *Biometrika*, **78** (1991), 197–205.
78. B. W. Silverman, *Density estimation for statistics and data analysis*, New York: Routledge, 1998. <https://doi.org/10.1201/9781315140919>
79. I. Soukariéh, S. Bouzebda, Weak convergence of the conditional  $U$ -statistics for locally stationary functional time series, *Stat. Inference Stoch. Process.*, **27** (2024), 227–304. <https://doi.org/10.1007/s11203-023-09305-y>
80. G. Strang, Wavelet transforms versus Fourier transforms, *Bull. Amer. Math. Soc.*, **28** (1993), 288–305.
81. M. Talagrand, Sharper bounds for Gaussian and empirical processes, *Ann. Probab.*, **22** (1994), 28–76.
82. V. A. Volkonskii, Y. A. Rozanov, Some limit theorems for random functions. I, *Theory Probab. Appl.*, **4** (1959), 178–197. <https://doi.org/10.1137/1104015>



83. V. A. Volkonskii, Y. A. Rozanov, Some limit theorems for random functions. II, *Theory Probab. Appl.*, **6** (1961), 186–198. <https://doi.org/10.1137/1106023>
84. G. Wahba, Spline models for observational data, In: *CBMS-NSF Regional conference series in applied mathematics*, Society for industrial and applied mathematics (SIAM), 1990. <https://doi.org/10.1137/1.9781611970128>
85. G. S. Watson, Smooth regression analysis, *Sankhyā*, **26** (1964), 359–372.
86. L. Xiao, Asymptotic theory of penalized splines, *Electron. J. Statist.*, **13** (2019), 747–794. <https://doi.org/10.1214/19-EJS1541>
87. L. Xiao, Asymptotic properties of penalized splines for functional data, *Bernoulli*, **26** (2020), 2847–2875. <https://doi.org/10.3150/20-BEJ1209>
88. F. Yao, H. G. Müller, J. L. Wang, Functional data analysis for sparse longitudinal data, *J. Amer. Statist. Assoc.*, **100** (2005), 577–590. <https://doi.org/10.1198/016214504000001745>
89. M. Yuan, T. T. Cai, A reproducing kernel Hilbert space approach to functional linear regression, *Ann. Statist.*, **38** (2010), 3412–3444. <https://doi.org/10.1214/09-AOS772>
90. E. Zeidler, *Nonlinear functional analysis and its applications. I*, New York: Springer, 1986.
91. J. Zhang, *Analysis of variance for functional data*, New York: Chapman and Hall/CRC, 2013, <https://doi.org/10.1201/b15005>
92. Z. Zhou, Z. Lin, Asymptotic normality of locally modelled regression estimator for functional data, *J. Nonparametr. Stat.*, **28** (2016), 116–131. <https://doi.org/10.1080/10485252.2015.1114112>

## A. Appendix

**Lemma A.1.** (See [92]) Under assumptions (H1), (H2-ii) and (H5)–(H7-i), we have

- (1)  $\mathbb{E}(K_1^j) = M_j \phi_x(h_K)$ , for  $j = 1, 2$ ;
- (2)  $\mathbb{E}(K_1^a \beta_1) = o(h_K \phi_x(h_K))$ , for all  $a > 0$ ;
- (3)  $\mathbb{E}(K_1^a \beta_1^b) = N(a, b) h_K^b \phi_x(h_K) + o(h_K^b \phi_x(h_K))$ , for all  $a > 0$  and  $b = 2, 4$ ;
- (4) For all  $(k, l) \in \mathbb{N}^* \times \mathbb{N}$ ,  $\mathbb{E}(K_1^k |\beta_1|^l) \leq C h_K^l \phi_x(h_K)$ ;
- (5) For all  $(a, c, l, s) \in \mathbb{N}^* \times \mathbb{N}^* \times \mathbb{N} \times \mathbb{N}$ ,  $\mathbb{E}(K_1^a K_2^c |\beta_1|^l |\beta_2|^s) \leq C h_K^{l+s} \psi_x(h_K)$ .

*Proof of Lemma 6.1.* To prove this lemma, we use the sufficient convergence condition in probability. For this, it suffices to prove the following formulas

$$\mathbb{E}(u_{n,0}) = 1, \mathbb{E}(u_{n,1}) \xrightarrow{\mathbb{P}} 0, \mathbb{E}(u_{n,p}) \xrightarrow{\mathbb{P}} \frac{N(1,p)}{M_1}, \mathbb{E}(u_{n,1}) \xrightarrow{\mathbb{P}} \frac{N(1,2p+1)}{M_1},$$

and

$$\text{Var}(u_{n,j}) \rightarrow 0, \text{ for } j \in \{0, 1, p, 2p+1\}. \quad (\text{A.1})$$

Using Lemma A.1, we get

$$\mathbb{E}(u_{n,0}) = 1, \mathbb{E}(u_{n,1}) = \frac{\mathbb{E}(h_K \beta_1 K_1)}{\mathbb{E}(K_1)} = o(1),$$

$$\mathbb{E}(u_{n,l}) = \frac{\mathbb{E}(h_K^{-l} \beta_1^l K_1)}{\mathbb{E}(K_1)} \rightarrow \frac{N(1,l)}{M_1} \text{ for } l \in \{p, 2p+1\}.$$

Concerning Eq (A.1), we can write

$$\begin{aligned} \text{Var}(u_{n,j}) &= \frac{1}{n^2 \mathbb{E}(K_1)^2} \left( \sum_{i=1}^n \text{Var}(\beta_i^j h_K^{-j} K_i) + 2 \sum_{1 \leq i \neq l \leq n} \text{Cov}(\beta_i^j h_K^{-j} K_i, \beta_l^j h_K^{-j} K_l) \right) \\ &= Q_{n1} + Q_{n2}. \end{aligned}$$

For the term  $Q_{n1}$ , we have

$$\begin{aligned} Q_{n1} &= \frac{1}{n^2 \mathbb{E}(K_1)^2} \sum_{i=1}^n \text{Var}(\beta_i^j h_K^{-j} K_i) \tag{A.2} \\ &\leq \frac{1}{n \mathbb{E}(K_1)^2} \mathbb{E}(\beta_1^{2j} h_K^{-2j} K_1^2) \\ &\leq \frac{1}{n \phi_x(h_K)}. \end{aligned}$$

Moreover, for computing the term of  $Q_{n2}$ , we used the same technique as in [62], we define the sets  $W_1$  and  $W_2$  as follows

$$W_1 = \{(i, l) \in \{1, \dots, n\} \text{ such that } 1 \leq |i - l| \leq u_n\},$$

and

$$W_2 = \{(i, l) \in \{1, \dots, n\} \text{ such that } u_n + 1 \leq |i - l| \leq n - 1\},$$

where the sequence  $u_n$  is selected in such a way that  $u_n \rightarrow +\infty$  as  $n \rightarrow +\infty$ . Based on the aforementioned splitting, we obtain

$$Q_{1,n} = \frac{1}{n^2 \mathbb{E}^2(K_1)} \sum_{W_1} \text{Cov}(h_K^{-j} \beta_i^j K_i, h_K^{-j} \beta_l^j K_l), \tag{A.3}$$

and

$$Q_{2,n} = \frac{1}{n^2 \mathbb{E}^2(K_1)} \sum_{W_2} \text{Cov}(\beta_k^c K_k, h_K^{-j} \beta_l^j K_l). \tag{A.4}$$

For the sum  $Q_{1,n}$ , by assumptions (H2-ii), (H5) and by Lemma of [92], we have

$$\begin{aligned} |\text{Cov}(h_K^{-2j} \beta_i^j K_i, \beta_l^j K_l)| &\leq |\mathbb{E}(h_K^{-2j} \beta_i^j K_i \beta_l^j K_l)| + \mathbb{E}^2(h_K^{-j} \beta_1^j K_1) \\ &\leq C(\mathbb{E}(K_i K_l) + \mathbb{E}^2(K_1)) \\ &\leq C(\phi_x^{1+\epsilon}(h_K) + \phi_x^2(h_K)) \\ &\leq C(\phi_x^{1+\epsilon}(h_K)). \end{aligned}$$

Then, we readily infer

$$|Q_{1,n}| \leq \frac{1}{n^2 \mathbb{E}^2(K_1)} \sum_{W_1} C \phi_x^{1+\epsilon}(h_K)$$

$$\leq \frac{u_n}{n} \phi_x^{-1+\epsilon}(h_K).$$

On the other hand, for computing the sum of covariance  $Q_{1,n}$ , we use the inequality for the bounded mixing processes, then for any  $l \neq i$ , we have

$$|\text{Cov}(h_K^{-j} \beta_i^j K_i, h_K^{-j} \beta_l^j K_l)| \leq C \alpha(|i-l|).$$

Next, using the inequality

$$\sum_{j \geq x+1} j^{-a} \leq \int_{u \geq x} u^{-a},$$

we derive

$$\sum_{i=1}^n \sum_{W_2} \alpha(|i-l|) \leq \frac{C n (u_n)^{1-a}}{a-1}. \quad (\text{A.5})$$

Therefore, we readily obtain

$$|Q_{2,n}| \leq \frac{C (u_n)^{1-a}}{n \phi_x^2(h_K)}.$$

Subsequently, by choosing  $u_n = \lfloor \phi_x^{1+\epsilon}(h_K) \rfloor^{-1/a}$  and by assumption (H7), we obtain

$$Q_{1,n} \rightarrow 0 \text{ and } Q_{2,n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence the proof is complete.  $\square$

*Proof of Lemma 6.2.* We consider a given vector of real numbers, denoted as  $\mathbf{a} =^\top (\alpha_0 \dots \alpha_p) \neq 0$ , then

$$\sqrt{n \phi_x(h_K)} \mathbf{a}^\top V_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \mathbb{E}(Z_i)) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}(Z_i), \quad (\text{A.6})$$

where

$$Z_i = \frac{\sqrt{n \phi_x(h_K)}}{\mathbb{E}(K_1)} \left( \sum_{j=0}^p \alpha_j h_K^{-j} \beta_i^j \right) K_i (Y_i - r(x)).$$

By the Cramér-Wold theorem and Slutsky's theorem, to show (A.6), it suffices to prove the following two claims:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}(Z_i) = \frac{\sqrt{n \phi_x(h_K)} \Psi'(0)}{M_1} \left( \alpha_0 o(h_K) + \sum_{j=1}^p \alpha_j h_K^{j+1} N(1, j+1) \right), \quad (\text{A.7})$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \mathbb{E}(Z_i)) \xrightarrow{D} N(0, \sigma^2(x) \mathbf{a}^\top \mathbf{V} \mathbf{a}). \quad (\text{A.8})$$

First, we proof Claim (A.7). We have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}(Z_i) = \sqrt{n} \mathbb{E}(Z_1)$$

$$\begin{aligned}
&= \frac{\sqrt{n\phi_x(h_K)}}{\mathbb{E}(K_1)} \mathbb{E} \left( \left( \alpha_0 + \sum_{j=1}^p \alpha_j h_K^{-j} \beta_1^j \right) K_1 (Y_1 - r(x)) \right) \\
&= \frac{n\phi_x(h_K)}{\mathbb{E}(K_1)} \underbrace{a_0 \mathbb{E}(K_1(Y_1 - r(x)))}_{S_1} + \frac{n\phi_x(h_K)}{\mathbb{E}(K_1)} \sum_{j=1}^p a_j \underbrace{\mathbb{E} \left( h_K^{-j} \beta_1^j K_1 (Y_1 - r(x)) \right)}_{S_2}.
\end{aligned}$$

Next, by the same arguments as those used by [72] and [40], we get

$$\begin{aligned}
S_1 &= \alpha_0 \Psi'(0) \mathbb{E}(K_1 \Psi(\beta_1)) + o(\mathbb{E}(K_1 \beta_1)) \\
&= \alpha_0 \Psi'(0) o(h_K \phi_x(h_K)),
\end{aligned}$$

and

$$S_2 = \sum_{i=1}^p a_j \Psi'(0) h_K^{j+1} N(1, j+1) \phi_x(h_K),$$

then, using Lemma A.1, we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}(Z_i) = \sqrt{n\phi_x(h_K)} \left( \frac{\alpha_0 \Psi'(0)}{M_1} o(h_K) + \sum_{i=1}^p a_j \Psi'(0) h_K^{j+1} N(1, j+1) \right). \quad (\text{A.9})$$

Now, we prove Claim (A.8). Using The CLT by [56] (see Corollary 2.2, Page 196), which rests on the asymptotic behavior of the quantity

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}[\Delta_i^2], \quad (\text{A.10})$$

where  $\Delta_i = \frac{1}{\sqrt{n}} (Z_i - \mathbb{E}(Z_i))$  in addition to the assumptions:

**Assumption 1.** (A1) There exists a sequence  $\tau_n = o(\sqrt{n})$ , such that

- (i)  $\tau_n = o(\sqrt{n})$  such that  $\tau_n \leq \left( \max_{i=1, \dots, n} C_i \right)^{-1}$ , where  $C_i = \text{ess sup}_{\omega \in \Omega} |\Delta_i|$ ,
- (ii) for all  $\epsilon > 0$ ,  $\frac{n}{\tau_n} \alpha(\epsilon \tau_n) \rightarrow 0$ , for all  $\epsilon > 0$ .

**Assumption 2.** (A2) There exists a sequence  $(m_n)$  of positive integers tending to  $\infty$ , such that

- (i)  $nm_n \gamma_n = o(1)$ , where  $\gamma_n := \max_{1 \leq i \neq j \leq n} (\mathbb{E}|\Delta_i \Delta_j|)$ ,
- (ii) for all  $\epsilon > 0$ ,  $\frac{n}{\tau_n} \alpha(\epsilon \tau_n) \rightarrow 0$ .

We start by computing the limit of (A.10). In order to do so, let us observe that

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}(\Delta_i^2) &= \text{Var}(Z_1) \\
&= \frac{\phi_x(h_K)}{\mathbb{E}^2(K_1)} \mathbb{E} \left( \left( \alpha_0 + \sum_{j=1}^p \alpha_j h_K^{-j} \beta_1^j \right) K_1^2 (Y_1 - r(x))^2 \right) \\
&\quad - \frac{\phi_x(h_K)}{\mathbb{E}^2(K_1)} \mathbb{E}^2 \left( \left( \alpha_0 + \sum_{j=1}^p \alpha_j h_K^{-j} \beta_1^j \right) K_1 (Y_1 - r(x)) \right) = A_1 - A_2.
\end{aligned}$$

Using the fact that

$$A_1 = \frac{\phi_x(h_K)}{\mathbb{E}^2(K_1)} \left( \underbrace{\alpha_0^2 \mathbb{E}(K_1^2(Y_1 - r(x))^2)}_{I_1} + 2\alpha_0 \underbrace{\sum_{j=1}^p \alpha_j \mathbb{E}(h_K^{-j} \beta_1^j K_1^2(Y_1 - r(x))^2)}_{I_2} \right) + \frac{\phi_x(h_K)}{\mathbb{E}^2(K_1)} \mathbb{E} \left( \underbrace{\left( \sum_{j=1}^p \alpha_j h_K^{-j} \beta_1^j \right)^2}_{I_3} K_1^2(Y_1 - r(x))^2 \right),$$

in combination with Lemma A.1, we readily find

$$I_1 = a_0^2 \sigma^2(x) M_1 \phi_x(h_K),$$

$$I_2 = a_0 \sum_{j=1}^p a_j \sigma^2(x) O(\phi_x(h_K)),$$

and

$$I_3 = \sum_{j=1}^p a_j^2 \mathbb{E}(h_K^{-2j} \beta_1^{2j} (K_1^2(Y_1 - r(X_1))^2)) + 2 \sum_{1 \leq i \neq j \leq p} a_j a_i \mathbb{E}(h_K^{-j-i} \beta_1^{j+i} (K_1^2(Y_1 - r(X_1))^2))$$

$$= \sigma^2(x) M_1 \sum_{j=1}^p a_j^2 N(2, 2j) \phi_x(h_K) + 2 \sigma^2(x) M_1 \sum_{1 \leq i \neq j \leq p} a_j a_i O(\phi_x(h_K)).$$

Under assumptions (H1), we have

$$\lim_{n \rightarrow +\infty} A_1 = \sigma^2(x) \frac{M_1}{M_2^2} \left( a_0^2 + \sum_{j=1}^p a_j N(2, 2j) \right)$$

$$= \sigma^2(x) a^\top \mathbf{V} a.$$

Subsequently, the assertion (A.7) suggests that  $\mathbb{E}(Z_1) \rightarrow 0$  as  $n \rightarrow \infty$ . Consequently, we infer that  $A_2 \rightarrow 0$  as  $n \rightarrow \infty$ . Concerning Assumption (1), the fact that the kernel  $K(\cdot)$  is bounded, allows us to infer that

$$C_i = O\left(\frac{1}{\sqrt{n\phi_x(h_K)}}\right).$$

Therefore, an appropriate choice is the following

$$\tau_n = \sqrt{\frac{n\phi_x(h_K)}{\log n}}.$$

Furthermore, this choice gives, for all  $\epsilon > 0$

$$\frac{n}{\tau_n} \alpha(\epsilon\tau_n) \leq C \left( n^{1-(a+1)/2} (\phi_x(h_K))^{-(a+1)/2} (\log n)^{(a+1)/2} \right)$$

$$\begin{aligned} &\leq Cn^{1-(a+1)/2+(a+1)/2(a-1)}(\log n)^{(a+1)/2} \\ &\leq Cn^{(3a-a^2)/2(a-1)}(\log n)^{(a+1)/2} \rightarrow 0 \text{ since } a > 3. \end{aligned}$$

Concerning Assumption (2), using assumptions (H2-ii) and (H3), we have

$$\begin{aligned} \mathbb{E}|\Delta_i \Delta_j| &\leq \frac{\phi_x(h_K)}{n\mathbb{E}(K_1)^2} \sum_{0 \leq j \neq i \leq p} |\alpha_i \alpha_j| \mathbb{E}(K_i K_j) \\ &\leq \frac{\sum_{0 \leq j \neq i \leq p} |\alpha_i \alpha_j|}{n\phi_x(h_K)} \sup_{i \neq j} \mathbb{P}((X_i, X_j) \in B(x, h_K) \times B(x, h_K)) \\ &\leq \frac{C\phi_x^\epsilon(h_K)}{n}. \end{aligned}$$

Elsewhere, using the fact that

$$\sum_{j \geq x+1} j^{-a} \leq \int_{u \geq x} u^{-a} = \left[ (a-1)x^{a-1} \right]^{-1}, \quad (\text{A.11})$$

we obtain

$$\sum_{j=m_n+1}^{\infty} \alpha(j) \leq \sum_{j=m_n}^{\infty} \alpha(j) \leq \int_{t \geq m_n} t^{-a} dt = \frac{m_n^{1-a}}{a-1},$$

thus,

$$\left( \sum_{j=m_n+1}^{\infty} \alpha(j) \right) \sum_{i=1}^n C_i = O\left( \frac{m_n^{1-a}}{a-1} \sqrt{\frac{n}{\phi_x(h_K)}} \right).$$

We select

$$m_n = \left\lfloor \left( \frac{\phi_x(h_K)}{n \log n} \right)^{1/(2(1-a))} \right\rfloor,$$

such that  $\lfloor \cdot \rfloor$  represents the function of the integer part. It is evident that, under assumption (H7),  $m_n \rightarrow \infty$ . Furthermore, by substituting the expression for  $m_n$ , we readily get

$$\sum_{j=m_n+1}^{\infty} \alpha(j) \sum_{i=1}^n C_i = o(1).$$

Once more, considering assumption (H2-i) and (H7), we have

$$\begin{aligned} m_n \gamma_n &\leq Cn^{-1-1/(2(1-a))} (\phi_x(h_K))^{1+1/(2(1-a))} (\log n)^{-1/(2(1-a))} \\ &\leq n^{(-3+2a)/(2(1-a))} (\phi_x(h_K))^{(3-2a)/(2(1-a))} (\log n)^{-1/(2(1-a))} \\ &\leq n^{-1+\eta(2a-3)/(2(1-a))} (\log n)^{-1/(2(1-a))} = o(n^{-1}). \end{aligned}$$

Hence the proof is complete.  $\square$

*Proof of Theorem 3.7.* According to the formula of (6.2), we deduce the following error formula

$$\widehat{c}(x) - c(x) = U_n^{-1}(x) V_n^*(x) - U_n^{-1}(x) e_n(x) + \frac{h_K^{p+1}}{(p+1)!} \Psi^{(p+1)}(0) U_n^{-1}(x) e_n(x) + O(h_K^{p+1}) U_n^{-1}(x) e_n(x),$$

where  $e_n(x) = {}^\top (u_{n,p+1}(x), \dots, u_{n,2p+1}(x))$ . Hence, Theorem 3.7 is a direct result of the following lemmas.

**Lemma A.2.** Assume that the conditions (U1), (U3), (U5) and (U6) are fulfilled. We have, as  $n \rightarrow \infty$ ,

$$\sup_{x \in S_{\mathcal{F}}} |u_{nj}(x) - \mathbb{E}(u_{nj}(x))| = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi_x(h_K)}} \right).$$

**Lemma A.3.** Assume that the conditions (U1)–(U6) are fulfilled. We have, as  $n \rightarrow \infty$ ,

$$\sup_{x \in S_{\mathcal{F}}} |v_{nj}^*(x) - \mathbb{E}(v_{nj}^*(x))| = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi_x(h_K)}} \right),$$

$$\sup_{x \in S_{\mathcal{F}}} |e_{nj}(x) - \mathbb{E}(e_{nj}(x))| = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi_x(h_K)}} \right).$$

**Lemma A.4.** Assume that the conditions (U1)–(U5) are fulfilled. We have, as  $n \rightarrow \infty$ ,

$$\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(v_{nj}^*(x))| = O(h_K^{p+1}), \quad \sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(e_{nj}(x))| = O(1).$$

□

*Proof of Lemma A.2.* The proof of this lemma relies on Proposition A.11-i in [38]. In order to apply the latter, we need to compute a certain quantity. In a first attempt, we evaluate the expression  $\mathbb{E}(u_{nj}(x))$ . For this, we use Lemma A.1 and by the fact that observations are identically distributed, we have

$$\mathbb{E}(u_{nj}(x)) = \frac{1}{h_K^j \mathbb{E}(K_1)} \mathbb{E}(\beta_1^j K_1) \tag{A.12}$$

$$= O(1). \tag{A.13}$$

Moreover, for the covariance term, we have :

$$S_{n,\alpha,l}^2(x) = \sum_{i=1}^n \sum_{i=1}^n |\text{Cov}(A_i^\alpha(x), A_j^l(x))|, \tag{A.14}$$

where

$$A_i^l(x) = \frac{1}{h_K^l} (\beta_i^l k_i(x) - \mathbb{E}(\beta_i^l(x) k_i(x))), \quad \text{for } l \in \{1, 2, \dots, p\}. \tag{A.15}$$

Under some of the dependence assumptions, we treat the term of  $S_{n,\alpha,l}^2(x)$ . Then we have

$$S_{n,\alpha,l}^2(x) = R_{n,1}(x) + R_{n,2}(x) + n \text{Var}(A_1^l(x)),$$

with

$$R_{n,1}(x) = \sum_{W_1} \text{Cov}(A_j^l(x), A_i^\alpha(x)); \quad W_1 = \{(i, j) \in \{1, 2, \dots, n\} \text{ such that } 1 \leq |i - l| \leq u_n\},$$

and

$$R_{n,2}(x) = \sum_{W_2} \text{Cov}(A_j^l(x), A_i^\alpha(x)); \quad W_2 = \{(i, j) \in \{1, 2, \dots, n\} \text{ such that } u_n + 1 \leq |i - l| \leq n - 1\}.$$

For the term variance using Lemma A.1 and we deduce that

$$n \text{Var}(A_1^l(x)) = O(n\phi_x(h_K)). \quad (\text{A.16})$$

Concerning the covariance term  $R_{n,1}(x)$ , we use the same idea as those used in the Eq (A.3) with the choice  $u_n = \phi_x^{-\epsilon}(h_K)$ , we get

$$R_{n,1}(x) = O(n\phi_x(h_K)), \quad (\text{A.17})$$

and for the term  $R_{n,1}(x)$  we applied Rio inequality, see Theorem B.3 (ii). So for this technique, we need to compute the absolute moments of the random variable  $A_j^l(x)$ . Then, using a Newton expansion, we have

$$\begin{aligned} \mathbb{E}[|A_j^l(x)|^m] &= \mathbb{E} \left| h_K^{-lm} \left( K_j(x) \beta_j^l(x) - \mathbb{E}[K_j(x) \beta_j^l(x)] \right)^m \right| \\ &= h_K^{-lm} \mathbb{E} \left| \sum_{i=0}^m C_m^i \left( K_j(x) \beta_j^l(x) \right)^i (x) \left( \mathbb{E}[K_j(x) \beta_j^l(x)] \right)^{m-i} (-1)^{m-i} \right| \\ &\leq h_K^{-lm} \sum_{i=0}^m C_m^i \mathbb{E} \left| K_j(x) \beta_j^l(x) \right|^i (x) \left| \mathbb{E}[K_j(x) \beta_j^l(x)] \right|^{m-i} \\ &\leq h_K^{-lm} \sum_{i=0}^m C_m^i \mathbb{E} \left| K_j^i(x) \beta_j^{li}(x) \right| \left| \mathbb{E}[K_j(x) \beta_j^l(x)] \right|^{m-i}, \end{aligned}$$

where  $C_m^i = \frac{m!}{i!(m-i)!}$ . Next, by applying Lemma A.1, we get

$$\mathbb{E}[|A_j^l|^m] = O(\phi_x(h_K)).$$

By following the same reasoning used to establish Eq (A.4) with the choice  $u_n = \phi_x^{-\epsilon}(h_K)$  permits to get:

$$R_{n,2}(x) = O(n\phi_x(h_K)). \quad (\text{A.18})$$

Finlay, by Eqs (A.16), (A.17) and (A.18), we deduce that

$$S_{n,\alpha,l}^2(x) = O(n\phi_x(h_K)). \quad (\text{A.19})$$

On the other hand, we can write

$$u_{nj}(x) - \mathbb{E}(u_{nj}(x)) = \frac{1}{n\phi_x(h_K)} \sum_{i=1}^n A_i^l(x), \quad (\text{A.20})$$



where  $A_i'(x)$  is defined in (A.15). Now, we apply the Definition 3.6 to show the uniform convergence of our estimator on a subset  $S$  of  $\mathcal{F}$ . For this aim, we take

$$z(x) = \arg \min_{j \in \{1, \dots, d_n\}} |\delta(x, x_j)|.$$

Let us consider the following decomposition:

$$\begin{aligned} \sup_{x \in S_{\mathcal{F}}} |u_{nj}(x) - \mathbb{E}(u_{nj}(x))| &\leq \underbrace{\sup_{x \in S_{\mathcal{F}}} |u_{nj}(x) - u_{nj}(x_{z(x)})|}_{T_1} \\ &+ \underbrace{\sup_{x \in S_{\mathcal{F}}} |u_{nj}(x_{z(x)}) - \mathbb{E}(u_{nj}(x_{z(x)}))|}_{T_2} \\ &+ \underbrace{\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(u_{nj}(x_{z(x)})) - \mathbb{E}(u_{nj}(x))|}_{T_3}. \end{aligned}$$

Concerning the term  $T_2$ , for all  $\epsilon > 0$ , we have

$$\begin{aligned} \mathbb{P}(T_2 > \epsilon) &= \mathbb{P}\left(\max_{j \in \{1, \dots, d_n\}} |u_{nj}(x) - \mathbb{E}(u_{nj}(x_{z(x)}))| > \epsilon\right) \\ &\leq d_n \max_{j \in \{1, \dots, d_n\}} \mathbb{P}\left(|u_{nj}(x) - \mathbb{E}(u_{nj}(x_{z(x)}))| > \epsilon\right) \\ &\leq d_n \max_{j \in \{1, \dots, d_n\}} \mathbb{P}\left(|A_{nj}(x_{z(x)})| > \epsilon n \phi(h_K)\right). \end{aligned}$$

Next, we can apply the Proposition A.11 of [38] for any  $m > 2$ ,  $\tau > 0$ ,  $\varpi \geq 1$  and for certain  $0 < C < \infty$ ,

$$\mathbb{P}(T_2 > \tau) < C(A_1 + A_2),$$

where

$$A_1 = d_n \left(1 + \frac{\tau^2 n^2 \phi(h_K)^2}{\varpi S_{n,\alpha,l}^2}\right)^{-\varpi/2}, \quad A_2 = d_n n \varpi^{-1} \left(\frac{\varpi}{\tau n \phi(h_K)}\right)^{(a+1)m/(a+m)}. \quad (\text{A.21})$$

By using equation of (A.19) and under assumption (U1), we obtain that

$$S_{n,\alpha,l}^2 = \sup_{x \in S_{\mathcal{F}}} S_{n,\alpha,l}^2(x) = O(n\phi(h_K)).$$

Now, for  $\eta > 0$ , we choose

$$\tau = \eta \sqrt{\frac{\log d_n}{n\phi(h_K)}} \quad \text{and} \quad \varpi = (\log(d_n))^2.$$

Therefore, by utilizing the fact that  $\Psi_{S_{\mathcal{F}}}(\epsilon) = \log(d_n)$  and by taking  $C \eta^2 = \lambda$ , we obtain under the assumption (U6) the following terms

$$A_1 = O(d_n^{1-\lambda}) \quad \text{and} \quad A_1 = O(n^{1-\lambda'}), \quad \text{for } \lambda, \lambda' > 0.$$

Finally, for  $\eta$  large enough, we infer

$$\mathbb{P} \left( T_2 > \eta \sqrt{\frac{\log d_n}{n\phi(h_K)}} \right) < C(d_n^{1-\lambda} + n^{1-\lambda'}).$$

Using the fact that

$$\sum_{n=1}^{\infty} d_n^{1-\lambda} < \infty \text{ and } \sum_{n=1}^{\infty} \frac{1}{n^{1+\lambda'}} < \infty,$$

we get

$$T_2 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h_K)}} \right).$$

For the term  $T_1$  we have

$$\begin{aligned} T_1 &\leq \frac{C}{nh_K^j \phi(h_K)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n K_i(x) \mathbf{1}_{B(x, h_K)}(X_i) | \beta_i^j(x) - \beta_i^j(x_{z_x}) \mathbf{1}_{B(x_{z_x}, h_K)}(X_i) | \\ &\quad + \frac{C}{nh_K^j \phi(h_K)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n \beta_i^j(x_{z_x}) \mathbf{1}_{B(x_{z_x}, h_K)}(X_i) | K_i(x) \mathbf{1}_{B(x, h_K)}(X_i) - K_i(x_{z_x}) | \\ &:= L_1^j + L_2^j. \end{aligned}$$

Evaluation of the term  $L_1^j$ . Under assumption (U3) we have

$$\mathbf{1}_{B(x, h_K)}(X_i) | \beta_i^j(x) - \beta_i^j(x_{z_x}) \mathbf{1}_{B(x_{z_x}, h_K)}(X_i) | \leq C \iota_n h_K^{j-1} \mathbf{1}_{B(x, h_K) \cap B(x_{z_x}, h_K)}(X_i) + C h_K^j \mathbf{1}_{B(x, h_K) \cap \overline{B(x_{z_x}, h_K)}}(X_i),$$

which gives the following expression

$$\begin{aligned} L_1^j &\leq \frac{C \iota_n}{nh_K \phi(h_K)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n K_i(x) \mathbf{1}_{B(x, h_K) \cap B(x_{z_x}, h_K)}(X_i) \\ &\quad + \frac{C}{n\phi(h_K)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n K_i(x) \mathbf{1}_{B(x, h_K) \cap \overline{B(x_{z_x}, h_K)}}(X_i). \end{aligned} \tag{A.22}$$

Making use of the Lipschitz condition on  $K(\cdot)$  and by hypotheses (U3) enables us to directly write

$$\begin{aligned} &| \beta_i^j(x_{z(x)}) | \mathbf{1}_{B(x_{z(x)}, h_K)}(X_i) | K_i(x) \mathbf{1}_{B(x, h_K)}(X_i) - K_i(x_{z(x)}) | \\ &\leq C h_K^j \iota_n \mathbf{1}_{B(x, h_K) \cap B(x_{z(x)}, h_K)}(X_i) + C h_K^j K_i(x_{z(x)}) \mathbf{1}_{\overline{B(x, h_K)} \cap B(x_{z(x)}, h_K)}(X_i). \end{aligned}$$

Therefore, we readily obtain

$$\begin{aligned} L_2^j &\leq \frac{C \iota_n}{n\phi(h_K)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n \mathbf{1}_{B(x, h_K) \cap B(x_{z(x)}, h_K)}(X_i) \\ &\quad + \frac{C}{n\phi(h_K)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n K_i(x_{z(x)}) \mathbf{1}_{\overline{B(x, h_K)} \cap B(x_{z(x)}, h_K)}(X_i). \end{aligned}$$

Under assumption (U5) en by combined the last inequality with (A.22) implies that

$$T_1 \leq \frac{Z_i}{n\phi(h_K)},$$

where

$$Z_i = \frac{C_{l_n}}{h_K} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n \mathbf{1}_{B(x, h_K) \cup B(x_{z(x)}, h_K)}(X_i).$$

Using a similar approach to the one employed in the proof of (A.14), under the assumptions (U1), (U2) and (U6), we obtain

$$S_n^2 = |\text{Cov}(Z_i, Z_j)| = O(n\phi(h_K)).$$

Next, by employing a similar proof to the one applied in the evaluation of  $T_2$ , we infer that

$$T_1 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h_K)}} \right).$$

Concerning the term of  $T_3$ , it is obvious that

$$T_3 < \mathbb{E} \left( \sup_{x \in S_{\mathcal{F}}} |u_{nj}(x) - u_{nj}(x_{z(x)})| \right),$$

we infer that

$$T_3 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h_K)}} \right).$$

This completes the proof. □

*Proof of Lemma A.4.* By conditioning on  $X_1$ , we have

$$\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(v_{nj}^*(x))| \leq \frac{1}{|\mathbb{E}(K_1)|} |\mathbb{E}((\beta_1 h_K^{-1})^j K_1)| \sup_{X_1 \in B(x, h_K)} |r(X_1) - r(x)|,$$

so under assumption (U4) and by using the uniform version of Lemma A.1, we obtain

$$\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(v_{nj}^*(x))| = O(h_K^{p+1}).$$

Concerning the term  $\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(e_{nj}(x))| = O(1)$  is already treated in Eq (A.12). □

*Proof of Theorem 4.1.* Using similar reasoning employed for the regression function, we show that

$$\sqrt{n\phi_x(h_K)}(B_n^* - B_L^p(x, y)) \xrightarrow{D} N(0, V_F(x, y)), \quad (\text{A.23})$$

where

$$B_n^* = {}^\top (\Upsilon_{n,0}^*, \dots, \Upsilon_{n,p}^*), \quad V_F(x, y) = F^x(y)(1 - F^x(y))a^\top \mathbf{V} a$$

and

$$\Upsilon_{n,j}^* = \frac{1}{n\mathbb{E}(K)} \sum_{i=1}^n \left( \frac{\beta(X_i, x)}{h_K} \right)^j K(h_K^{-1}(\delta(X_i, x)))(L(h_L^{-1}(Y_i - y)) - F^x(y)).$$

For any given vector of real numbers  $\mathbf{a} = {}^\top (\alpha_0, \dots, \alpha_p) \neq 0$ , we have

$$\sqrt{n\phi_x(h_K)}a^\top B_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n (R_i - \mathbb{E}(R_i)) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}(R_i),$$

where

$$R_i = \frac{\sqrt{\phi_x(h_K)}}{\mathbb{E}(K_1)} \left( \sum_{l=0}^p \alpha_l h_K^{-l} \beta_l^l \right) K_i (L_i - F^x(y)). \quad (\text{A.24})$$

According to the Cramér-Wold theorem and Eq (A.24), Eq (A.23), will be verified if we prove the following two statements:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (R_i - \mathbb{E}(R_i)) \xrightarrow{D} N(0, F^x(y)(1 - F^x(y)) a^\top \mathbf{V} a), \quad (\text{A.25})$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbb{E}(R_i)) = \sqrt{n\phi_x(h_K)} B_L^p(x, y). \quad (\text{A.26})$$

□

*Proof of statement A.25.* We utilize Bernstein's big-block method, as employed in Theorem 3.1 of [55]. The set  $(1, 2, \dots, n)$  is partitioned into  $2\nu_n + 1$  subsets, each containing large blocks of size  $(w_n)$  and small blocks of size  $(q_n)$ , by considering

$$\nu = \left\lfloor \frac{n}{w_n + q_n} \right\rfloor.$$

Assumption (M6) allows us to define the size of the large block as follows

$$w_n = \left\lfloor \frac{(n\phi_x(h_K))^{\frac{1}{2}}}{r_n} \right\rfloor.$$

Afterward, under the same assumptions and employing straightforward calculations, we obtain:

$$\lim_{n \rightarrow +\infty} \frac{q_n}{w_n} = 0, \quad \lim_{n \rightarrow +\infty} \frac{w_n}{n} = 0, \quad \lim_{n \rightarrow +\infty} \frac{w_n}{\sqrt{n\phi_x(h_K)}} = 0 \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{n}{w_n} \alpha(q_n) = 0. \quad (\text{A.27})$$

It can be easily inferred that, as  $n$  tends to infinity,

$$\frac{\nu q_n}{n} \simeq \left( \frac{n}{r_n + \nu_n} \right) \frac{q_n}{n} \simeq \frac{q_n}{w_n + q_n} \simeq \frac{q_n}{w_n} = 0. \quad (\text{A.28})$$

Presently, we partition the sum  $\sum_{i=1}^n (R_i - \mathbb{E}(R_i))$  into distinct large and small blocks as outlined below.

Let

$$I_j = (j-1)(w+q) + 1, l_j = (j-1)(w+q) + w + 1, j = 1, 2, \dots, \nu.$$

One can see that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \underbrace{(R_i - \mathbb{E}(R_i))}_{E_i} &= \sum_{j=1}^{\nu} \frac{F_j}{\sqrt{n}} + \sum_{j=1}^{\nu} \frac{F'_j}{\sqrt{n}} + \frac{F_n}{\sqrt{n}} \\ &=: n^{-1/2} (S_{1,n} + S_{2,n} + S_{3,n}), \end{aligned}$$

where the random variables  $F_j, F'_j$  and  $F_n$  are defined by

$$F_j = \sum_{i=l_j}^{l_j+w-1} E_i, \quad F'_j = \sum_{i=l_j}^{l_j+q-1} E_i, \quad F_n = \sum_{i=u(w+q)+1}^n E_i. \quad (\text{A.29})$$

Subsequently, making use of Slutsky's theorem, we establish that the two terms  $\frac{1}{\sqrt{n}}S_{2,n}$  and  $\frac{1}{\sqrt{n}}S_{3,n}$  converge to zero in probability. Before applying Lindeberg-Feller conditions to establish the asymptotic normality of the term  $\frac{1}{\sqrt{n}}S_{1,n}$ , it is imperative to demonstrate that the variables  $F_j$  are asymptotically independent. Therefore, we obtain

$$\frac{\mathbb{E}(S_{2,n}^2)}{n} = \sum_{j=1}^u \frac{1}{n} \text{Var}(F'_j) + \sum_{j=1}^u \sum_{\substack{i=1 \\ i \neq j}}^u \frac{2}{n} \text{Cov}(F'_j, F'_i) =: \frac{1}{n}(A_1 + A_2).$$

Using second-order stationarity, we get

$$\text{Var}(F'_j) = q \text{Var}(E_1) + 2 \sum_{i \neq j}^q \text{Cov}(E_i, E_j).$$

Therefore

$$\frac{A_1}{n} \leq \frac{uq}{n} \underbrace{\text{Var}(E_1)}_{V_F(x,y)} + \frac{2}{n} \sum_{i \neq j}^n \text{Cov}(E_i, E_j).$$

To derive the covariance term, we adopt the identical procedures outlined in the verification of Eq (A.2). Then we set

$$T_1 = \{(i, j) \in \{1, 2, \dots, n\} \text{ such that } 1 \leq |i - j| \leq c_n\},$$

and

$$T_2 = \{(i, j) \in \{1, 2, \dots, n\} \text{ such that } c_n + 1 \leq |i - j| \leq n - 1\}.$$

Based on the aforementioned splitting, we obtain

$$\frac{2}{n} \sum_{i \neq j}^n \text{Cov}(E_i, E_j) = \frac{2}{n} \sum_{G_1} \text{Cov}(R_i, R_j) + \frac{2}{n} \sum_{G_2} \text{Cov}(R_i, R_j) =: U_{1,n} + U_{2,n}.$$

$$\begin{aligned} \text{Cov}(R_i, R_j) &= \frac{\phi_x(h_K)}{\mathbb{E}^2(K_1)} \sum_{l=0}^p \sum_{m=0}^p \alpha_l \alpha_m \mathbb{E} \left( h_K^{-l} h_K^{-m} \beta_j^l \beta_j^m K_i K_j \mathbb{E}((L_i - F^x(y))(L_j - F^x(y)) \mid (X_i, X_j)) \right) \\ &\quad - \frac{\phi_x(h_K)}{\mathbb{E}^2(K_1)} \mathbb{E} \left( \sum_{l=0}^p \alpha_l h_K^{-l} \beta_j^l K_i (L_i - F^x(y)) \right)^2. \end{aligned}$$

Utilizing the inequality

$$|L_i(y) - F^x(y)| \leq 1, \quad (\text{A.30})$$

stated in Lemma A.1 and given the conditions (H2-ii) and (H4-ii), we obtain the following

$$\begin{aligned} |\text{Cov}(R_i, R_j)| &\leq \frac{\phi_x(h_k)}{\mathbb{E}^2(K_1)} \left( \sum_{l=0}^p \sum_{m=0}^p |\alpha_l \alpha_m| \mathbb{E}(K_i K_j) + \left( \sum_{l=0}^p |\alpha_l| \right)^2 \mathbb{E}^2(K_i) \right) \\ &\leq \frac{\phi_x(h_k)}{\mathbb{E}^2(K_1)} \left( C\mathbb{P}((X_i, X_j) \in B(x, h_k) \times B(x, h_k)) + C' M_1^2 \phi_x^2(h_k) \right) \\ &\leq C\phi_x(h_k) \frac{\psi_x(h_k)}{\phi_x^2(h_k)} + C' M_1^2 n \phi_x(h_k). \end{aligned}$$

Then, we readily infer

$$|U_{1,n}| \leq \left( 2Cn^{-1} \phi_x(h_k) \frac{\psi_x(h_k)}{\phi_x^2(h_k)} + 2C'n^{-1} M_1^2 \phi_x(h_k) \right) \underbrace{\sum_{G_1} 1}_{nc_n}.$$

In relation to the summation over the set  $G_2$ , the utilization of Theorem B.3 (ii) gives

$$\sum_{G_2} |\text{Cov}(R_i, R_j)| \leq \sum_{G_2} C(\alpha|j-i|)^{\frac{1}{p}} (\mathbb{E}|R_i|^q)^{\frac{1}{q}} (\mathbb{E}|R_j|^r)^{\frac{1}{r}}.$$

Conditioned on  $X_i$ , and using (A.30) along with assumption (M3), we obtain

$$\mathbb{E}(|R_i|^q) = \frac{\sqrt{\phi_x(h_k)}}{M_1 \phi_x(h_k)} \mathbb{E}(|K_i|^q \mathbb{E}(|L_i - F^x(y)|^q | X_i)) \leq C(\phi_x(h_k))^{-\frac{1}{2} + \frac{1}{q}}.$$

We readily infer that

$$|U_{2,n}| \leq 2 \sum_{G_2} C(\alpha|j-i|)^{\frac{1}{p}} (\phi_x(h_k))^{-1 + \frac{1}{q} + \frac{1}{r}} \quad (\text{A.31})$$

$$\leq 2c_n^{-\delta} (\phi_x(h_k))^{-\frac{1}{p}} \sum_{|l| > c_n} l^\delta (\alpha(|l|))^{\frac{1}{p}}. \quad (\text{A.32})$$

Using the fact that  $\frac{\psi_x(h_k)}{\phi_x^2(h_k)}$  is bounded, we select  $c_n = \lfloor \phi_x(h_k) \rfloor^{-\frac{1}{p\delta}}$  and by assumptions (M4), and (M5) we obtain

$$\frac{2}{n} \sum_{i \neq j}^n \text{Cov}(E_i, E_j) = o(1).$$

Furthermore, for the term  $A_2$ , we have

$$\frac{|A_2|}{n} \leq \sum_{1 \leq i < j \leq n} \frac{2}{n} \text{Cov}(F'_j, F'_i) = \underbrace{\sum_{1 \leq i < j \leq n} \frac{2}{n} |\text{Cov}(R_j, R_i)|}_{o(1)}.$$

Consequently by (A.27), we get

$$\frac{\mathbb{E}(S_{2,n}^2)}{n} = 0, \text{ as } n \rightarrow \infty.$$

Let's now examine the term of  $S_{2,n}^2$

$$\begin{aligned} \frac{\mathbb{E}(S_{3,n}^2)}{n} &\leq \frac{1}{n} \sum_{i=v(w+q)+1}^n \text{Var}(E_i) + \frac{2}{n} \sum_{1 \leq i < j \leq n} |\text{Cov}(E_i, E_j)| \\ &\leq \frac{n - v(w+q)}{n} \text{Var}(E_1) + \underbrace{\frac{2}{n} \sum_{1 \leq i < j \leq n} |\text{Cov}(E_i, E_j)|}_{o(1)}. \end{aligned}$$

Besides, from (A.27) and (A.28), we get  $\frac{vw}{n} \rightarrow 1$ . Then, we obtain

$$\frac{\mathbb{E}(S_{3,n}^2)}{n} = 0, \text{ as } n \rightarrow \infty.$$

In order to demonstrate the asymptotic independence of the variables  $F_j$ , we employ Lemma B.1. By setting  $V_j = \exp\left(\frac{itF_j}{\sqrt{n}}\right)$ , the ensuing relationship

$$\begin{aligned} \left| \mathbb{E} \left( \exp \left( it \frac{S_{1,n}}{\sqrt{n}} \right) \right) - \prod_{j=1}^v \mathbb{E} \left( \exp \left( it \frac{F_j}{\sqrt{n}} \right) \right) \right| &\leq 16v\alpha(q) \\ &\cong \frac{n}{w} \alpha(q) \rightarrow 0. \end{aligned} \quad (\text{A.33})$$

Therefore, according to formula (A.33), the term variance  $S_{1,n}$  is computed as follows:

$$\text{Var}(S_{1,n}) = \frac{vw}{n} \text{Var}(R_1),$$

where

$$\text{Var}(R_1) = \frac{\phi_x(h_K)}{M_1^2 \phi_x^2(h_K)} \sum_{l=0}^p \alpha_l^2 h_K^{-2l} \text{Var}(\beta_l^l K_l (L_l - F^x(y))).$$

Using Lemma 6.3 of [14], we obtain

$$\text{Var}(\beta_l^l K_l (L_l - F^x(y))) = \mathbb{E}(\beta_l^{2l} K_l^2) F^x(y) (1 - F^x(y)).$$

Then, by using the fact that  $\frac{vw}{n} \rightarrow 1$  and under Lemma A.1, we get

$$\text{Var}(R_1) = \alpha_0^2 \frac{M_2}{M_1^2} + \sum_{l=1}^p \alpha_l^2 \frac{N(2, 2l)}{M_1^2} F^x(y) (1 - F^x(y)) = F^x(y) (1 - F^x(y)) a^\top \mathbf{V} a.$$

Finally, our attention turns to Lindeberg's central limit theorem concerning  $F_j$ . It is then enough to demonstrate that for any  $\epsilon > 0$ ,

$$\frac{1}{n} \sum_{j=1}^v \mathbb{E} \left[ F_j^2 \mathbf{1}_{|F_j| > \epsilon \sqrt{n} V_F(x,y)} \right] \rightarrow 0 \text{ as } n \rightarrow +\infty.$$

Under assumption (M2), we have

$$\frac{|F_j|}{\sqrt{n}} \leq \sum_{l=0}^p |\alpha_l| \frac{w}{\sqrt{n\phi_x(h_K)}}.$$

According to (A.27), we obtain

$$\frac{|F_j|}{\sqrt{n}} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Hence, for any given  $\epsilon$  and when  $n$  is sufficiently large, the set  $\{|F_j| > \epsilon \sqrt{n} V_F(x, y)\}$  becomes empty. Consequently, the demonstration of (A.25) is now concluded.  $\square$

*Proof of statement A.26.* We have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbb{E}(R_i)) &= \sqrt{n} \mathbb{E}(R_1) \\ &= \sqrt{n\phi_x(h_K)} \left( \frac{\alpha_0}{\mathbb{E}(K_1)}, \frac{\alpha_1 h_K^{-1}}{\mathbb{E}(K_1)}, \dots, \frac{\alpha_p h_K^{-p}}{\mathbb{E}(K_1)} \right) \cdot \left( \begin{array}{c} \mathbb{E}(K_1 L_1) - F^x(y) \mathbb{E}(K_1) \\ \mathbb{E}(K_1 \beta_1 L_1) - F^x(y) \mathbb{E}(K_1 \beta_1) \\ \vdots \\ \underbrace{\mathbb{E}(K_1 \beta_1^p L_1) - F^x(y) \mathbb{E}(K_1 \beta_1^p)}_{N(x,y)} \end{array} \right). \end{aligned}$$

Next, we compute the term  $N(x, y)$ . Notice that

$$\mathbb{E}(K_1 \beta_1^p L_1) - F^x(y) \mathbb{E}(K_1 \beta_1^p) = \mathbb{E}(K_1 \beta_1^p ((\mathbb{E}(L_1 | X_1) - F^x(y))).$$

Making use of the Taylor expansion under the assumptions (M2) and (M3), we obtain

$$\mathbb{E}(L_1 | X_1) = \sum_{k=0}^p \varphi_{2k}(X_1, y) \frac{h_L^{2k}}{(2k)!} \int_{\mathbb{R}} t^{2k} L'(t) dt + o(h_L^{2p}).$$

Therefore, we readily obtain

$$\begin{aligned} &\mathbb{E}(K_1 \beta_1^p ((\mathbb{E}(L_1 | X_1) - F^x(y))) \\ &= \sum_{k=0}^p \mathbb{E}(K_1 \beta_1^p \varphi_k(X_1, y)) \frac{h_L^{2k}}{(2k)!} \int_{\mathbb{R}} t^{2k} L'(t) dt - \varphi_0(x, y) \mathbb{E}(K_1 \beta_1^p) + o(h_L^{2p} E(K_1 \beta_1^p)). \end{aligned} \quad (\text{A.34})$$

Observe that  $\psi_k(0) = 0$ , we get

$$\mathbb{E}(K_1 \beta_1^p \varphi_k(X_1, y)) = \varphi_k(x, y) \mathbb{E}(K_1 \beta_1^p) + \sum_{a=1}^p \frac{\psi_k^{(a)}(0)}{a!} \mathbb{E}(K_1 \beta_1^{p+a}). \quad (\text{A.35})$$

By combining Eqs (A.34) and (A.35), we can derive

$$\begin{aligned} &\mathbb{E}(K_1 \beta_1^p ((\mathbb{E}(L_1 | X_1) - F^x(y))) \\ &= \sum_{k=0}^p \frac{h_L^{2k}}{(2k)!} \int_{\mathbb{R}} t^{2k} L'(t) dt \left[ \varphi_{2k}(x, y) E(K_1 \beta_1^p) + \sum_{a=1}^p \frac{\psi_k^{(a)}(0)}{a!} \mathbb{E}(K_1 \beta_1^{p+a}) \right] \\ &\quad - \sum_{k=0}^p \frac{h_L^{2k}}{(2k)!} \int_{\mathbb{R}} t^{2k} L'(t) dt \varphi_0(x, y) \mathbb{E}(K_1 \beta_1^p) + o(h_L^{2p} \mathbb{E}(K_1 \beta_1^p)). \end{aligned}$$

So the statement (A.26) is proved.  $\square$



## B. Appendix

This appendix contains supplementary information that is an essential part of providing a more comprehensive understanding of the paper.

**Lemma B.1.** [82] Let  $V_1, \dots, V_L$  be random variables measurable with respect to the  $\sigma$ -algebras  $\mathcal{F}_{i_1}^{j_1}, \dots, \mathcal{F}_{i_L}^{j_L}$  respectively with  $1 \leq i_1 < j_1 < i_2 < \dots < j_L \leq n$ ,  $i_{l+1} - j_l \geq w \geq 1$  and  $|V_j| \leq 1$  for  $j = 1, \dots, L$ . Then

$$\left| \mathbb{E} \left( \prod_{j=1}^L V_j \right) - \prod_{j=1}^L \mathbb{E}(V_j) \right| \leq 16(L-1)\alpha(w),$$

where  $\alpha(w)$  is the strongly mixing coefficient.

**Theorem B.2.** (Lindeberg central limit theorem). For each  $n \geq 1$ , let  $\{U_{n1}, \dots, U_{nr_n}\}$  be a collection of independent random variables such that  $\mathbb{E}(U_{nj}) = 0$  and  $\text{Var}(U_{nj}) < \infty$  for  $j = 1, \dots, r_n$ .

$$\tilde{U}_{nj} = \frac{U_{nj}}{\sqrt{\sum_{k=1}^{r_n} \text{Var} U_{nk}}}, \quad j = 1, \dots, r_n.$$

Then

$$\sum_{j=1}^{r_n} \tilde{U}_{nj} \rightarrow N(0, 1) \text{ in distribution as } n \rightarrow \infty,$$

if for every  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{r_n} \mathbb{E} |\tilde{U}_{nj}|^2 \mathbf{1}(|\tilde{U}_{nj}| > \epsilon) = 0.$$

The following theorem is Proposition A.10. in [38].

**Theorem B.3.** Assume that  $(T_n)_{n \in \mathbb{Z}}$  is  $\alpha$ -mixing. Let us, for some  $k \in \mathbb{Z}$ , consider a real variable  $\mathcal{T}$  (resp.  $\mathcal{T}'$ ) which is  $\mathcal{A}_{-\infty}^k$ -measurable (resp.  $\mathcal{A}_{n+k}^{+\infty}$ -measurable).

(i) If  $\mathcal{T}$  and  $\mathcal{T}'$  are bounded, then:

$$\exists C, 0 < C < +\infty, \text{Cov}(\mathcal{T}, \mathcal{T}') \leq C\alpha(n).$$

(ii) If, for some positive numbers  $p, q, r$  such that  $p^{-1} + q^{-1} + r^{-1} = 1$ , we have  $E\mathcal{T}^p < \infty$  and  $E\mathcal{T}'^q < \infty$ , then:

$$\exists C, 0 < C < +\infty, \text{Cov}(\mathcal{T}, \mathcal{T}') \leq C (E\mathcal{T}^p)^{\frac{1}{p}} (E\mathcal{T}'^q)^{\frac{1}{q}} \alpha(n)^{\frac{1}{r}}.$$



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)