mathematics

Article

Wavelet Density and Regression Estimators for Functional Stationary and Ergodic Data: Discrete Time

Sultana DIDI [1,†], Ahoud AL HARBY [2,†] and Salim BOUZEBDA [3,*,†]

1 Department of Statistics, College of Sciences, Qassim University, P.O. Box 6688, Buraydah 51452, Saudi Arabia
2 Department of Mathematics, College of Sciences, Qassim University, P.O. Box 6688, Buraydah 51452, Saudi Arabia a.almaklafi@qu.edu.sa
3 LMAC (Laboratory of Applied Mathematics of Compiègne), Université de Technologie de Compiègne, 60200 Compiègne, France
* Correspondence: salim.bouzebda@utc.fr
† These authors contributed equally to this work.

Abstract: The nonparametric estimation of density and regression function based on functional stationary processes using wavelet bases for Hilbert spaces of functions is investigated in this paper. The mean integrated square error over adapted decomposition spaces is given. To obtain the asymptotic properties of wavelet density and regression estimators, the Martingale method is used. These results are obtained under some mild conditions on the model; aside from ergodicity, no other assumptions are imposed on the data. This paper extends the scope of some previous results for wavelet density and regression estimators by relaxing the independence or the mixing condition to the ergodicity. Potential applications include the conditional distribution, curve discrimination, and time series prediction from a continuous set of past values.

Keywords: multivariate regression estimation; multivariate density estimation; stationarity; ergodicity; rates of strong convergence; wavelet-based estimators; martingale differences; conditional distribution; curve discrimination

MSC: 62G07; 62G08; 62G05; 62G20; 62H05; 60G42; 60G46

check for updates

Citation: DIDI, S.; AL HARBY, A.; BOUZEBDA, S. Wavelet Density and Regression Estimators for Functional Stationary and Ergodic Data: Discrete Time. *Mathematics* **2022**, *10*, 3433. https://doi.org/10.3390/math10193433

Academic Editor: Manuel Alberto M. Ferreira

Received: 6 August 2022
Accepted: 17 September 2022
Published: 21 September 2022

1. Introduction

The statistical literature has recently become more interested in statistical issues concerning studying functional random variables or variables with values in an infinite-dimensional space. The availability of data measured on ever-finer temporal/spatial grids, such as in meteorology, medicine, satellite images, and many other research disciplines, is driving the growth of this research issue, statistically modeling these data as random functions revealed many complex theoretical and numerical research challenges. The reader may consult the monographs for a summary of the theoretical and practical aspects of functional data analysis. The work in Bosq [1] concerns linear models for random variables with values in a Hilbert space, Ramsay and Silverman [2] discussed scalar-on-function and function-on-function linear models, functional principal component analysis, and parametric discriminant analysis. The work in [3], on the other hand, concentrates on nonparametric methods, particularly kernel-type estimation for scalar-on-function nonlinear regression models. Such tools were extended to classification and discrimination analysis. Horváth and Kokoszka [4] discussed the application of several interesting statistical concepts to the functional data framework, including goodness-of-fit tests, portmanteau tests, and change point problems. The work in [5] focuses on analyzing variance for functional data, whereas that in [6] is more concerned with regression analysis for Gaussian processes. Recent studies and surveys on functional data modeling and analysis can be found in [7–22].

Motivated by diverse applications and their helpful role in statistical inference, the problem of estimating conditional models has been subjected to a wide range of statistical literature, employing many types of estimation approaches, the most common of which are the traditional kernel methods. Such methods, however, may have some limitations when estimating compactly supported or discontinuous curves at boundary points. Alternative wavelet methods are prominent due to their adaptability to discontinuities in the curve to be estimated. In practice, the wavelet procedure provides a simple estimation algorithm to implement and compute. For more information on wavelets theory, we refer to [23–26] and others. The work in [27] discusses wavelet approximation properties in detail, and surveys the use of wavelets in various curve estimation problems. Some wavelet theory applications are discussed in [28] by considering the estimation of the integrated squared derivative density function in the independent unidimensional case. The results in [28] were then extended by [29] to estimate the derivatives of a density for negatively and positively associated sequences, respectively. Rao [30] proposed wavelet estimators for the partial derivatives of a multivariate probability density function, where the rates of almost certain convergence for the independence case are obtained. We cite [31] on estimating partial derivatives of a multivariate probability density function in the presence of additive noise. At this point, we refer to [32]. In the i.i.d. framework, Ref. [33] investigated the density and regression estimation problems for functional data. The authors [33] developed a new adaptive procedure based on the term-by-term selection of wavelet coefficient estimators using wavelet bases for Hilbert spaces of functions. The primary goal of this paper is to extend the previous reference to stationary ergodic processes. To our knowledge, the consideration of the general dependence framework for wavelet analysis is unexplored, which motivates this study. Allow $\{\mathbf{X}_n, n \in \mathbb{Z}\}$ to be a stationary sequence. Consider the backward field $\mathscr{A}_n = \sigma(\mathbf{X}_k : k \leq n)$ and the forward field $\mathscr{B}_n = \sigma(\mathbf{X}_k : k \geq n)$. The sequence is strongly mixing if

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

The sequence is ergodic if

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

where τ is the time-evolution or shift transformation. The naming of strong mixing in the above definition is more stringent than what is ordinarily referred to (when using the vocabulary of measure-preserving dynamical systems) as strong mixing, namely to that

$$\lim_{n\to\infty} \mathbb{P}(A \cap \tau^{-n} B) = \mathbb{P}(A)\mathbb{P}(B)$$

for any two measurable sets A, B, see, for instance, Ref. [34]. As a result, strong mixing implies ergodicity, whereas the converse is not always true (see, for example, Remark 2.6 on page 50 concerning Proposition 2.8 on page 51 in [35]). Some reasons for considering ergodic dependence structure in data rather than a mixing structure are discussed in [36–44], where details on the definition of ergodic property of processes are given, as well as illustrative examples of such processes. One of the arguments used in [45] to justify the ergodic setting is that for certain classes of processes, proving ergodic properties rather than the mixing condition can be much easier. As a result, the ergodicity hypothesis appears to be the best fit and offers a better framework for studying data series generated by noisy chaos. The work in [45] provided an example of an ergodic but non-mixing process in their discussion, which can be summarized as follows: Let $(T_i, \lambda_i) : i \in \mathbb{Z}$ be a strictly stationary process such that $T_i \mid \mathcal{T}_{i-1}$ is a Poisson process with parameter λi, where $\mathcal{T}_i$ is the *sigma*-field generated by $(T_i, \lambda_i, T_{i-1}, \ldots)$. Assume that

$$\lambda_i = f(\lambda_{i-1}, T_{i-1}),$$

and $f : [0, \infty] \times \mathbb{N} \to (0, \infty)$ is a given function. This process is not mixing in general (see Remark 3 of [46]). It is known that any sequence $(\varepsilon_i)_{i \in \mathbb{Z}}$ of i.i.d. random variables is ergodic. Hence, according to Proposition 2.10 in [35], it is easy to see that $(Y_i)_{i \in \mathbb{Z}}$ with

$$Y_i = \vartheta((\ldots, \varepsilon_{i-1}, \varepsilon_i), (\varepsilon_{i+1}, \varepsilon_{i+2}, \ldots)),$$

for some Borel-measurable function $\vartheta(\cdot)$.

The primary goal of this paper is to provide the first complete theoretical justification for wavelet-based functional density and regression function estimation for stationary processes. To our knowledge, the mean integrated square error over adapted decomposition spaces using wavelet estimators in functional ergodic data frameworks has not yet been considered in the literature and thus remains a fundamentally unsolved open problem. By combining several Martingale theory techniques used in the mathematical development of the proofs, we hope to fill this gap in the literature. These tools are not the same as those used in regression estimation under strong mixing or in an independent setting. However, as we will see later, combining existing ideas and results is not enough to solve the problem. Dealing with wavelet estimators in an ergodic setting will involve detailed mathematical derivations.

The following is how the paper is structured. The multiresolution analysis is introduced in Section 2. The main results for density estimation are presented in Section 3. The main results for the regression estimation are presented in Section 4. Some potential applications are listed in Section 5. Section 6 contains some concluding remarks. Section 7 contains a collection of all proofs.

2. Multiresolution Analysis

We will now introduce some basic notation to define wavelet bases for Hilbert spaces of functions following [33,47] with some changes necessary for our setting. In our work, we consider nonlinear, thresholded, wavelet-based estimators. Firstly, we initiate our study by describing elements of the basic theory of wavelet methods and introducing nonlinear wavelet-based estimators; the interested reader may refer to [23,24], see also [48,49] and the references therein, although the wavelet bases on a separable Hilbert space **H** of real or complex-valued functions on a complete separable metric space was introduced later by [47], which we briefly recall here for the sake of reader's convenience. Let **H** be a separable Hilbert space of real-valued functions defined on a complete separable metric space **S**. Since the space **H** is separable, it has an orthonormal basis

$$\mathcal{E} = \{e_j : j \in \Delta\},$$

where Δ is a countable index set. The space **H** is equipped with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\| \cdot \|$.

Consider the sequence of subsets $\{\mathcal{I}_k; k \geq 0\}$ an increasing sequence of finite subsets of Δ such that

$$\bigcup_{k \geq 0} \mathcal{I}_k = \Delta.$$

The subset $\mathcal{J}_k$ denotes the orthogonal complement of $\mathcal{I}_k$ in $\mathcal{I}_{k+1}$, i.e.,

$$\mathcal{J}_k = \mathcal{I}_{k+1} / \mathcal{I}_k.$$

Choose, for any $k \geq 0$, $\zeta_{k,\ell} \in \mathbf{S}$, $\ell \in \mathcal{I}_k$ and $\eta_{k,\ell} \in \mathbf{S}$, $\ell \in \mathcal{J}_k$, such that the following matrices

$$A_k = \left(e_j(\zeta_{k,\ell})\right)_{(j,\ell)\in\mathcal{I}_k\times\mathcal{I}_k}, \quad B_k = \left(e_j(\eta_{k,\ell})\right)_{(j,\ell)\in\mathcal{J}_k\times\mathcal{J}_k}, \tag{1}$$

satisfy one of the two following conditions, for instance, see [33,47] and the references therein.

(A.1) $A_k^* A_k = \mathrm{diag}(a_{k,\ell})_{\ell\in\mathcal{I}_k}$ and $B_k^* B_k = \mathrm{diag}(b_{k,\ell})_{\ell'\in\mathcal{J}_k}$ where $a_{k,\ell}$ and $b_{k,\ell}$ for $\ell \in \mathcal{I}_k$ and $\ell' \in \mathcal{J}_k$ are positive constants.

(A.2) $A_k A_k^* = \mathrm{diag}(c_{k,\ell})_{\ell\in\mathcal{I}_k}$ and $B_k B_k^* = \mathrm{diag}(d_{k,\ell})_{\ell'\in\mathcal{J}_k}$ where $c_{k,\ell}$ and $d_{k,\ell}$ for $\ell \in \mathcal{I}_k$ and $\ell' \in \mathcal{J}_k$ are positive constants.

The condition **(A.1)** implies that

$$a_{k,\ell} = \sum_{j\in\mathcal{I}_k} |e_j(\zeta_{k,\ell})|^2, \quad \ell \in \mathcal{I}_k, \quad \text{and} \quad b_{k,\ell} = \sum_{j\in\mathcal{J}_k} |e_j(\eta_{k,\ell})|^2, \ell \in \mathcal{I}_k, \tag{2}$$

which means that all the columns of A_k and B_k are not the zero vector. As for **(A.2)**, it gives

$$c_{k,\ell} = \sum_{\ell\in\mathcal{I}_k} |e_j(\zeta_{k,\ell})|^2, \quad j \in \mathcal{I}_k, \quad \text{and} \quad d_{k,\ell} = \sum_{\ell\in\mathcal{J}_k} |e_j(\eta_{k,\ell})|^2, j \in \mathcal{I}_k, \tag{3}$$

indicating that all the rows of A_k and B_k are not the zero vector. For any $\mathbf{x} \in \mathbf{S}$, we set

$$\begin{cases} \phi_k(\cdot;\zeta_{k,\ell}) = \sum_{j\in\mathcal{I}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} \overline{e_j(\zeta_{k,\ell})} e_j(\cdot), \\ \psi_k(\cdot;\eta_{k,\ell}) = \sum_{j\in\mathcal{J}_k} \frac{1}{\sqrt{h_{j,k,\ell}}} \overline{e_j(\eta_{k,\ell})} e_j(\cdot), \end{cases} \tag{4}$$

where

$$g_{j,k,\ell} = \begin{cases} a_{k,\ell} & \text{if } \textbf{(A.1)}, \\ c_{k,\ell} & \text{if } \textbf{(A.2)}, \end{cases} \qquad h_{j,k,\ell} = \begin{cases} b_{k,\ell} & \text{if } \textbf{(A.1)}, \\ d_{k,\ell} & \text{if } \textbf{(A.2)}, \end{cases} \tag{5}$$

The following collection form an orthonormal basis of $\mathbf{H}$ (see Theorem 2 of [47]):

$$\mathcal{B} = \left\{\phi_0(x,\zeta_{0,\ell}), \ell \in \mathcal{I}_0; \psi_k(x,\eta_{k,\ell}), k \geq 0, \ell \in \mathcal{J}_k\right\}. \tag{6}$$

For more details, see [33,47,50]. Hence, we conclude that for any $f \in \mathbf{H}$, we have

$$f(\mathbf{x}) = \sum_{\ell\in\mathcal{I}_0} \alpha_{0,\ell}\phi_0(\mathbf{x};\zeta_{0,\ell}) + \sum_{k\geq 0}\sum_{\ell\in\mathcal{J}_k} \beta_{k,\ell}\psi_k(\mathbf{x};\eta_{k;\ell}), \tag{7}$$

where

$$\alpha_{0,\ell} = \langle f, \phi_0(\cdot;\zeta_{0,\ell})\rangle, \quad \beta_{k,\ell} = \langle f, \psi_k(\cdot;\eta_{k;\ell})\rangle. \tag{8}$$

In the following, we add two assumptions on the orthonormal basis $\mathcal{E}$:

(E.1) There exists a constant $C_1 > 0$ such that, for any integer $k \geq 0$, one has

(i)

$$\sum_{j\in\mathcal{I}_k} \frac{1}{g_{j,k,\ell}} \left|e_j(\zeta_{k,\ell})\right|^2 \leq C_1,$$

(ii)

$$\sum_{j\in\mathcal{I}_k} \frac{1}{h_{j,k,\ell}} \left|e_j(\eta_{k,\ell})\right|^2 \leq C_1.$$

(E.2) There exists a constant $C_2 > 0$ such that, for any integer $k \geq 0$, one has

$$\sup_{\mathbf{x}\in\mathcal{S}} \sum_{j\in\mathcal{J}_k} \left|e_j(\mathbf{x})\right|^2 \leq C_2|\mathcal{J}_k|.$$

Remark 1. *Clearly, one can see that assumption* **(E.1)** *is satisfied under assumption* **(A.1)** *when taking* $C_1 = 1$*, we may also refer to [47], Section 4, Example 2 and its applications for more details. [47,50] have presented three examples verifying the assumption* **(E.2)** *taking*

$$\sup_{\mathbf{x}\in\mathcal{S}} \sum_{j\in\mathcal{J}_k} \left|e_j(\mathbf{x})\right|^2 \leq 1,$$

see also [50] Theorem 3.2. Moreover, Ref. [33] have used both assumptions in the case of i.i.d and functional data.

Besov Space

Over the years, many researchers have addressed, from a statistical point of view, the following question: given an estimation method and a prescribed estimation rate for a given loss function, what is the maximal space over which this rate is achieved, for instance, see [27,51] and the references therein. We are interested in the estimation methods based on wavelet bases' thresholding procedures in a natural setting. It is well known that wavelet bases provide characterizations of smoothness spaces such as the Hölder spaces C^s, Sobolev spaces $W^s(L_p)$, and Besov spaces $B^s_q(L_p)$ for a range of indices s that depend both on the smoothness properties of ψ and its dual function $\tilde{\psi}$, for instance, we refer to [51] for more detail and examples, at this point, we may refer to [52]. From a statistical point of view, the following definition is used in approximation theory for the study of nonlinear procedures such as thresholding and greedy algorithms, for instance we refer to [27,49,51,53].

Definition 1 (Besov space)**.** *Let* $s > 0$*. We say that the function* $f \in \mathbf{H}$*, defined by (7), belongs to the Besov space* $\mathcal{B}^s_\infty(\mathbf{H})$ *if and only if:*

$$\sup_{m\geq 0}|\mathcal{J}_m|^{2s} \sum_{k\geq m} \sum_{\ell\in\mathcal{J}_k} |\beta_{k,\ell}|^2 < \infty. \tag{9}$$

Definition 2 (Weak Besov space)**.** *Let* $r > 0$*. We say that the function* $f \in \mathbf{H}$*, defined by (7), belongs to the weak Besov space* $\mathcal{W}^r(\mathbf{H})$ *if and only if:*

$$\sup_{\lambda\geq 0}\lambda^r \sum_{k\geq 0} \sum_{\ell\in\mathcal{J}_k} \mathbb{1}_{\{|\beta_{k,\ell}|\geq\lambda\}} < \infty. \tag{10}$$

3. Problem Definition of the Density Estimation

Let $\{\mathbf{X}_i, Y_i\}_{i\geq 1}$ be a sequence strictly stationary ergodic pairs of random elements where Y_i is a real or complex-valued variable and $\mathbf{X}_i$ takes values in a complete separable metric space of Hilbert space $\mathbf{S}$ associated with the corresponding Borel σ-algebra $\mathcal{B}$. Let $\mathbb{P}_{\mathbf{X}}$ be the probability measure induced by $\mathbf{X}_1$ on $(\mathbf{S}, \mathcal{B})$. Suppose that there exists σ-finite measure ν on the measurable space $(\mathbf{S}, \mathcal{B})$ such that $\mathbb{P}_{\mathbf{X}}$ is dominated by ν. The Radon–Nikodym theorem ensures the existence of a non-negative measurable function $f(\cdot)$ such that

$$\mathbb{P}_{\mathbf{X}}(B) = \int_B f(\mathbf{x})\nu(d\mathbf{x}), B \in \mathcal{B}. \tag{11}$$

In this context, we aim to estimate $f(\cdot)$ based on n observed functional data $\mathbf{X}_1, \ldots, \mathbf{X}_n$– Examples of such random elements $\mathbf{X}_1$ are stochastic processes with continuous sample paths on a finite interval $[a, b]$ with $\mathbf{S} = C[a, b]$ associated with supremum norm and processes with square integrable sample paths on the real line when $\mathbf{S} = L_2(\mathbb{R})$. We suppose that $f \in \mathbf{H}$, where $\mathbf{H}$ is a separable Hilbert space of real or complex-valued functions defined on $\mathbf{S}$ and square-integrable with respect to the σ-finite measure ν. In this paper, we are particularly interested in the wavelet estimation procedures developed in the 1990s; see Meyer's work for the functional data of a Hilbert space, more precisely, the nonlinear estimators. The majority of the approaches carried out in this model consist in introducing kernel estimators techniques to estimate the model's functional part, refer to [54]. Let $f(\cdot)$ be the common density function of the sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, which is assumed to be

(F.1) $\exists C_f > 0$ a known constant such that

$$\sup_{\mathbf{x} \in \mathbf{S}} f(\mathbf{x}) \leq C_f. \tag{12}$$

3.1. Density Function Estimator

From now, we assume that the density function $f(\cdot) \in \mathbf{H}$, a separable Hilbert space. Then, $f(\cdot)$ fulfills the wavelet representation (7). Suppose that we observe a sequence $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ of copies of $(\mathbf{X}, Y)$ that is assumed to be functional stationary and ergodic with $\mathbf{X}$ admitting the density function $f(\cdot)$. We study density estimation through wavelet bases for Hilbert spaces of functions developed by [47]. We consider the estimates of the coefficients $\{\alpha_{k,\ell}\}$ and $\{\beta_{k,\ell}\}$ given, respectively, by (14) and (15). For any $j_0 \leq m$. Here, the resolution level $m = m(n) \to \infty$ at a rate specified below, since we assume that $\phi(\cdot)$ and $\psi_i(\cdot)$ have a compact support so that the summations in (7) are finite for each fixed $\mathbf{x}$ (note that in this case the support of $\phi(\cdot)$ and $\psi_i(\cdot)$ is a monotonically increasing function of their degree of differentiability [24]). We focus our attention on the nonlinear estimators (13) which will be studied in the mean integrated squared error over adapted decomposition spaces, in a similar way as in [33] in the setting of the i.i.d functional processes. The density wavelet hard thresholding estimator $\widehat{f}(\cdot)$ is defined, for all $\mathbf{x} \in \mathbf{S}$, by

$$\widehat{f}_n(\mathbf{x}) = \sum_{\ell \in \mathcal{I}_0} \widehat{\alpha}_{0,\ell} \phi_0(\mathbf{x}; \zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \widehat{\beta}_{k,\ell} \mathbb{1}_{\left\{|\widehat{\beta}_{k,\ell}| \geq \kappa \sqrt{\frac{\ln n}{n}}\right\}} \psi_k(\mathbf{x}; \eta_{k;\ell}), \tag{13}$$

where

$$\widehat{\alpha}_{k,\ell} = \frac{1}{n} \sum_{i=1}^n \phi_k(\mathbf{X}_i; \zeta_{k,\ell}), \tag{14}$$

$$\widehat{\beta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^n \psi_k(\mathbf{X}_i; \eta_{k,\ell}). \tag{15}$$

Here, κ is a large enough constant and m_n is the integer satisfying

$$\frac{1}{2} \frac{n}{\ln n} \leq |\mathcal{J}_{m_n}| \leq \frac{n}{\ln n}. \tag{16}$$

3.2. Estimation Procedure Steps

Our estimation method is divided into three steps:

1. Estimation of the wavelet coefficients $\alpha_{k,\ell}$ and $\beta_{k,\ell}$, see (8), by the estimators $\widehat{\alpha}_{k,\ell}$ and $\widehat{\beta}_{k,\ell}$ defined by Equations (14) and (15);
2. Applying the hard thresholding to select the greatest $\widehat{\beta}_{k,\ell}$;
3. Reconstructing the selected elements of the initial wavelet basis.

It is important to note that our choice is the universal threshold $\kappa \left(\frac{\ln n}{n}\right)^{1/2}$ and the definition of m_n is based on theoretical considerations. The considered estimator does not depend on the smoothness of $f(\cdot)$; we may refer the reader to [50] for more details in the case of the linear wavelet estimator of $f(\cdot)$. Furthermore, for more details on the case of $\mathbf{H} = \mathbb{L}([a, b])$ and more standard nonparametric models, see [27,55]. To state the results, we need some notation. Throughout the paper, we will denote by $\mathcal{F}_i$ the $\sigma-$field generated by

$\{\mathbf{X}_j : 0 \le j \le i\}$ and $\mathcal{G}_i$ the $\sigma-$field generated by $\{(\mathbf{X}_j, Y_j) : \mathbf{X}_{i+1}, 0 \le j \le i\}$. Let $B \in \mathcal{B}$, be an open set of the Borel $\sigma-$algebra $\mathcal{B}$. For any $i = 1, \ldots, n$ define $f_{\mathbf{X}_i}^{\mathcal{F}_{i-1}}(\cdot)$ as the conditional density of $\mathbf{X}_i$ given the $\sigma-$field $\mathcal{F}_{i-1}$. Define

$$F_{\mathbf{X}_i} = \mathbb{P}(\mathbf{X}_i \in B) = \mathbb{P}_X(B), \quad \text{see (11)},$$

and

$$F_{\mathbf{X}_i}^{\mathcal{F}_{i-1}} = \mathbb{P}(\mathbf{X}_i \in B | \mathcal{F}_{i-1})$$

as the distribution function and the conditional distribution function, given the $\sigma-$field $\mathcal{F}_{i-1}$, respectively. The following assumptions will be needed throughout the paper.

(C.0) There is a non-negative measurable function $f^{\mathcal{F}_{i-1}}$ such that

$$\mathbb{P}_{\mathbf{X}}^{\mathcal{F}_{i-1}}(B) = \int_B f^{\mathcal{F}_{i-1}}(\mathbf{x}) \nu(d\mathbf{x}), B \in \mathcal{B}. \tag{17}$$

(C.1) For any $\mathbf{x} \in \mathbf{S}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f^{\mathcal{F}_{i-1}}(\mathbf{x}) = f(\mathbf{x}), \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense.}$$

At this point, we may refer to [56] for further details.

Theorem 1. *Under the conditions* **(C.0)**, **(C.1)**, **(F.1)**, **(E.1)** *and* **(E.2)**, *and (16) for any* $\theta \in (0, 1)$,

$$f \in \mathcal{B}_{\infty}^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H),$$

there exists a constant $C_1 > 0$ *in such a way that*

$$\mathbb{E}\left(\left\|\widehat{f}(\mathbf{x}) - f(\mathbf{x})\right\|^2\right) \le C_1 \left(\frac{\ln n}{n}\right)^{\theta}, \tag{18}$$

for large enough n.

A direct consequence is the following upper bound result: for $s > 0$, if

$$f \in \mathcal{B}_{\infty}^{s/(2s+1)}(\mathbf{H}) \cap \mathcal{W}^{2/(2s+1)}(\mathbf{H}),$$

then there exists a constant $C_2 > 0$ such that

$$\mathbb{E}\left(\|\widehat{f} - f\|^2\right) \le C_2 \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

This rate of convergence corresponds to the near optimal one in the "standard" minimax setting (see, e.g., [27]). Moreover, applying [[49], Theorem 3.2], one can see that $\mathcal{B}_{\infty}^{\theta/2}(\mathbf{H}) \cap \mathcal{W}^{2(1-\theta)}(\mathbf{H})$ is the "maxiset" associated to $\widehat{f}(\cdot)$ at the rate of convergence $(\ln n/n)^{\theta}$, i.e.,

$$\lim_{n\to\infty} \left(\frac{n}{\ln n}\right)^{\theta} \mathbb{E}\left(\|\widehat{f} - f\|^2\right) < \infty \Leftrightarrow f \in \mathcal{B}_{\infty}^{\theta/2}(\mathbf{H}) \cap \mathcal{W}^{2(1-\theta)}(\mathbf{H}).$$

4. Problem Definition of the Regression Estimation

For a measurable function $\rho : \mathbb{R}^q \to \mathbb{R}$, we define the regression function $\mathbf{m}(\cdot, \rho)$ by

$$\rho(Y) = \mathbf{m}(\mathbf{X}, \rho) + \epsilon, \tag{19}$$

where ϵ is a random variable independent of $\mathbf{X}$ with $\mathcal{N}(0,1)$. We suppose that $\mathbf{m}(\cdot,\rho) \in \mathbf{H}$, where $\mathbf{H}$ is a separable Hilbert space of real or complex-valued functions defined on $\mathbf{S}$ and square integrable with respect to the σ-finite measure ν. We shall suppose that there exist a known constant and $C_{\mathbf{m}} > 0$ such that

$$\sup_{\mathbf{x}\in\mathbf{S}} \quad \mathbf{m}(\mathbf{x},\rho) \leq C_{\mathbf{m}}. \tag{20}$$

In this context, we redefine the probability measure $\mathbb{P}_{\mathbf{X}}$ in (11) and suppose that $f(\cdot)$ is a non-negative measurable **known** function.

(M.1) We shall suppose that there exist two known constant $C_{\mathbf{m}} > 0$ such that

$$\sup_{\mathbf{x}\in\mathbf{S}} \mathbf{m}(\mathbf{x};\rho) \leq C_{\mathbf{m}}.$$

(M.2) We shall suppose that there exist two known constant $c_f > 0$ such that

$$\inf_{\mathbf{x}\in\mathbf{S}} f(\mathbf{x}) \geq c_f.$$

Regression Function Estimator

In this context, we aim to estimate $\mathbf{m}(\cdot,\rho)$ based on n observed functional data $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$. The kernel estimator of the regression function for functional data was proposed by [57]

$$\widehat{\mathbf{m}}_{n;h_n}(\mathbf{x},\rho) := \frac{\sum_{i=1}^{n} \rho(Y_i) K(d(\mathbf{x},\mathbf{X}_i)/h_n)}{\sum_{i=1}^{n} K(d(\mathbf{x},\mathbf{X}_i)/h_n)}.$$

As combined with the work of [55], we define the wavelet hard thresholding estimator $\widehat{\mathbf{m}}(\cdot,\rho)$, for all $\mathbf{x} \in \mathbf{S}$, by

$$\widehat{\mathbf{m}}(\mathbf{x},\rho) = \sum_{\ell\in\mathcal{I}_0} \widehat{\eta}_{0,\ell}\phi_0(\mathbf{x};\zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell\in\mathcal{J}_k} \widehat{\theta}_{k,\ell} \mathbb{1}_{\left\{ |\widehat{\theta}_{k,\ell}| \geq \kappa\sqrt{\frac{\ln n}{n}} \right\}} \psi_k(\mathbf{x};\eta_{k;\ell}), \tag{21}$$

where

$$\widehat{\eta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^{n} \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \phi_k(\mathbf{X}_i;\zeta_{k,\ell}), \tag{22}$$

$$\widehat{\theta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^{n} \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \psi_k(\mathbf{X}_i;\eta_{k,\ell}). \tag{23}$$

where κ is large enough constant and m_n is the integer satisfying

$$\frac{1}{2}\frac{n}{(\ln n)^2} \leq |\mathcal{J}_{m_n}| \leq \frac{n}{(\ln n)^2}. \tag{24}$$

Theorem 2. *Under the conditions* **(E.1)**, **(E.2) (M.1)–(M.2)**, **(C.0)** *and* **(C.1)**, *combined with the assumption* (24), *for any* $\theta \in (0,1)$, $\mathbf{m}(\cdot,\rho) \in \mathcal{B}_{\infty}^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$, *there exists a constant* $C_3 > 0$ *in such a way that*

$$\mathbb{E}\left(\|\widehat{\mathbf{m}}(\cdot,\rho) - \mathbf{m}(\cdot;\rho)\|^2 \right) \leq C \left(\frac{\ln n}{n} \right)^{\theta}, \tag{25}$$

for large enough n.

Suppose that $\mathbf{m}(\cdot,\rho)$ and f satisfy **(M.1)** and, for any $\theta \in (0,1)$, $\mathbf{m}(\cdot,\rho) \in \mathcal{B}_{\infty}^{\theta/2}(\mathbf{H}) \cap \mathcal{W}^{2(1-\theta)}(\mathbf{H})$, where $\mathcal{B}_{\infty}^{\theta/2}(\mathbf{H})$ is in (Definition 1) with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(\mathbf{H})$ is in (Definition 2) with $r = 2(1-\theta)$. Then, there exists a constant $C_3 > 0$ such that

$$\mathbb{E}\Big(\|\widehat{\mathbf{m}}(\cdot,\rho) - \mathbf{m}(\cdot;\rho)\|^2\Big) \leq C_3\left(\frac{(\ln n)^2}{n}\right)^{\theta}$$

for n large enough. Again, note that, for $s > 0$, if

$$\mathbf{m}(\cdot,\rho) \in \mathcal{B}_{\infty}^{s/(2s+1)}(\mathbf{H}) \cap \mathcal{W}^{2/(2s+1)}(\mathbf{H}),$$

then there exists a constant $C_4 > 0$ such that

$$\mathbb{E}\Big(\|\widehat{\mathbf{m}}(\cdot,\rho) - \mathbf{m}(\cdot;\rho)\|^2\Big) \leq C_4\left(\frac{(\ln n)^2}{n}\right)^{2s/(2s+1)}.$$

Up to an additional logarithmic term, this rate of convergence corresponds to the near-optimal one in the "standard" minimax setting (see, for example, [27]). Theorem 2 is the first to investigate an adaptive wavelet-based estimator for functional data in the context of nonparametric regression for ergodic processes.

Since the coefficients defined by (22) and (23) depend on the unknown function $f(\cdot)$, one can use

$$\widetilde{\eta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\widehat{f}(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell}), \tag{26}$$

$$\widetilde{\theta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{\widehat{f}(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell}). \tag{27}$$

We define the wavelet hard thresholding estimator $\widetilde{\mathbf{m}}(\cdot,\rho)$, for all $\mathbf{x} \in \mathbf{S}$, by

$$\widetilde{\mathbf{m}}(\mathbf{x},\rho) = \sum_{\ell\in\mathcal{I}_0}\widetilde{\eta}_{0,\ell}\phi_0(\mathbf{x};\zeta_{0,\ell}) + \sum_{k=0}^{m_n}\sum_{\ell\in\mathcal{J}_k}\widetilde{\theta}_{k,\ell}\mathbb{1}_{\left\{|\widetilde{\theta}_{k,\ell}| \geq \kappa\sqrt{\frac{\ln n}{n}}\right\}}\psi_k(\mathbf{x};\eta_{k,\ell}). \tag{28}$$

Recall the following elementary observation

$$\frac{1}{\widehat{f}(\cdot)} = \frac{1}{f(\cdot)} + \frac{(f(\cdot)-\widehat{f}(\cdot))}{f(\cdot)\widehat{f}(\cdot)}.$$

From the last equation, we can infer the following:

$$\widetilde{\eta}_{k,\ell} = \widehat{\eta}_{k,\ell} + \frac{1}{n}\sum_{i=1}^{n}\frac{(f(\mathbf{X}_i)-\widehat{f}(\mathbf{X}_i))}{f(\mathbf{X}_i)\widehat{f}(\mathbf{X}_i)}\rho(Y_i)\phi_k(\mathbf{X}_i;\zeta_{k,\ell}),$$

$$\widetilde{\theta}_{k,\ell} = \widehat{\theta}_{k,\ell} + \frac{1}{n}\sum_{i=1}^{n}\frac{(f(\mathbf{X}_i)-\widehat{f}(\mathbf{X}_i))}{f(\mathbf{X}_i)\widehat{f}(\mathbf{X}_i)}\rho(Y_i)\psi_k(\mathbf{X}_i;\eta_{k,\ell}).$$

By combining Theorem 1 with Theorem 2, we obtain the following corollary.

Corollary 1. *Under the conditions of Theorems 1 and 2, there exists a constant $C_5 > 0$ such that*

$$\mathbb{E}\Big(\|\widetilde{\mathbf{m}}(\mathbf{x};\rho) - \mathbf{m}(\mathbf{x};\rho)\|^2\Big) \leq C_5\left(\frac{\ln n}{n}\right)^{\theta}, \tag{29}$$

for large enough n.

Remark 2. *In our previous paper [40], we were concerned with the nonparametric estimation of the density and the regression function in a finite-dimensional setting using orthonormal wavelet bases. Our findings differ significantly from those presented in the present paper. In [40], we provided the strong uniform consistency properties with rates of these estimators, over compact subsets of $\mathbb{R}^d$, under a general ergodic condition on the underlying processes. We also establish the asymptotic normality of wavelet-based estimators. We used the Burkholder–Rosenthal inequality as the main ingredient in this paper, which is a more complicated tool than the exponential inequality used in the previous paper. More importantly, in the present paper, we look into the mean integrated square error over compact subsets, which is entirely different from the results of the previous paper.*

5. Applications

5.1. The Conditional Distribution

Our result can be used to investigate the conditional distribution $F(y \mid x)$ for $y \in \mathbb{R}^d$. To be more precise, let $\rho(y) = \mathbb{1}\{y \leq t\}$. We define the wavelet hard thresholding estimator $\breve{\mathbf{m}}(\cdot, \rho)$, for all $\mathbf{x} \in \mathbf{S}$, by

$$\widehat{F}(y \mid x) = \sum_{\ell \in \mathcal{I}_0} \breve{\eta}_{0,\ell} \phi_0(\mathbf{x}; \zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \breve{\theta}_{k,\ell} \mathbb{1}_{\left\{ |\breve{\theta}_{k,\ell}| \geq \kappa \sqrt{\frac{\ln n}{n}} \right\}} \psi_k(\mathbf{x}; \eta_{k;\ell}), \tag{30}$$

where

$$\breve{\eta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{1}\{Y_i \leq t\}}{f(\mathbf{X}_i)} \phi_k(\mathbf{X}_i; \zeta_{k,\ell}), \tag{31}$$

$$\breve{\theta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{1}\{Y_i \leq t\}}{f(\mathbf{X}_i)} \psi_k(\mathbf{X}_i; \eta_{k,\ell}). \tag{32}$$

A direct consequence of Theorem 2 is

$$\mathbb{E}\left(\left\| \widehat{F}(y \mid x) - F(y \mid x) \right\|^2 \right) \leq C \left(\frac{\ln n}{n} \right)^{\theta}. \tag{33}$$

5.2. The Curve Discrimination

We can state the curve discrimination problem in the following way. Let $\{\mathbf{X}_i\}_{i=1,\ldots,n}$ be a sample of curves, and each of them is known to belong to one among $\mathbf{G}$ groups $\boldsymbol{\iota} = 1, \ldots, \mathbf{G}$. Let us denote by T_i the group of the curve $\mathbf{X}_i$. Assume that each pair of variables $(\mathbf{X}_i, T_i)$ has the same distribution as the pair $(\mathbf{X}, T)$. Given a new curve $\mathbf{x}$, the question is to know its class membership, and for that, we will estimate, for any $\boldsymbol{\iota} = 1, \ldots, \mathbf{G}$, the conditional probability:

$$p_{\boldsymbol{\iota}}(x) = \mathbb{P}(T = \boldsymbol{\iota} \mid \mathbf{X} = \mathbf{x}).$$

Following the idea proposed in [58,59] permitting the estimation of these probabilities by

$$\widehat{p}_{\boldsymbol{\iota}}(x) = \sum_{\ell \in \mathcal{I}_0} \breve{\breve{\eta}}_{0,\ell} \phi_0(\mathbf{x}; \zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \breve{\breve{\theta}}_{k,\ell} \mathbb{1}_{\left\{ |\breve{\breve{\theta}}_{k,\ell}| \geq \kappa \sqrt{\frac{\ln n}{n}} \right\}} \psi_k(\mathbf{x}; \eta_{k;\ell}), \tag{34}$$

where

$$\breve{\breve{\eta}}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}\{T_i = \iota\}}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell}), \tag{35}$$

$$\breve{\breve{\theta}}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}\{T_i = \iota\}}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell}). \tag{36}$$

A direct consequence of Theorem 2 is

$$\mathbb{E}\left(\left\|\widehat{F}(y \mid x) - F(y \mid x)\right\|^2\right) \leq C\left(\frac{\ln n}{n}\right)^{\theta}. \tag{37}$$

As remarked by [58,59], for each ι, we make use of the notation

$$Y = \begin{cases} 1 & \text{if } T = \iota \\ 0 & \text{otherwise}, \end{cases}$$

then, we can write

$$p_\iota(x) = \mathbb{E}(Y \mid X = x).$$

An application of Theorem 2 is

$$\mathbb{E}\left(\|\widehat{p}_\iota(x) - p_\iota(x\|^2\right) \leq C\left(\frac{\ln n}{n}\right)^{\theta}.$$

5.3. Time Series Prediction from Continuous Set of Past Values

A direct consequence of our results is the prediction of future values of some real time series, which we will follow from [59]. One of the purposes of the proposed functional method is to predict the future from a continuous set of past values of the process. Let $\{Z(t)\}_{t\in\mathbb{R}}$ denote a real-valued process and s denote a fixed non-negative real number. We are interested in the prediction's problem of a future value $Z(\tau)$ given some past values $Z(t)$ for $\tau - T \leq t < \tau$, at some time $\tau > 0$. Our goal can be seen as the estimation of the operator r:

$$Z(\tau + s) = r(Z(t) \text{ for } \tau - T \leq t < \tau) + \varepsilon,$$

whenever it exists. Let us describe the model. Suppose that the process has been observed from $t = 0$ until $t = t_{\max}$ and without loss of generality, assume that

$$t_{\max} = nT + s < \tau.$$

The methodology consists of splitting the observed process into n pieces of fixed length. Each piece of the process is denoted by

$$\mathbf{X}_i = \{Z(t), (i-1)T \leq t < iT\}.$$

Let us denote the response value $Y_i = Z(iT + s)$. This can be formulated by a regression problem in the following way:

$$Y_i = \mathbf{m}(\mathbf{X}_i, Id) + \varepsilon_i \quad \text{for } i = 1, \ldots, n,$$

where Id denotes the identity function. To fit the theoretical setting of the present paper, we assume that such a function $\mathbf{m}(\cdot, Id)$ does not depend on i; this is the case for the stationary processes. Hence, at time τ, we can use for predicting the value at time $\tau + s$ the following predictor, which is directly derived from (21) :

$$\widehat{\mathbf{m}}(\mathbf{x}, Id) = \sum_{\ell \in \mathcal{I}_0} \widehat{\eta}_{0,\ell}\phi_0(\mathbf{x};\zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \widehat{\theta}_{k,\ell} \mathbb{1}_{\left\{|\widehat{\theta}_{k,\ell}| \geq \kappa\sqrt{\frac{\ln n}{n}}\right\}} \psi_k(\mathbf{x};\eta_{k;\ell}), \tag{38}$$

where

$$\widehat{\eta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z(iT+s)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell}), \tag{39}$$

$$\widehat{\theta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z(iT+s)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell}), \tag{40}$$

where $\mathbf{x} = \{Z(t), \text{for } \tau - T \leq t < \tau\}$. Our Theorem 2 gives mathematical support to this nonparametric functional predictor, and provides a different way of solving the prediction problem investigated in [3,59].

6. Concluding Remarks

In this work, we have investigated the nonparametric estimation of the density and the regression function based on the functional stationary processes, using wavelet bases for Hilbert spaces of functions. We have characterized the mean integrated square error over compact subsets. The asymptotic properties of these estimators are obtained employing the Martingale approach, which is completely different from the mixing and the independent setting. The assumption on the dependence of the process is ergodicity. To motivate the present paper, we have presented the conditional distribution, the curve discrimination and the time series prediction from a continuous set of past values. Extending the nonparametric functional ideas to the local stationary process is a somewhat underdeveloped field. It would be interesting to extend our work to the case of the functional local stationary process, which requires nontrivial mathematics; this would go well beyond the scope of the present paper.

7. Proofs

7.1. Proof of Theorem 1

In this paper, we need an upper bound inequality for partial sums of unbounded martingale differences that we use to derive the asymptotic results for the density and the regression functions estimates built upon functional strictly stationary and ergodic data. Here and in the sequel, we denote by "C" a positive constant that may be different from line to line. This inequality is given in the following lemmas. This lemma is stated following Notation **1** in [60].

Lemma 1 (Burkholder-Rosenthal inequality). *Let $(X_i)_{i\geq 1}$ be a stationary Martingale adapted to the filtration $(\mathcal{F}_i)_{i\geq 1}$, define $(d_i)_{i\geq 1}$ is the sequence of Martingale differences adapted to $(\mathcal{F}_i)_{i\geq 1}$ and*

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

Then, for any positive integer n,

$$\|\max_{1\leq j\leq n} |S_j|)\|_p \ll n^{1/p}\|d_1\|_p + \left\|\sum_{k=1}^{n} \mathbb{E}(d_k^2/\mathcal{F}_{k-1})\right\|_{p/2}^{1/2}, \quad \textit{for any} \quad p \geq 2; \tag{41}$$

where, as usual, the norm $\|\cdot\|_p = (\mathbb{E}[|\cdot|^p])^{1/p}$.

Lemma 2 ([61])**.** *Let $\{Z_i, i \geq 1\}$ be a sequence of Martingale differences such that*

$$|Z_i| \leq B, \quad a.s.,$$

then, for all $\epsilon > 0$ and all sufficiently large n, we have

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} Z_i\right| > \epsilon\right\} \leq 2\exp\left\{-\frac{\epsilon^2}{2nB^2}\right\}.$$

The following lemmas describe the asymptotic behavior of the estimators $\widehat{\alpha}_{k,\ell}$ and $\widehat{\beta}_{k,\ell}$.

Lemma 3. *For any $k \in \{0, \ldots, m_n\}$ and any $\ell \in \mathcal{I}_k$, under assumptions* **(C.0)**, **(C.1)**, **(F.1)** *and* **(E.1)(i)**, *there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(|\widehat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2\right) \leq C\left(\frac{\ln n}{n}\right). \tag{42}$$

Lemma 4. *For any $k \in \{0, \ldots, m_n\}$ and any $\ell \in \mathcal{J}_k$, and under assumptions* **(C.0)**, **(C.1)**, **(F.1)**, **(E.1)** *and* **(E.2)**, *and condition (16), there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right|^4\right) = C\left(\frac{\ln n}{n}\right)^2, \quad a.s. \tag{43}$$

Lemma 5. *For any $k \in \{0, \ldots, m_n\}$ and any $\ell \in \mathcal{J}_k$, for $\kappa > 0$ large enough and under assumptions* **(C.0)**, **(C.1)**, **(F.1)**, **(E.1)** *and* **(E.2)**, *and condition (16), there exists a constant $C > 0$ such that*

$$\mathbb{P}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq C\left(\frac{\ln n}{n}\right)^2. \tag{44}$$

7.1.1. Proof of Theorem 1

Observe that the proof of Theorem 1 is a direct application of ([49], Theorem 3.1) with $c(n) = (\ln n/n)^{1/2}$, $\sigma_i = 1$, $r = 2$ and the Lemmas 3–5. We adapted and extended the method of demonstration of [33], Theorem 3.1 to the stationary ergodic process. □

7.1.2. Proof of Lemmas

Proof of Lemma 3

Consider the following decomposition

$$\begin{aligned} \widehat{\alpha}_{k,\ell} - \alpha_{k,\ell} &= \widehat{\alpha}_{k,\ell} - \widetilde{\alpha}_{k,\ell} + \widetilde{\alpha}_{k,\ell} - \alpha_{k,\ell} \\ &= A_{k,\ell,1} + A_{k,\ell,2}, \end{aligned} \tag{45}$$

where

$$\widetilde{\alpha}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\phi_k(\mathbf{x}_i; \zeta_{k,\ell})|\mathcal{F}_{i-1}].$$

Under the assumptions **(C.0)** and **(C.2)**, we have

$$\begin{aligned}
\widetilde{\alpha}_{k,\ell} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{S}}\phi_k(\mathbf{x};\zeta_{k,\ell})f^{\mathcal{F}_{i-1}}(\mathbf{x})\nu(d\mathbf{x}) \\
&= \int_{\mathbf{S}}\phi_k(\mathbf{x};\zeta_{k,\ell})\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)\nu(d\mathbf{x}) \\
&= \int_{\mathbf{S}}\phi_k(\mathbf{x};\zeta_{k,\ell})(f(\mathbf{x})+o(1))\nu(d\mathbf{x}) \\
&= \int_{\mathbf{S}}\phi_k(\mathbf{x};\zeta_{k,\ell})f(\mathbf{x})\nu(d\mathbf{x})+o(1) \\
&= \alpha_{k,\ell}+o(1).
\end{aligned}$$

We readily obtain

$$\widetilde{\alpha}_{k,\ell}=\alpha_{k,\ell},\quad \text{as},\quad n\rightarrow\infty, \tag{46}$$

implying that

$$A_{k,\ell,2}=o(1),\quad \text{a.s.} \tag{47}$$

Therefore, we infer that

$$\widehat{\alpha}_{k,\ell}-\alpha_{k,\ell} = A_{k,\ell,1}+o(1),\quad \text{a.s.}$$

Let us now consider the term $A_{k,\ell,1}$. We have

$$\begin{aligned}
A_{k,\ell,1} &= \widehat{\alpha}_{k,\ell}-\widetilde{\alpha}_{k,\ell} \\
&= \frac{1}{n}\sum_{i=1}^{n}(\phi_k(\mathbf{x}_i;\zeta_{k,\ell})-\mathbb{E}[\phi_k(\mathbf{x}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}]) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell}).
\end{aligned}$$

Notice that $(\Phi_k(\mathbf{x}_i;\zeta_{k,\ell}))_{0\leq k\leq m_n}$ is a sequence of Martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_i)_{0\leq k\leq m_n}$. It is obvious, by Lemma 1 inequality, to see that

$$\mathbb{E}\left[|A_{k,\ell,1}|^2\right] = \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell})\right|^2\right].$$

Applying the Burkholder–Rosenthal inequality (1), for $p=2$, we obtain

$$\begin{aligned}
&\left(E\left[\left(\left|\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell})\right|^2\right)\right]\right)^{\frac{1}{2}} \\
\leq\ & n^{1/2}\parallel\Phi_k(\mathbf{x}_1;\zeta_{k,\ell})\parallel_2+\parallel\sum_{i=1}^{n}E\left[\Phi_k^2(\mathbf{x}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right]\parallel_1^{1/2} \\
=\ & \Phi_{(1)}+\Phi_{(2)}
\end{aligned} \tag{48}$$

On one hand, using a very famous decomposition combined with the fact that $\mathcal{F}_0$ is the trivial $\sigma-$field, we obtain

$$
\begin{aligned}
\frac{1}{n}\Phi_{(1)}^2 &= \| \Phi_k(\mathbf{x}_1;\zeta_{k,\ell}) \|_2^2 \\
&= E\Big[\big|\phi_k((\mathbf{x}_1;\zeta_{k,\ell})) - E[\phi_k(\mathbf{x}_1;\zeta_{k,\ell})|\mathcal{F}_0]\big|^2\Big] \\
&\leq E\left[\sum_{j=0}^{2} | \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |^j \, (E[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |])^{2-j}\right] \\
&= \sum_{j=0}^{2} C_2^j E\Big[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |^j\Big].(E[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |])^{2-j} \\
&= C_2^2 E\Big[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |^2\Big] + C_2^1 (E[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |])^2 + C_2^0 E\Big[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |^2\Big], \qquad (49)
\end{aligned}
$$

By the Cauchy–Schwarz inequality, together with assumptions **(E.1)**(i), **(E.2)** and the condition (16), we obtain

$$
\begin{aligned}
\sup_{\mathbf{x}\in\mathbf{S}} \big|\phi_k(\mathbf{x}_i;\zeta_{k,\ell})\big| &\leq \sup_{\mathbf{x}\in\mathbf{S}} \sum_{j\in\mathcal{J}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} |e_j(\zeta_{k,\ell})||e_j(\mathbf{x})| \\
&\leq \left(\sum_{j\in\mathcal{J}_k} \frac{1}{g_{j,k,\ell}} |e_j(\zeta_{k,\ell})|^2\right)^{1/2} \left(\sup_{\mathbf{x}\in\mathbf{S}} \sum_{j\in\mathcal{J}_k} |e_j(\mathbf{x})|^2\right)^{1/2} \\
&\leq C_1^{1/2} C_2^{1/2} \sqrt{|\mathcal{J}_k|} \\
&\leq C_3 \sqrt{|\mathcal{J}_{m_n}|} \\
&\leq C_3 \sqrt{\frac{n}{\ln n}}. \qquad (50)
\end{aligned}
$$

Observe that, under the assumptions **(F.1)** and **(E.1)**(i) and the fact that $\mathcal{E}$ is an orthonormal basis of $\mathbf{H}$, we have

$$
\begin{aligned}
\mathbb{E}\Big[\big|\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\big|^2\Big] &= \int_{\mathbf{S}} \big|\phi_k(\mathbf{x};\zeta_{k,\ell})\big|^2 f(\mathbf{x})\nu(d\mathbf{x}) \\
&\leq C_f \int_{\mathbf{S}} \big|\phi_k(\mathbf{x};\zeta_{k,\ell})\big|^2 \nu(d\mathbf{x}) \\
&= C_f \int_{\mathbf{S}} \left|\sum_{j\in\mathcal{I}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} e_j(\zeta_{k,\ell}) e_j(\mathbf{x})\right|^2 \nu(d\mathbf{x}) \\
&= C_f \int_{\mathbf{S}} \sum_{j\in\mathcal{I}_k} \frac{1}{g_{j,k,\ell}} \big|e_j(\zeta_{k,\ell})\big|^2 \nu(d\mathbf{x}) \\
&\leq C_f C_1. \qquad (51)
\end{aligned}
$$

where C_1 is a positive constant,

$$
E[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |] = O\left(\sqrt{\frac{n}{\ln n}}\right), \qquad (52)
$$

$$
E\Big[| \phi_k(\mathbf{x}_1;\zeta_{k,\ell}) |^2\Big] = O(1) \qquad (53)
$$

therefore,

$$
\Phi_{(1)} = O(n^{1/2}). \qquad (54)
$$

On the other hand, we consider the second term of decomposition (48), observe that

$$\begin{aligned}\Phi_2 &= \left(\mathbb{E}\left(\sum_{i=1}^{n}\mathbb{E}\left[\Phi_k^2(\mathbf{x}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right]\right)\right)^{1/2} \\ &= \left(\sum_{i=1}^{n}\mathbb{E}\left(\mathbb{E}\left[\Phi_k^2(\mathbf{x}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right]\right)\right)^{1/2} \\ &= \left(\sum_{i=1}^{n}\mathbb{E}\left[\Phi_k^2(\mathbf{x}_i;\zeta_{k,\ell})\right]\right)^{1/2}\end{aligned}$$

using the notable identity, we obtain

$$\begin{aligned}&\mathbb{E}\left[\Phi_k^2(\mathbf{x}_i;\zeta_{k,\ell})\right] \\ &= \mathbb{E}\left[\left(|\phi_k(\mathbf{x}_i;\zeta_{k,\ell}) - \mathbb{E}[\phi_k(\mathbf{x}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}]|\right)^2\right] \\ &\leq \mathbb{E}\Big[|\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |^2 + 2\ |\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |\ \mathbb{E}[|\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |\ |\mathcal{F}_{i-1}] \\ &\quad + \mathbb{E}\left[|\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |^2\ |\mathcal{F}_{i-1}\right]\Big] \\ &\leq 2\mathbb{E}\left[|\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |^2\right] + 2\mathbb{E}\left[\mathbb{E}\left[|\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |^2\ |\mathcal{F}_{i-1}\right]\right] \\ &\leq 4\mathbb{E}\left[|\ \phi_k(\mathbf{x}_i;\zeta_{k,\ell})\ |^2\right]\end{aligned}$$

observe that, using (50) and (51), we obtain

$$\Phi_2 = O(n^{1/2}). \tag{55}$$

therefore, we combine (54) and (55) to obtain

$$\left(E\left[\left|\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell})\right|^2\right]\right)^{1/2} = O(n^{1/2}).$$

Hence,

$$\begin{aligned}\mathbb{E}\left[|A_{k,\ell,1}|^2\right] &= \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell})\right|^2\right] \\ &= \frac{1}{n^2}O(n) \\ &\leq C\left(\frac{\ln n}{n}\right)\end{aligned}$$

Therefore, there exists a constant $C = C_f C_1 > 0$, such that

$$\mathbb{E}\left(\left|\widehat{\alpha}_{k,\ell} - \alpha_{k,\ell}\right|^2\right) \leq \frac{4C}{n} \leq 4C\left(\frac{\ln n}{n}\right). \tag{56}$$

Hence, the proof is complete. □

Proof of Lemma 4

Consider the following decomposition

$$\begin{aligned}\widehat{\beta}_{k,\ell} - \beta_{k,\ell} &= \widehat{\beta}_{k,\ell} - \widetilde{\beta}_{k,\ell} + \widetilde{\beta}_{k,\ell} - \beta_{k,\ell} \\ &= B_{k,\ell,1} + B_{k,\ell,2},\end{aligned} \tag{57}$$

where

$$\widetilde{\beta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}].$$

Observe that, under the assumptions **(F.1)** and **(E.1)**(i) and the fact that $\mathcal{E}$ is an orthonormal basis of **H**, and proceeding in a similar way as in (46), we show that

$$\widetilde{\beta}_{k,\ell} = \beta_{k,\ell}, \quad \text{as,} \quad n \to \infty. \tag{58}$$

This, in turn, implies that

$$B_{k,\ell,2} = o(1), \quad \text{a.s.} \tag{59}$$

Therefore, we obtain

$$\widehat{\beta}_{k,\ell} - \beta_{k,\ell} \quad = \quad B_{k,\ell,1} + o(1), \quad \text{a.s.}$$

Hence, we readily infer

$$\mathbb{E}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right|^4\right) \quad = \quad \frac{1}{n^4}\mathbb{E}\left(\left|\sum_{i=1}^{n} \Psi_{i,k,\ell}\right|^4\right), \tag{60}$$

where

$$\Psi_{i,k,\ell} = \psi_k(\mathbf{x}_i;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}].$$

Notice that $(\Psi_{i,k,\ell})_{0\leq k\leq n}$ is a sequence of Martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_i)_{0\leq k\leq n}$, applying the Burkholder–Rosenthal inequality for $p=4$ (see Lemma 1), we obtain

$$\begin{aligned} \left(\mathbb{E}\left(\left|\sum_{i=1}^{n} \Psi_{i,k,\ell}\right|^4\right)\right)^{1/4} \quad &\leq \quad \left\|\max_{1\leq j\leq n}\left|\sum_{i=1}^{j} \Psi_{i,k,\ell}\right|\right\|_4 \\ &\ll \quad n^{1/4}\|\Psi_{1,k,\ell}\|_4 + \left\|\sum_{i=1}^{n} \mathbb{E}(\Psi^2_{i,k,\ell}|\mathcal{F}_{i-1})\right\|^{1/2}_{4/2} \\ &= \quad \Psi^{(1)}_{k,\ell} + \Psi^{(2)}_{k,\ell}. \end{aligned} \tag{61}$$

Consider the first term of Equation (61). We have

$$\begin{aligned} \frac{1}{n}\left(\Psi^{(1)}_{k,\ell}\right)^4 \quad &= \quad \|\Psi_{1,k,\ell}\|_4^4 \\ &= \quad \mathbb{E}\left(\left|\psi_k(\mathbf{x}_1;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_1;\eta_{k,\ell})|\mathcal{F}_0]\right|^4\right) \\ &\leq \quad \mathbb{E}\left[\left(\left|\psi_k(\mathbf{x}_1;\eta_{k,\ell})\right| + \mathbb{E}\left[\left|\psi_k(\mathbf{x}_1;\eta_{k,\ell})\right|\right]\right)^4\right]. \end{aligned}$$

Using the classical identity

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}$$

in connection with the Jensen inequality and taking $n = 4$, we obtain

$$\begin{aligned}&\left(|\psi_k(\mathbf{x}_1;\eta_{k,\ell})| + \mathbb{E}[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|]\right)^4\\ &\quad = \sum_{k=0}^{4} C_4^k |\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^k \left(\mathbb{E}[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|]\right)^{4-k}\\ &\quad \leq \sum_{k=0}^{4} C_4^k |\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^k \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^{4-k}\right].\end{aligned}$$

This gives that

$$\begin{aligned}\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 &\leq \mathbb{E}\left[\sum_{k=0}^{4} C_4^k |\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^k \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^{4-k}\right]\right]\\ &= \sum_{k=0}^{4} C_4^k \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^k\right] \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^{4-k}\right]. \qquad (62)\end{aligned}$$

By proceeding in a similar way as in (51) and making use of the assumptions **(F.1)** and **(E.1)**(i), we infer that

$$\mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^2\right] \leq C, \qquad (63)$$

where C is a positive constant. Moreover, by the Cauchy–Schwarz inequality together with assumptions **(E.1)**(ii), **(E.2)** and the condition (16), we obtain

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbf{S}} |\psi_k(\mathbf{x};\eta_{k,\ell})| &\leq \sup_{\mathbf{x}\in\mathbf{S}} \sum_{j\in\mathcal{J}_k} \frac{1}{\sqrt{h_{j,k,\ell}}} |e_j(\eta_{k,\ell})||e_j(\mathbf{x})|\\ &\leq \left(\sum_{j\in\mathcal{J}_k} \frac{1}{h_{j,k,\ell}} |e_j(\eta_{k,\ell})|^2\right)^{1/2} \left(\sup_{\mathbf{x}\in\mathbf{S}} \sum_{j\in\mathcal{J}_k} |e_j(\mathbf{x})|^2\right)^{1/2}\\ &\leq C_1^{1/2} C_2^{1/2} \sqrt{|\mathcal{J}_k|}\\ &\leq C_4 \sqrt{|\mathcal{J}_{m_n}|}\\ &\leq C_4 \sqrt{\frac{n}{\ln n}}. \qquad (64)\end{aligned}$$

We then obtain

$$\begin{aligned}\mathbb{E}[\,|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|\,] &= O\left(\sqrt{\frac{n}{\ln n}}\right), \qquad (65)\\ \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^2\right] &= O(1), \qquad (66)\\ \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^3\right] &= O\left(\sqrt{\frac{n}{\ln n}}\right), \qquad (67)\\ \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|^4\right] &= O\left(\frac{n}{\ln n}\right). \qquad (68)\end{aligned}$$

Observe that the largest term is (68); now, using (68) in Equation (62), we deduce that

$$\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 = O\left(\frac{n}{\ln n}\right).$$

This implies that

$$\Psi_{k,\ell}^{(1)} = O\left(\frac{n^{1/2}}{(\ln n)^{1/4}}\right). \qquad (69)$$

Let us now investigate the upper bound of $\Psi^{(2)}_{k,\ell}$ in (61). Observe that

$$\begin{aligned}\Psi^{(2)}_{k,\ell} &= \left\|\sum_{i=1}^{n}\mathbb{E}(\Psi^2_{i,k,\ell}/\mathcal{F}_{i-1})\right\|_2^{1/2}\\ &= \left(\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{E}\left[\Psi^2_{i,k,\ell}|\mathcal{F}_{i-1}\right]\right)^2\right]\right)^{1/4},\end{aligned}$$

for all $i = 1,\ldots,n$. Making use of the Jensen inequality with the fact that $(a-b)^2 = a^2 - 2ab + b^2$, it follows that

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[\Psi^2_{i,k,\ell}|\mathcal{F}_{i-1}\right] &= \sum_{i=1}^{n}\left(\mathbb{E}\left[(\psi_k(\mathbf{x}_i;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}])^2|\mathcal{F}_{i-1}\right]\right)\\ &\leq 4\sum_{i=1}^{n}\mathbb{E}\left[(\psi_k(\mathbf{x}_i;\eta_{k,\ell}))^2|\mathcal{F}_{i-1}\right].\end{aligned}$$

Observe, under the assumptions **(F.1)**, **(E.1)**(i), **(C.0)** and **(C.1)** and (63):

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[(\psi_k(\mathbf{x}_i;\eta_{k,\ell}))^2|\mathcal{F}_{i-1}\right] &= n\int_{\mathbf{S}}\left|\psi_k(\mathbf{x};\eta_{k,\ell})\right|^2\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)\nu(d\mathbf{x})\\ &= n\int_{\mathbf{S}}\left|\psi_k(\mathbf{x};\eta_{k,\ell})\right|^2(f(\mathbf{x})+o(1))\nu(d\mathbf{x})\\ &\leq n\left(C_f+o(1)\right)\int_{\mathbf{S}}\left|\psi_k(\mathbf{x};\zeta_{k,\ell})\right|^2\nu(d\mathbf{x})\\ &\leq nC_fC_1.\end{aligned} \tag{70}$$

It follows that

$$\Psi^{(2)}_{k,\ell} = O\left(n^{1/2}\right). \tag{71}$$

Combining (61), (69) and (71), we obtain

$$\mathbb{E}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right|^4\right) = O\left(\frac{n^2}{\ln n}\right) + O\left(n^2\right).$$

We conclude that

$$\mathbb{E}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right|^4\right) = O\left(\frac{1}{n^2\ln n}\right) + O\left(\frac{1}{n^2}\right). \tag{72}$$

This implies that there exists a constant $C > 0$, such that

$$\mathbb{E}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right|^4\right) \leq C\left(\frac{\ln n}{n}\right)^2. \tag{73}$$

The proof is achieved. □

Proof of Lemma 5

Consider the previous decomposition in Lemma 4, to write that

$$\begin{aligned}\widehat{\beta}_{k,\ell} - \beta_{k,\ell} &= \left(\widehat{\beta}_{k,\ell} - \widetilde{\beta}_{k,\ell}\right) + \left(\widetilde{\beta}_{k,\ell} - \beta_{k,\ell}\right)\\ &= B_{k,\ell,1} + B_{k,\ell,2},\end{aligned}$$

where

$$
\begin{aligned}
B_{k,\ell,1} &= \frac{1}{n}\sum_{i=1}^{n}\Psi_{i,k,\ell} = \frac{1}{n}\sum_{i=1}^{n}(\psi_k(\mathbf{x}_i;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}]), \\
B_{k,\ell,2} &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}] - \beta_{k,\ell}.
\end{aligned}
$$

By using (59), we obtain the desired result for the term $B_{k,\ell,2}$

$$
\widehat{\beta}_{k,\ell} - \beta_{k,\ell} = B_{k,\ell,1} + o(1).
$$

Now, observe that

$$
\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\Psi_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{n\ln n}\right).
$$

An application of Lemma 2 implies that

$$
\begin{aligned}
|\Psi_{i,k,\ell}| &= |\psi_k(\mathbf{x}_i;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}]| \\
&\leq 2\sup_{\mathbf{x}\in\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})| \\
&\leq C_3\sqrt{\frac{n}{\ln n}} \\
&\leq C_3\sqrt{n}. \qquad (74)
\end{aligned}
$$

Let $B = C_3\sqrt{n}$. Then, for all $\epsilon_n = \frac{\kappa}{2}\sqrt{n\ln n}$ where n is sufficiently large, we have

$$
\begin{aligned}
\mathbb{P}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{n\ln n}\right) &\leq 2\exp\left\{-\frac{\epsilon_n^2}{2nB^2}\right\} \\
&= 2\exp\left\{-\frac{\left(\frac{\kappa}{2}\sqrt{n\ln n}\right)^2}{2n\left(C_3\sqrt{n}\right)^2}\right\} \\
&= 2\exp\left\{-\frac{\kappa^2\ln n}{8C_3^2 n}\right\} \\
&= 2\exp\left\{\ln n^{-\frac{\kappa^2}{8C_3^2 n}}\right\} \\
&= 2n^{-w(\kappa,n)}, \qquad (75)
\end{aligned}
$$

where

$$
w(\kappa,n) = \frac{\kappa^2}{8C_3^2 n}.
$$

By choosing κ such that $w(n,\kappa) = 2$, we have

$$
\begin{aligned}
\mathbb{P}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) &\leq C\frac{1}{n^2} + o(1) \\
&\leq C\left(\frac{\ln n}{n}\right)^2. \qquad (76)
\end{aligned}
$$

The proof of (44) is achieved. □

7.1.3. Proof of Theorem 2

Recall that

$$\widehat{\mathbf{m}}(\mathbf{x},\rho) = \sum_{\ell\in\mathcal{I}_0} \widehat{\eta}_{0,\ell}\phi_0(\mathbf{x};\zeta_{0,\ell}) + \sum_{k=0}^{m_n}\sum_{\ell\in\mathcal{J}_k} \widehat{\theta}_{k,\ell}\mathbb{1}_{\left\{|\widehat{\theta}_{k,\ell}|\geq\kappa\sqrt{\frac{\ln n}{n}}\right\}}\psi_k(\mathbf{x};\eta_{k;\ell}),$$

where

$$\widehat{\eta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell}), \quad \widehat{\theta}_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell}).$$

Lemma 6. *For any $k\in\{0,\ldots,m_n\}$ and any $\ell\in\mathcal{I}_k$ and under the assumptions* **(E.1)(i)**, **(M.1)-(M.2)**, **(C.0)** *and* **(C.1)**, *there exists a constant $C>0$ such that*

$$\mathbb{E}\left(|\widehat{\eta}_{k,\ell}-\eta_{k,\ell}|^2\right) \leq C\left(\frac{\ln n}{n}\right). \tag{77}$$

Lemma 7. *For any $k\in\{0,\ldots,m_n\}$ and any $\ell\in\mathcal{J}_k$, and under the assumptions* **(E.1)**, **(E.2)** **(M.1)-(M.2)**, **(C.0)** *and* **(C.1)**, *combined with the condition (24), there exists a constant $C>0$ such that*

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right) = C\left(\frac{\ln n}{n}\right)^2, \quad a.s. \tag{78}$$

Lemma 8. *For any $k\in\{0,\ldots,m_n\}$ and any $\ell\in\mathcal{J}_k$, for $\kappa>0$ large enough,* **(E.1)**, **(E.2)** **(M.1)–(M.2)**, **(C.0)** *and* **(C.1)**, *combined with the condition (24), there exists a constant $C>0$ such that*

$$\mathbb{P}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq C\left(\frac{\ln n}{n}\right)^2. \tag{79}$$

7.1.4. Proof of Theorem 2

Observe that the proof of Theorem 2 is a direct application of ([49], Theorem 3.1) with $c(n) = (\ln n/n)^{1/2}$, $\sigma_i = 1$, $r = 2$ and the following Lemmas 6–8. We extended the method of the proof in [33], Theorem 4.1. □

7.1.5. Proof of Lemmas

Proof of Lemma 6

Consider the following decomposition

$$\begin{aligned}\widehat{\eta}_{k,\ell}-\eta_{k,\ell} &= \widehat{\eta}_{k,\ell}-\widetilde{\eta}_{k,\ell}+\widetilde{\eta}_{k,\ell}-\eta_{k,\ell} \\ &= A_{k,\ell,1}+A_{k,\ell,2},\end{aligned} \tag{80}$$

where

$$\begin{aligned}\widetilde{\eta}_{k,\ell} &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{(\mathbf{m}(\mathbf{X}_i,\rho)+\epsilon_i)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbf{m}(\mathbf{X}_i,\rho)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\epsilon_i}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right].\end{aligned}$$

From the independence between ϵ_i and $\mathbf{X}_i$, we have

$$\begin{aligned}\mathbb{E}[\epsilon_i|\mathcal{G}_{i-1}] &= \mathbb{E}[\epsilon_i|\mathbf{X}_i] \\ &= \mathbb{E}[\epsilon_i] \\ &= 0. \end{aligned} \tag{81}$$

Observe that

$$\begin{aligned}\mathbb{E}\left[\frac{\epsilon}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] &= \mathbb{E}\left[\frac{\mathbb{E}[\epsilon|\mathcal{G}_{i-1}]}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[\epsilon|\mathbf{X}_i]}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[\epsilon]}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &= 0.\end{aligned}$$

This implies that

$$\begin{aligned}\widetilde{\eta}_{k,\ell} &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbf{m}(\mathbf{X}_i,\rho)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right].\end{aligned}$$

Making use of the assumptions **(M.1)–(M.2)**, **(C.0)** and **(C.1)**, we have

$$\begin{aligned}\widetilde{\eta}_{k,\ell} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{S}}\frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})f^{\mathcal{F}_{i-1}}(\mathbf{x})\nu(d\mathbf{x}) \\ &= \int_{\mathbf{S}}\frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)\nu(d\mathbf{x}) \\ &= \int_{\mathbf{S}}\frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})(f(\mathbf{x})+o(1))\nu(d\mathbf{x}) \\ &= \int_{\mathbf{S}}\frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})f(\mathbf{x})\nu(d\mathbf{x})+o(1) \\ &= \eta_{k,\ell}+o(1).\end{aligned}$$

We readily obtain that

$$\widetilde{\eta}_{k,\ell} = \eta_{k,\ell}, \quad \text{as}, \quad n \to \infty. \tag{82}$$

implying that

$$A_{k,\ell,2} = o(1), \quad \text{a.s.} \tag{83}$$

Therefore,

$$\widehat{\eta}_{k,\ell} - \eta_{k,\ell} = A_{k,\ell,1} + o(1), \quad \text{a.s.}$$

Let us now turn our attention to the term $A_{k,\ell,1}$ in (80), we have

$$\begin{aligned}A_{k,\ell,1} &= \widehat{\eta}_{k,\ell} - \widetilde{\eta}_{k,\ell} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\phi_k(\mathbf{X}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\right]\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\Phi_k(\mathbf{X}_i;\zeta_{k,\ell}).\end{aligned}$$

Notice that $(\Phi_k(\mathbf{x}_i;\zeta_{k,\ell}))_{0\le k\le m_n}$ is a sequence of Martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_i)_{0\le k\le m_n}$. It is obvious, proceeding as the proof of (56), that

$$\mathbb{E}\Big[|A_{k,\ell,1}|^2\Big] \quad = \quad \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_k(\mathbf{X}_i;\zeta_{k,\ell})\right|^2\right]$$

where

$$\begin{aligned}&\left(\mathbb{E}\left[\left(\left|\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell})\right|^2\right)\right]\right)^{\frac{1}{2}}\\ &\le \quad n^{1/2}\parallel \Phi_k(\mathbf{x}_1;\zeta_{k,\ell})\parallel_2 + \parallel \sum_{i=1}^{n}\mathbb{E}\Big[\Phi_k^2(\mathbf{x}_i;\zeta_{k,\ell})|\mathcal{F}_{i-1}\Big]\parallel_1^{1/2}\\ &= \quad \Phi_{(1)}+\Phi_{(2)}\end{aligned} \tag{84}$$

On the one hand, using a very famous decomposition combined with the fact that $\mathcal{F}_0$ is the trivial $\sigma-$field, we obtain

$$\begin{aligned}\frac{1}{n}\Phi_{(1)}^2 &= \parallel \Phi_k(\mathbf{x}_1;\zeta_{k,\ell})\parallel_2^2\\ &= \mathbb{E}\left[\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\phi_k((\mathbf{x}_1;\zeta_{k,\ell}))-\mathbb{E}\left[\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})|\mathcal{F}_0\right]\right|^2\right]\\ &\le \mathbb{E}\left[\sum_{j=0}^{2}\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|^j\left(\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|\right]\right)^{2-j}\right]\\ &= \sum_{j=0}^{2}C_2^j\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|^j\right].\left(\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|\right]\right)^{2-j}\\ &= C_2^2\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|^2\right]+C_2^1\left(\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|\right]\right)^2\\ &\quad +C_2^0\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{x}_1;\zeta_{k,\ell})\right|^2\right],\end{aligned}$$

It follows, from assumptions **(M.1)** and **(M.2)**, that

$$|\rho(Y_i)|\le C_{\mathbf{m}}+|\epsilon_i|, \tag{85}$$

combined with the independence between $\mathbf{X}_1$ and ϵ_1, $\mathbb{E}[\epsilon_1^2]=1$. Observe that, under assumption **(E.1)**(i) and the fact that $\mathcal{E}$ is an orthonormal basis of $\mathbf{H}$, we have

$$
\begin{aligned}
\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{X}_1;\zeta_{k,\ell})\right|^2\right] &\leq \frac{(C_{\mathbf{m}}^2+1)}{c_f}\mathbb{E}\left[\left|\frac{1}{f(\mathbf{X}_1)}\right||\phi_k(\mathbf{x}_1;\zeta_{k,\ell})|^2\right] \\
&\leq \frac{(C_{\mathbf{m}}^2+1)}{c_f}\int_{\mathbf{S}}\frac{1}{f(\mathbf{x})}|\phi_k(\mathbf{x};\zeta_{k,\ell})|^2 f(\mathbf{x})\nu(d\mathbf{x}) \\
&= \frac{(C_{\mathbf{m}}^2+1)}{c_f}\int_{\mathbf{S}}\left|\sum_{j\in\mathcal{I}_k}\frac{1}{\sqrt{g_{j,k,\ell}}}e_j(\zeta_{k,\ell})e_j(\mathbf{x})\right|^2\nu(d\mathbf{x}) \\
&= \frac{(C_{\mathbf{m}}^2+1)}{c_f}\int_{\mathbf{S}}\sum_{j\in\mathcal{I}_k}\frac{1}{g_{j,k,\ell}}|e_j(\zeta_{k,\ell})|^2\nu(d\mathbf{x}) \\
&\leq \frac{(C_{\mathbf{m}}^2+1)}{c_f}C_1 = O(1).
\end{aligned} \tag{86}
$$

Moreover, from Assumptions **(M.1)** and **(M.2)** and using (52), (85), combined with the independence between $\mathbf{X}_1$ and ϵ_1, $\mathbb{E}[\epsilon_1] = 0$. we have

$$
\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\phi_k(\mathbf{X}_1;\zeta_{k,\ell})\right|\right] = O\left(\sqrt{\frac{n}{\ln n}}\right). \tag{87}
$$

Therefore,

$$
\Phi_{(1)} = O(n^{1/2}). \tag{88}
$$

On the other hand, we consider the second term of decomposition (84), and proceeding as in the proof of (55) and considering (86) and (87)

$$
\Phi_2 = O(n^{1/2}). \tag{89}
$$

therefore, combining (88) and (89) to obtain

$$
\left(E\left[\left|\sum_{i=1}^{n}\Phi_k(\mathbf{X}_i;\zeta_{k,\ell})\right|^2\right]\right)^{1/2} = O(n^{1/2}).
$$

Hence,

$$
\begin{aligned}
\mathbb{E}\left[|A_{k,\ell,1}|^2\right] &= \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_k(\mathbf{x}_i;\zeta_{k,\ell})\right|^2\right] \\
&= \frac{1}{n^2}O(n) \\
&\leq C\left(\frac{\ln n}{n}\right)
\end{aligned} \tag{90}
$$

Therefore, combining (83) and (90), there exists a constant C such as

$$
\mathbb{E}\left(|\widehat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2\right) \leq \frac{C}{n} \leq C\left(\frac{\ln n}{n}\right). \tag{91}
$$

Hence, the proof is complete. □

Proof of Lemma 7

Consider the following decomposition

$$\begin{aligned}\widehat{\theta}_{k,\ell}-\theta_{k,\ell} &= \widehat{\theta}_{k,\ell}-\breve{\theta}_{k,\ell}+\breve{\theta}_{k,\ell}-\theta_{k,\ell} \\ &= B_{k,\ell,1}+B_{k,\ell,2},\end{aligned} \tag{92}$$

where

$$\breve{\theta}_{k,\ell}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}\right].$$

Observe that, under the assumptions **(M.1)–(M.2)** and **(C.1)**, the Equation (81) and the fact that $\mathcal{E}$ is an orthonormal basis of $\mathbf{H}$, by proceeding as in (82), we show that

$$\breve{\theta}_{k,\ell}=\theta_{k,\ell},\quad \text{as},\quad n\rightarrow\infty, \tag{93}$$

implying that

$$B_{k,\ell,2}=o(1),\quad \text{a.s.} \tag{94}$$

Therefore,

$$\widehat{\theta}_{k,\ell}-\theta_{k,\ell} = B_{k,\ell,1}+o(1),\quad \text{a.s.}$$

Hence, we obtain

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right) = \frac{1}{n^4}\mathbb{E}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right|^4\right), \tag{95}$$

where

$$\Psi_{i,k,\ell}=\frac{Y_i}{f(\mathbf{x}_i)}\psi_k(\mathbf{x}_i;\eta_{k,\ell})-\mathbb{E}\left[\frac{Y_i}{f(\mathbf{x}_i)}\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}\right].$$

Notice that $(\Psi_{i,k,\ell})_{0\leq k\leq n}$ is a sequence of Martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_i)_{0\leq k\leq n}$, applying the Burkholder–Rosenthal inequality for $p=4$ (see Lemma 1), we obtain

$$\begin{aligned}\left(\mathbb{E}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right|^4\right)\right)^{1/4} &\leq \left\|\max_{1\leq j\leq n}\left|\sum_{i=1}^{j}\Psi_{i,k,\ell}\right|\right\|_4 \\ &\ll n^{1/4}\|\Psi_{1,k,\ell}\|_4+\left\|\sum_{i=1}^{n}\mathbb{E}(\Psi^2_{i,k,\ell}|\mathcal{F}_{i-1})\right\|^{1/2}_{4/2} \\ &= \Psi^{(1)}_{k,\ell}+\Psi^{(2)}_{k,\ell}.\end{aligned} \tag{96}$$

Consider the first term of Equation (96),

$$\begin{aligned}\frac{1}{n}\left(\Psi^{(1)}_{k,\ell}\right)^4 &= \|\Psi_{1,k,\ell}\|_4^4 \\ &= \mathbb{E}\left(\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})-\mathbb{E}\left[\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})|\mathcal{F}_0\right]\right|^4\right) \\ &\leq \mathbb{E}\left[\left(\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|+\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|\right]\right)^4\right].\end{aligned}$$

By combining the identity

$$(a+b)^n=\sum_{k=0}^{n}C_n^k a^k b^{n-k}$$

and the Jensen inequality and taking $n = 4$, we obtain

$$\begin{aligned}
&\left(\left|\frac{\rho(Y_1)}{f(\mathbf{x}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right| + \mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|\right]\right)^4 \\
&= \sum_{k=0}^{4} C_4^k \left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^k \left(\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|\right]\right)^{4-k} \\
&\leq \sum_{k=0}^{4} C_4^k \left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^k \mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^{4-k}\right].
\end{aligned}$$

Then, we have

$$\begin{aligned}
\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 &\leq \mathbb{E}\left[\sum_{k=0}^{4} C_4^k \left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^k \mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^{4-k}\right]\right] \\
&= \sum_{k=0}^{4} C_4^k \mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^k\right] \mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^{4-k}\right]. \qquad (97)
\end{aligned}$$

By following the same reasoning as in (86) and under the same assumptions **(M.1)**-**(M.2)** and **(E.1)**(i), we have

$$\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^2\right] \leq C, \qquad (98)$$

where C is a positive constant. Moreover, by the Cauchy–Schwarz inequality together with assumptions **(E.1)**(ii), **(E.2)** and condition (24), we obtain

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathbf{S}}\left|\psi_k(\mathbf{x};\eta_{k,\ell})\right| &\leq \sup_{\mathbf{x}\in\mathbf{S}}\sum_{j\in\mathcal{J}_k}\frac{1}{\sqrt{h_{j,k,\ell}}}|e_j(\eta_{k,\ell})||e_j(\mathbf{x})| \\
&\leq \left(\sum_{j\in\mathcal{J}_k}\frac{1}{h_{j,k,\ell}}|e_j(\eta_{k,\ell})|^2\right)^{1/2}\left(\sup_{\mathbf{x}\in\mathbf{S}}\sum_{j\in\mathcal{J}_k}|e_j(\mathbf{x})|^2\right)^{1/2} \\
&\leq C_1^{1/2}C_2^{1/2}\sqrt{|\mathcal{J}_k|} \\
&\leq C_3\sqrt{|\mathcal{J}_{m_n}|} \\
&\leq C_3\sqrt{\frac{n}{(\ln n)^2}}. \qquad (99)
\end{aligned}$$

Hence, we infer that

$$\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|\right] = O\left(\sqrt{\frac{n}{(\ln n)^2}}\right), \qquad (100)$$

$$\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^2\right] = O(1), \qquad (101)$$

$$\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^3\right] = O\left(\sqrt{\frac{n}{(\ln n)^2}}\right), \qquad (102)$$

$$\mathbb{E}\left[\left|\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^4\right] = O\left(\frac{n}{(\ln n)^2}\right). \qquad (103)$$

Observe that the largest term is (103), now, using that same statement (103) in Equation (62), we deduce that

$$\begin{aligned} \frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 &\leq C_4 \mathbb{E}\left[\left|\psi_k(\mathbf{X}_1;\eta_{k,\ell})\right|^4\right] \\ &= O\left(\frac{n}{(\ln n)^2}\right). \end{aligned}$$

It follows that

$$\Psi_{k,\ell}^{(1)} = O\left(\frac{n}{\ln n}\right)^{1/2}. \tag{104}$$

Let us now investigate the upper bound of $\Psi_{k,\ell}^{(2)}$ of (61). Observe that

$$\begin{aligned} \Psi_{k,\ell}^{(2)} &= \left\|\sum_{i=1}^{n}\mathbb{E}(\Psi_{i,k,\ell}^2/\mathcal{F}_{i-1})\right\|_2^{1/2} \\ &= \left(\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{E}\left[\Psi_{i,k,\ell}^2|\mathcal{F}_{i-1}\right]\right)^2\right]\right)^{1/4}, \end{aligned}$$

for all $i = 1, \ldots, n$, using the Jensen inequality and the fact that

$$(a-b)^2 = a^2 - 2ab + b^2,$$

it follows

$$\begin{aligned} &\sum_{i=1}^{n}\mathbb{E}\left[|\Psi_{i,k,\ell}|^2|\mathcal{F}_{i-1}\right] \\ &= \sum_{i=1}^{n}\left(\mathbb{E}\left[\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{x}_i;\eta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{x}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}\right]\right|^2|\mathcal{F}_{i-1}\right]\right) \\ &\leq 4\sum_{i=1}^{n}\mathbb{E}\left[\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\right|^2|\mathcal{F}_{i-1}\right]. \end{aligned}$$

From the independence between ϵ_i and $\mathbf{X}_i$, we have

$$\begin{aligned} \mathbb{E}\left[\epsilon^2|\mathcal{G}_{i-1}\right] &= \mathbb{E}\left[\epsilon^2|\mathbf{X}_i\right] \\ &= \mathbb{E}\left[\epsilon^2\right] = 1. \end{aligned}$$

Under assumptions **(M.1)–(M.2)** and **(E.1)**(i) and **(C.1)**, (63) and (105), we have

$$\begin{aligned} &\sum_{i=1}^{n}\mathbb{E}\left[\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\right|^2|\mathcal{F}_{i-1}\right] \\ &= \sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\right|^2|\mathcal{G}_{i-1}\right]|\mathcal{F}_{i-1}\right] \\ &\leq \frac{n(C_m^2+1)}{c_f}\int_{\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})|^2\left|\frac{\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})}{f(\mathbf{x})}\right|\nu(d\mathbf{x}) \\ &= \frac{n(C_m^2+1)}{c_f}(1+o(1))\int_{\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})|^2\nu(d\mathbf{x}) \\ &\leq nC, \end{aligned} \tag{105}$$

where

$$C=\frac{(C_m^2+1)}{c_f}(1+o(1)).$$

It follows

$$\Psi^{(2)}_{k,\ell} \;=\; Cn^{1/2}. \tag{106}$$

Combining (61), (69) and (71), we obtain

$$\mathbb{E}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right|^4\right) \;=\; O\left(\left(\frac{n}{\ln n}\right)^2\right)+O\left(n^2\right),$$

combining this with (95), we conclude

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right) \;=\; O\left(\left(\frac{1}{n^4}\right)\left(\frac{n}{\ln n}\right)^2\right)+O\left(\left(\frac{1}{n^4}\right)n^2\right). \tag{107}$$

Hence, there exists a constant $C>0$, such that

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right) \;\leq\; C\left(\frac{1}{n}\right)^2\leq C\left(\frac{\ln n}{n}\right)^2. \tag{108}$$

The proof is achieved. □

Proof of Lemma 8

Considering the previous decomposition (92) in Lemma 7, we have

$$\begin{aligned}\widehat{\theta}_{k,\ell}-\theta_{k,\ell} &= \left(\widehat{\theta}_{k,\ell}-\breve{\theta}_{k,\ell}\right)+\left(\breve{\theta}_{k,\ell}-\theta_{k,\ell}\right)\\ &= B_{k,\ell,1}+B_{k,\ell,2},\end{aligned}$$

where

$$\begin{aligned}B_{k,\ell,1} &= \frac{1}{n}\sum_{i=1}^{n}\Psi_{i,k,\ell}\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})-\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}\right]\right),\\ B_{k,\ell,2} &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})|\mathcal{F}_{i-1}\right]-\theta_{k,\ell}.\end{aligned}$$

Statement (94) achieves the desired result for the term $B_{k,\ell,2}$

$$\widehat{\theta}_{k,\ell}-\theta_{k,\ell} \;=\; B_{k,\ell,1}+o(1).$$

We consider the next decomposition

$$\Psi_{i,k,\ell} \;=\; V_{i,k,\ell}+W_{i,k,\ell}, \tag{109}$$

where

$$\begin{aligned}V_{i,k,\ell} &= \left(\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{x}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i}-\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i}|\mathcal{F}_{i-1}\right]\right),\\ W_{i,k,\ell} &= \left(\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}^c_{\mathcal{A}_i}-\mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}^c_{\mathcal{A}_i}|\mathcal{F}_{i-1}\right]\right),\end{aligned}$$

and

$$\mathbb{1}_{\mathcal{A}_i} = \left\{|\epsilon_i| \geq c_* \sqrt{\ln n}\right\},$$

and c_* denotes a constant which will be chosen later. Now, observe that

$$\begin{aligned}
\mathbb{P}\left(\left|\widehat{\theta}_{k,\ell} - \breve{\theta}_{k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) &\leq \mathbb{P}\left(|B_{k,\ell,1}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) + o(1) \\
&= \mathbb{P}\left(\left|\sum_{i=1}^{n} \Psi_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{n \ln n}\right) + o(1) \\
&= I_1 + I_2 + o(1), \qquad (110)
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \mathbb{P}\left(\left|\sum_{i=1}^{n} V_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{n \ln n}\right), \\
I_2 &= \mathbb{P}\left(\left|\sum_{i=1}^{n} W_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{n \ln n}\right).
\end{aligned}$$

First, we aim to bound the term I_1 of Equation (110). The Markov inequality and the Cauchy–Schwarz inequality yield

$$\begin{aligned}
I_1 &\leq \frac{2}{\kappa\sqrt{n \ln n}} \mathbb{E}\left(\left|\sum_{i=1}^{n} V_{i,k,\ell}\right|\right) \\
&\leq \frac{2}{\kappa\sqrt{n \ln n}} \sum_{i=1}^{n} \mathbb{E}\left(|V_{i,k,\ell}|\right). \qquad (111)
\end{aligned}$$

Observe that

$$\begin{aligned}
&\mathbb{E}\left(|V_{i,k,\ell}|\right) \\
&= \mathbb{E}\left(\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i} - \mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i}|\mathcal{F}_{i-1}\right]\right|\right) \\
&\leq \mathbb{E}\left(\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i}\right|\right) + \mathbb{E}\left(\mathbb{E}\left[\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i}\right||\mathcal{F}_{i-1}\right]\right) \qquad (112) \\
&\leq 2\mathbb{E}\left(\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_i}\right|\right) \\
&\leq 2\left(\mathbb{E}\left(\left|\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\right|^2\right)\right)^{1/2} (\mathbb{P}(\mathcal{A}_i))^{1/2}. \qquad (113)
\end{aligned}$$

We use (98) combined with an elementary Gaussian inequality and take c_* to have

$$\frac{c_*^2}{4} - 1/2 = 2.$$

We obtain

$$\begin{aligned}
I_1 &\leq \frac{2C}{\kappa}\sqrt{\frac{n}{\ln n}} \exp\left\{-\frac{c_*^2 \ln n}{4}\right\} \\
&\leq \frac{2C}{\kappa}\frac{n^{-\left(\frac{c_*^2}{4}-1/2\right)}}{\sqrt{\ln n}} \\
&\leq C_\kappa \frac{1}{n^2}, \qquad (114)
\end{aligned}$$

where $C_\kappa = \frac{2C}{\kappa}$. We now intend to investigate an upper bound for I_2 of decomposition (110). We start by verifying the condition of Lemma 2. Suppose that assumptions **(M.1)** and **(M.2)** are satisfied combined with (99), we obtain

$$\begin{aligned} |Y_i \mathbb{1}_{\mathcal{A}_i^c}| &\leq C_m + c_* \sqrt{\ln n} \\ &\leq C\sqrt{\ln n}, \end{aligned} \tag{115}$$

which implies

$$\begin{aligned} |W_{i,k,\ell}| &\leq \left| \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \psi_k(\mathbf{X}_i; \eta_{k,\ell}) \mathbb{1}_{\mathcal{A}_i^c} - \mathbb{E}\left[\frac{\rho(Y_i)}{f(\mathbf{X}_i)} \psi_k(\mathbf{X}_i; \eta_{k,\ell}) \mathbb{1}_{\mathcal{A}_i^c} \mathbb{1}_{\mathcal{A}_i^c} | \mathcal{F}_{i-1} \right] \right| \\ &\leq \frac{2C\sqrt{\ln n}}{c_f} \sup_{\mathbf{x} \in \mathbf{S}} |\psi_k(\mathbf{x}; \eta_{k,\ell})| \\ &\leq \frac{2C}{c_f} \sqrt{\ln n} \sqrt{\frac{n}{(\ln n)^2}} \\ &\leq C_3 \sqrt{\frac{n}{\ln n}} \\ &\leq C_3 \sqrt{n}, \end{aligned} \tag{116}$$

where $C_3 = \frac{2C}{C_f}$, let $B = C_3\sqrt{n}$, then, for all $\epsilon_n = \frac{\kappa}{2}\sqrt{n \ln n}$ sufficiently large n, we have

$$\begin{aligned} I_2 = \mathbb{P}\left(\left| \sum_{i=1}^{n} W_{i,k,\ell} \right| \geq \frac{\kappa}{2} \sqrt{n \ln n} \right) &\leq 2 \exp\left\{ -\frac{\epsilon_n^2}{2nB^2} \right\} \\ &= 2 \exp\left\{ -\frac{\left(\frac{\kappa}{2} \sqrt{n \ln n} \right)^2}{2n (C_3 \sqrt{n})^2} \right\} \\ &= 2 \exp\left\{ -\frac{\kappa^2 \ln n}{4 C_3^2 n} \right\} \\ &= 2 \exp\left\{ \ln n^{-\frac{\kappa^2}{4 C_3^2 n}} \right\} \\ &= 2 n^{-w(\kappa, n)}, \end{aligned} \tag{117}$$

where

$$w(\kappa, n) = \frac{\kappa^2}{4C_3^2 n}.$$

Taking κ such that $w(n, \kappa) = 2$, we have

$$I_2 \leq C \frac{1}{n^2}. \tag{118}$$

It follows from (110), (114) and (118) that

$$\begin{aligned} \mathbb{P}\left(\left| \widehat{\theta}_{k,\ell} - \widetilde{\theta}_{k,\ell} \right| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) &\leq C \frac{1}{n^2} + o(1) && (119) \\ &\leq C \left(\frac{\ln n}{n} \right)^2. && (120) \end{aligned}$$

The proof of (79) is achieved. □

Author Contributions: S.D., A.A.H. and S.B.: conceptualization, methodology, investigation, writing-original draft, writing—review and editing. All authors contributed equally to the writing of this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the Special Issue Editor of the Special Issue on "Functional Data Analysis: Theory and Applications to Different Scenarios", Mustapha Rachdi for the invitation. The authors are indebted to the Editor-in-Chief and the two referees for their very generous comments and suggestions on the first version of our article, which helped us to improve content, presentation and layout of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bosq, D. *Linear Processes in Function Spaces*; Volume 149 of Lecture Notes in Statistics; Theory and applications; Springer: New York, NY, USA, 2000.
2. Ramsay, J.; Silverman, B. *Functional Data Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2005.
3. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Springer Series in Statistics; Springer: New York, NY, USA, 2006.
4. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer Series in Statistics; Springer: New York, NY, USA, 2012.
5. Zhang, J. *Analysis of Variance for Functional Data*; Volume 127 of Monographs on Statistics and Applied Probability; CRC Press: Boca Raton, FL, USA, 2014.
6. Qing, S.; Choi, T. *Gaussian Process Regression Analysis for Functional Data*; CRC Press: Boca Raton, FL, USA, 2011.
7. Geenens, G. Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.* **2011**, *5*, 30–43. [CrossRef]
8. Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* **2014**, *147*, 1–23. [CrossRef]
9. Shang, H. A survey of functional principal component analysis. *AStA Adv. Stat. Anal.* **2014**, *98*, 121–142. [CrossRef]
10. Horváth, L.; Rice, G. An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.* **2015**, *28*, 505–548. [CrossRef]
11. Müller, H. Peter hall, functional data analysis and random objects. *Ann. Statist.* **2016**, *44*, 1867–1887. [CrossRef]
12. Nagy, S. An overview of consistency results for depth functionals. In *Functional Statistics and Related Fields*; Contrib. Stat.; Springer: Cham, Switzerland, 2017; pp. 189–196.
13. Vieu, P. On dimension reduction models for functional data. *Statist. Probab. Lett.* **2018**, *136*, 134–138. [CrossRef]
14. Aneiros-Pérez, G.; Cao, R.; Fraiman, R.; Genest, C.; Vieu, P. Recent advances in functional data analysis and high-dimensional statistics. *J. Multivar. Anal.* **2018**, *170*, 3–9. [CrossRef]
15. Ling, N.; Vieu, P. Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics* **2018**, *52*, 934–949. [CrossRef]
16. Goia, A.; Vieu, P. An introduction to recent advances in high/infinite dimensional statistics. *J. Multivar. Anal.* **2016**, *146*, 1–6. [CrossRef]
17. Almanjahie, I.M.; Bouzebda, S.; ChikrElmezouar, Z.; Laksaci, A. The functional *k*NN estimator of the conditional expectile: uniform consistency in number of neighbors. *Stat. Risk Model.* **2022**, *38*, 47–63. [CrossRef]
18. Almanjahie, I.M.; Bouzebda, S.; Kaid, Z.; Laksaci, A. Nonparametric estimation of expectile regression in functional dependent data. *J. Nonparametr. Stat.* **2022**, *34*, 250–281. [CrossRef]
19. Bouzebda, S.; Nezzal, A. 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.* **2022**, 1–103. [CrossRef]
20. Bouzebda, S.; Nemouchi, B. Weak-convergence of empirical conditional processes and conditional *U*-processes involving functional mixing data. *Stat. Inference Stoch. Process.* **2022**, 1–56. [CrossRef]
21. Bouzebda, S.; Nemouchi, B. Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional *U*-statistics involving functional data. *J. Nonparametr. Stat.* **2020**, *32*, 452–509. [CrossRef]
22. Almanjahie, I.M.; Kaid, Z.; Laksaci, A.; Rachdi, M. Estimating the conditional density in scalar-on-function regression structure: *k*-NN local linear approach. *Mathematics* **2022**, *10*, 902. [CrossRef]
23. Meyer, Y. Wavelets and operators. In *Different Perspectives on Wavelets (San Antonio, TX, 1993)*; Volume 47 of Proc. Sympos. Appl. Math.; Amer. Math. Soc.: Providence, RI, USA, 1993; pp. 35–58.
24. Daubechies, I. *Ten Lectures on Wavelets*; Volume 61 of CBMS-NSF Regional Conference Series in Applied Mathematics; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 1992.
25. Mallat, S. *A Wavelet Tour of Signal Processing*, 3rd ed.; Elsevier/Academic Press: Amsterdam, The Netherlands, 2009;
26. Vidakovic, B. *Statistical Modeling by Wavelets*; Wiley Series in Probability and Statistics: Applied Probability and Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1999.

27. Härdle, W.; Kerkyacharian, G.; Picard, D.; Tsybakov, A. *Wavelets, Approximation, and Statistical Applications*; Volume 129 of Lecture Notes in Statistics; Springer: New York, NY, USA, 1998.
28. Rao, B.L.S.P. Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cybernet.* **1996**, *28*, 91–100. [CrossRef]
29. Chaubey, Y.P.; Doosti, H.; Rao, B.P. Wavelet based estimation of the derivatives of a density for a negatively associated process. *J. Stat. Theory Pract.* **2008**, *2*, 453–463. [CrossRef]
30. Rao, B.L.S.P. *Nonparametric Estimation of Partial Derivatives of a Multivariate Probability Density by the Method of Wavelets*; De Gruyter: Berlin, Germany, 2018; pp. 321–330.
31. Prakasa Rao, B. L.S. Wavelet estimation for derivative of a density in the presence of additive noise. *Braz. J. Probab. Stat.* **2018**, *32*, 834–850. [CrossRef]
32. Allaoui, S.; Bouzebda, S.; Chesneau, C.; Liu, J. Uniform almost sure convergence and asymptotic distribution of the wavelet-based estimators of partial derivatives of multivariate density function under weak dependence. *J. Nonparametr. Stat.* **2021**, *33*, 170–196. [CrossRef]
33. Chesneau, C.; Kachour, M.; Maillot, B. Nonparametric estimation for functional data by wavelet thresholding. *REVSTAT* **2013**, *11*, 211–230.
34. Rosenblatt, M. Uniform ergodicity and strong mixing. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **1972**, *24*, 79–84. [CrossRef]
35. Bradley, R. *Introduction to Strong Mixing Conditions*; Kendrick Press: Heber City, UT, USA, 2007; Volume 1.
36. Laïb, N.; Louani, D. Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *J. Multivariate Anal.* **2010**, *101*, 2266–2281. [CrossRef]
37. Bouzebda, S.; Didi, S.; ElHajj, L. Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: Asymptotic results. *Math. Methods Statist.* **2015**, *24*, 163–199. [CrossRef]
38. Delecroix, M.; Rosa, A.C. Nonparametric estimation of a regression function and its derivatives under an ergodic hypothesis. *J. Nonparametr. Statist.* **1996**, *6*, 367–382. [CrossRef]
39. Bouzebda, S.; Didi, S. Additive regression model for stationary and ergodic continuous time processes. *Comm. Statist. Theory Methods* **2017**, *46*, 2454–2493. [CrossRef]
40. Bouzebda, S.; Didi, S. Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: Asymptotic results. *Comm. Statist. Theory Methods* **2017**, *46*, 1367–1406. [CrossRef]
41. Bouzebda, S.; Didi, S. Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes. *Rev. Mat. Complut.* **2021**, *34*, 811–852. [CrossRef]
42. Bouzebda, S.; Didi, S. Some results about kernel estimators for function derivatives based on stationary and ergodic continuous time processes with applications. *Comm. Statist. Theory Methods* **2022**, *51*, 3886–3933. [CrossRef]
43. Bouzebda, S.; Chaouch, M. Uniform limit theorems for a class of conditional z-estimators when covariates are functions. *J. Multivar. Anal.* **2022**, *189*, 104872. [CrossRef]
44. Bouzebda, S.; Chaouch, M.; Didi, S. Asymptotics for function derivatives estimators based on stationary and ergodic discrete time processes. *Ann. Inst. Statist. Math.* **2022**, *74*, 1–35. [CrossRef]
45. Leucht, A.; Neumann, M. Degenerate U- and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.* **2013**, *65*, 349–386. [CrossRef]
46. Neumann, M.H. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **2011**, *17*, 1268–1284. [CrossRef]
47. Goh, S.S. Wavelet bases for Hilbert spaces of functions. *Complex Var. Elliptic Equ.* **2007**, *52*, 245–260. [CrossRef]
48. Kerkyacharian, G.; Picard, D. Density estimation in Besov spaces. *Statist. Probab. Lett.* **1992**, *13*, 15–24. [CrossRef]
49. Kerkyacharian, G.; Picard, D. Thresholding algorithms, maxisets and well-concentrated bases. *Test* **2000**, *9*, 283–344; [CrossRef]
50. Prakasa, Rao, B.L.S. Nonparametric density estimation for functional data via wavelets. *Commun. Stat. Theory Methods* **2010**, *39*, 1608–1618. [CrossRef]
51. Cohen, A.; DeVore, R.; Kerkyacharian, G.; Picard, D. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.* **2001**, *11*, 167–191. [CrossRef]
52. DeVore, R.A. Nonlinear approximation. In *Acta Numerica, 1998*; Volume 7 of Acta Numer; Cambridge Univ. Press: Cambridge, UK, 1998; pp. 51–150.
53. Autin, F. Point de vue Maxiset en Estimation Non Paramétrique. Ph.D. Theses, Université Paris-Diderot-Paris VII, Paris, France, 2004.
54. Gasser, T.; Hall, P.; Presnell, B. Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1998**, *60*, 681–691. [CrossRef]
55. Donoho, D.L.; Johnstone, I.M.; Kerkyacharian, G.; Picard, D. Density estimation by wavelet thresholding. *Ann. Statist.* **1996**, *24*, 508–539. [CrossRef]
56. Peškir, G. The uniform mean-square ergodic theorem for wide sense stationary processes. *Stoch. Anal. Appl.* **1998**, *16*, 697–720. [CrossRef]
57. Ferraty, F.; Vieu, P. Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Nonparametric Stat.* **2004**, *16*, 111–125. [CrossRef]
58. Ferraty, F.; Vieu, P. The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **2002**, *17*, 545–564. [CrossRef]

59. Ouassou, I.; Rachdi, M. Regression operator estimation by delta-sequences method for functional data and its applications. *AStA Adv. Stat. Anal.* **2012**, *96*, 451–465. [CrossRef]
60. Burkholder, D.L. Distribution function inequalities for martingales. *Ann. Probab.* **1973**, *1*, 19–42. [CrossRef]
61. de la Peña, V.H.; Giné, E. *Decoupling*; Probability and Its Applications (New York); From dependence to independence, Randomly stopped processes. *U*-statistics and processes. Martingales and beyond; Springer: New York, NY, USA, 1999;