Wavelet Density and Regression Estimators for Continuous Time Functional Stationary and Ergodic Processes

Sultana Didi, Salim Bouzebda

mathematics

MDPI

Article

Wavelet Density and Regression Estimators for Continuous Time Functional Stationary and Ergodic Processes

Sultana Didi [1,†] **and Salim Bouzebda** [2,*,†]

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

Abstract: In this study, we look at the wavelet basis for the nonparametric estimation of density and regression functions for continuous functional stationary processes in Hilbert space. The mean integrated squared error for a small subset is established. We employ a martingale approach to obtain the asymptotic properties of these wavelet estimators. These findings are established under rather broad assumptions. All we assume about the data is that they are ergodic, but beyond that, we make no assumptions. In this paper, the mean integrated squared error findings in the independence or mixing setting were generalized to the ergodic setting. The theoretical results presented in this study are (or will be) valuable resources for various cutting-edge functional data analysis applications. Applications include conditional distribution, conditional quantile, entropy, and curve discrimination.

Keywords: multivariate regression estimation; multivariate density estimation; stationarity; ergodicity; rates of strong convergence; wavelet-based estimators; martingale differences; continuous time series

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

Citation: Didi, S.; Bouzebda, S. Wavelet Density and Regression Estimators for Continuous Time Functional Stationary and Ergodic Processes. *Mathematics* **2022**, *10*, 4356. https://doi.org/10.3390/math10224356

Academic Editors: Mustapha Rachdi and Leonid V. Bogachev

Received: 10 October 2022
Accepted: 17 November 2022
Published: 19 November 2022

1. Introduction and Motivations

In recent years, the statistical literature has become increasingly interested in statistical issues pertaining to the study of 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, and satellite images, is driving the expansion of this research topic; statistical modeling of these data as random functions uncovered numerous complex theoretical and numerical research challenges. The reader may consult the monographs to summarize functional data analysis's theoretical and practical aspects [1] for linear models for random variables with values in a Hilbert space, [2] for scalar-on-function and function-on-function linear models, functional principal component analysis, and parametric discriminant analysis [3], on the other hand, concentrated on nonparametric methods, particularly kernel-type estimation for scalar-on-function nonlinear regression models. Such tools were extended to classification and discrimination analysis [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 [5] was interested in analyzing variance for functional data, whereas [6] was more concerned with regression analysis for gaussian processes. Recent studies and surveys on functional data modeling and analysis can be found in [7–22].

The subject of estimating conditional models has been extensively studied in the statistical literature, utilizing several estimation techniques, the most prevalent of which is the conventional kernel method. This is because these methods have broad applications and play an important role in statistical inference. However, such methods may have certain

drawbacks when predicting compactly supported or discontinuous curves at boundary locations. The prevalence of alternative wavelet approaches can be attributed to the adaptability of these methods to discontinuities in the curve that is to be approximated. In application, the wavelet method offers a straightforward estimating algorithm to both implement and compute. Refs. [23–26] provided more information on wavelet theory. We quote the work of [27], which discusses the properties of wavelet approximation and analyzes the application of wavelets in various curve estimation issues in depth. In [28], numerous applications of wavelet theory are described for the independent unidimensional case, with an emphasis on calculating the integrated squared derivative density function. The findings of [28] were later extended by [29] to estimate the density derivatives for negatively and positively related sequences, respectively. In [30], wavelet estimators for partial derivatives of a multivariate probability density function were constructed, and convergence rates for an independence scenario could be determined. In detail, [31] discusses the estimation of partial derivatives of a multivariate probability density function in the presence of additive noise. We could consult [32,33] for the most recent information on this subject. Using the independent and identically distributed paradigm, [34] examined density and regression estimation issues unique to functional data. Ref. [34] proposed a novel adaptive method based on the term-by-term selection of wavelet coefficient estimators and wavelet bases for Hilbert spaces of functions.

The primary purpose of this paper is to provide further context for the earlier discussion of stationary ergodic processes. The lack of research on the general dependence framework for wavelet analysis prompted us to conduct the present study. Several arguments for contemplating an ergodic dependency structure in that data as opposed to a mixing structure are presented in [35–42], where further information on the notion of the ergodic property of processes and examples of such processes are provided. In [43], one of the arguments used to justify the ergodic setting is that establishing ergodic characteristics rather than the mixing condition can be significantly simpler for some classes of processes. Therefore, the ergodicity hypothesis provides the ideal framework for examining data series created by chaotic noise.

In the body of statistical research that has been published, estimation issues based on discretely sampled observations as well as continuous time have been investigated. In finance, even if the underlying process is continuous across time, only its values at a finite number of points are known. The first one is more important from the standpoint of financial econometrics (see [44,45]). On the other hand, the technical development of statistical inference using a continuous time record of observations is far easier to accomplish than an inference based on discretely observed data. It makes it possible to advance further in the model's statistical analysis and find solutions to various concerns concerning discretely observed diffusions that had not been resolved until this point. Further, remember that the discrete-time model hits its limits when the discretization step equals zero, which is when the continuous-time model becomes the limit (see [46]). As a result, the asymptotic behavior of estimate processes in actuality may be close to the asymptotic behavior established theoretically for continuous-time observations if the available data are "dense enough" relative to the observation period. The knowledge of the most accurate estimate derived from continuous-time data is also of practical significance. Ref. [47] investigated the challenge of estimating an invariant density function for a continuous-time Markov process and established the mean-square consistency of kernel-type estimators. Ref. [48] demonstrated the homogeneity and consistency of these estimators. Ref. [48] showed that a family of smoothed estimators, including kernel-type estimators, are asymptotically normal with a rate of convergence $T^{1/2}$ for continuous-time processes. This finding is unexpected because the invariant density estimators operate differently in this instance than in discrete-time processes, where the convergence rate is often smaller than $1/2$. Numerous physical phenomena, such as seismic waves, the Earth's magnetic field, and isotherms and isobars in meteorology, are functional variables observable in continuous time, as is evident.

In [49], we have considered the wavelet basis for the nonparametric estimation of density and regression functions for continuous functional stationary processes in Hilbert space. We have characterized the mean integrated square errors. This research aims to provide the first complete theoretical rationale for wavelet-based functional density and regression function estimation for continuous stationary processes by extending our earlier work [49] to the continuous setting. The employment of wavelet estimators in functional ergodic data frames and the resulting difficulty of establishing the mean integrated square error over suitable decomposition spaces is, to the best of our knowledge, still an unresolved topic in the literature. By merging different martingale theory techniques utilized in the mathematical construction of the proofs, we aim to contribute to the literature by addressing this gap. In the independent or mixing setting, the tools employed for regression estimation change significantly in the ergodic setting. Nevertheless, we shall see that more than merely mixing preexisting ideas and outcomes is needed to address the issue. In order to deal with wavelet estimators in an ergodic setting, one must resort to complex mathematical derivations.

The paper's structure is as follows: The multiresolution analysis is described in Section 2. The principal density estimate results are presented in Section 3. Section 4 summarizes the principal outcomes for regression estimation. Section 5 provides some examples of potential applications. There are some concluding observations in Section 6. All proof is compiled in Section 7.

2. Multiresolution Analysis

Following [34,49,50], we will now introduce some basic notations for defining wavelet bases for Hilbert spaces of functions with a few modifications to accommodate our context. In this study, nonlinear, thresholded, wavelet-based estimators are examined. Beginning with a description of the fundamental theory of wavelet approaches, we then introduce nonlinear wavelet-based estimators. The interested reader should consult [23,24], see also [51,52] and the references therein, despite the fact that the wavelet is based on a separable Hilbert space $\mathbf{H}$ of real or complex-valued functions on a complete separable metric space. Let $\mathbf{H}$ represent the separable Hilbert space of real-valued functions defined on a separable complete metric space $\mathbf{S}$. Given that $\mathbf{H}$ is separable; it possesses an orthonormal basis

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

where Δ is a countable index set. The space $\mathbf{H}$ 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

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

fulfill one of the two following conditions, for instance, see [34,50] 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.

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 zero vectors. As for (**A.2**), it gives

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

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

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

where

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

The following collection serves as the orthonormal basis for $\mathbf{H}$ (see Theorem 2 of [50]):

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

For further details, see [34,50,53]. Hence, we infer 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

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

Adding two additional assumptions to the orthonormal basis $\mathcal{E}$:

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

(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 $k \geq 0$, we have

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

Remark 1. *Clearly, the assumption (**E.1**) is satisfied under assumption (**A.1**) whenever $C_1 = 1$; we may also consult [50], Example 2 and its applications for further details. [50,53] have provided three examples satisfying condition (**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 [53] Theorem 3.2. Moreover, [34] has used both assumptions in the independent and identically distributed functional data.

Besov Space

Over the years, numerous statisticians have pondered the following question: given an estimating method and a required estimation rate for a certain loss function, what is the largest space over which this rate can be accomplished, for example, see [27,54] and the references therein. We are interested in estimating methods based on thresholding methods given by wavelet bases in a natural situation. It is common knowledge that wavelet bases serve to characterize the smoothness of 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 on both the smoothness properties that are dependent on both the smoothness of ψ and its dual function $\tilde{\psi}$, for instance, see [54] for more detail and examples, at this point, we may consult [55]. The following statistical definition is used in approximation theory for the study of nonlinear procedures, such as thresholding and greedy algorithms; see [27,52,54,56].

Definition 1 (Besov space)**.** *Let $s > 0$. We say that the function $f \in \mathbf{H}$, defined by statement (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 statement (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. The Density Estimation

Let $\{\mathbf{X}_t, Y_t\}_{t\geq 0}$ denote a sequence of strictly stationary ergodic pairs of random elements, where $\mathbf{X}_t$ takes values in a complete separable metric space of Hilbert space $\mathbf{S}$ associated with the corresponding Borel σ-algebra $\mathcal{B}$ and Y_i is a real or complex-valued variable. Let $\mathbb{P}_{\mathbf{X}}$ denote the probability measure induced by $\mathbf{X}_0$ on $(\mathbf{S}, \mathcal{B})$. Assume that there exists σ-finite measure ν on the measurable space $(\mathbf{S}, \mathcal{B})$ in such a way that $\mathbb{P}_{\mathbf{X}}$ is dominated by ν. The Radon–Nikodym theorem guarantees the existence of a measurable function that is nonnegative $f(\cdot)$ in such a way that

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

In this framework, we intend to estimate $f(\cdot)$ on the basis of n observed functional data $(\mathbf{X}_t)_{\{0\leq t\leq T\}}$. We assume 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 research, we are especially interested in the wavelet estimation processes created in the 1990s, see Meyer's work for the functional data of a Hilbert space and, more specifically, the nonlinear estimators. The majority of this model's approaches involve introducing kernel estimator techniques to estimate the functional component of the model, see [57]. Let $f(\cdot)$ be the sample's common density function $(\mathbf{X}_t)_{\{0\leq t\leq T\}}$, which is assumed to be

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

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

Density Function Estimator

From now, suppose that the density function $f(\cdot) \in \mathbf{H}$ is a separable Hilbert space. Then $f(\cdot)$ satisfies the wavelet representation (7). Assume that we observe a sequence $\{(\mathbf{X}_t, Y_t)\}_{0\leq t\leq T}$ of copies of $(\mathbf{X}, Y)$ that is supposed to be functional stationary and ergodic with $\mathbf{X}$ with the density function $f(\cdot)$. We examine density estimation utilizing wavelet bases for Hilbert spaces of functions of [50]. We consider the estimated coefficients $\{\alpha_{k,\ell}\}$ and $\{\beta_{k,\ell}\}$ given, respectively, by (14) and (15), for any $j_0 \leq m$. Here the resolution level $m = m(T) \to \infty$ at the 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 [34] in the setting of the independent and identically distributed functional processes, in particular, we refer to [49]. The density wavelet hard thresholding estimator $\widehat{f}(\cdot)$ is defined for all $\mathbf{x} \in \mathbf{S}$, by

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

where

$$\widehat{\alpha}_{k,\ell} = \frac{1}{T}\int_0^T \phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt, \tag{14}$$

$$\widehat{\beta}_{k,\ell} = \frac{1}{T}\int_0^T \psi_k(\mathbf{X}_t;\eta_{k,\ell})dt. \tag{15}$$

Here κ denotes a large-enough constant, and m_T is the integer satisfying

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

See [36,38] for further details on the multivariate case.

Comments on the Method of Estimation

Our estimation method is divided into three steps:

1. Estimation of the wavelet coefficients $\alpha_{k,\ell}$ and $\beta_{k,\ell}$, see statement (8), $\widehat{\alpha}_{k,\ell}$ and $\widehat{\beta}_{k,\ell}$ given by Equations (14) and (15);
2. The greatest $\widehat{\beta}_{k,\ell}$ is selected by applying hard thresholding;
3. Reconstruction of the selected elements of the initial wavelet basis.

It is essential to highlight that our choice is the universal threshold $\kappa\left(\frac{\ln T}{T}\right)^{1/2}$ and the definition of m_T is based on theoretical considerations. The estimator under consideration is not dependent on the smoothness of $f(\cdot)$; see [53] for more details in the case of the linear wavelet estimator of $f(\cdot)$. Moreover, for additional information on the case of $\mathbf{H} = \mathbb{L}([a, b])$ and more standard nonparametric models, refer to [27,58]. Notation is necessary for stating the results. Throughout the paper, we shall denote by $B \in \mathcal{B}$, the open set of the Borel $\sigma-$algebra $\mathcal{B}$. For any $0 \leq t \leq T$ and $\delta > 0$ small real, we define

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

and

$$F_{\mathbf{X}_t}^{\mathcal{F}_{t-\delta}} = \mathbb{P}(\mathbf{X}_t \in B|\mathcal{F}_{t-\delta})$$

as the distribution function and the conditional distribution function, given the $\sigma-$field $\mathcal{F}_{t-\delta}$, respectively. Before presenting our results, we present supplementary notation and our hypotheses. For the remainder of the paper, for a positive real δ, we will denote by

$$n = \frac{T}{\delta} \in \mathcal{N}, \quad \text{and} \quad T_j = j\delta, \quad \text{for} \quad j = 1, \ldots, n.$$

Let $\mathcal{F}_{t-\delta}$ be the $\sigma-$field generated by

$$\{(X_s, Y_s) : 0 \leq s < t - \delta\}$$

Let $\mathcal{F}_j$ be the $\sigma-$field generated by

$$\{(X_s, Y_s) : 0 \leq s \leq T_j\}$$

and

$$\mathcal{S}_t = \sigma\{(X_s, Y_s); (X_r) : 0 \leq s \leq t, t \leq r \leq t + \delta\}.$$

The following assumptions are required for the entire paper.

(**C.0**) There is a nonnegative measurable function $f_t^{\mathcal{F}_{t-\delta}}$ such that

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

We may refer to [3,35,59] for further details.

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

$$\lim_{n\to\infty} \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x}) = f(\mathbf{x}), \quad \text{in the } a.s. \text{ and } L^2 \text{ sense.}$$

At this point, we may refer to [60] for further details.

Comments on hypotheses. Approximating the integral $\int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt$ by its Riemann's sum, we have

$$\frac{1}{T}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt = \frac{1}{T}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt \backsimeq \frac{1}{n}\sum_{i=1}^{n} f_{T_i}^{\mathcal{F}_{T_{i-1}}}(\mathbf{x}).$$

By the fact that process $(X_{T_j})_{j\geq 1}$ is stationary and ergodic, in a similar way as in [61] (see, Lemma 4 and Corollary 1 together with their proofs), one may establish that the sequence

$$\left(f_{T_i}^{\mathcal{F}_{T_{i-1}}}(\mathbf{x})\right)_{i\geq 1} = (f_{i\delta,(i-1)\delta}(\mathbf{x}))_{i\geq 1}$$

of random functions is stationary and ergodic. Indeed, it suffices to substitute the conditional densities in the work of Delecroix with $f_{i\delta,(i-1)\delta}$ and the density by function $f(\cdot)$. Refer to [35,62] and the references therein.

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

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

there is a constant $C_1 > 0$ *such that*

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

for a large-enough T.

The following upper bound finding is an immediate consequence: if $f \in \mathcal{B}_{\infty}^{s/(2s+1)}(\mathbf{H}) \cap \mathcal{W}^{2/(2s+1)}(\mathbf{H})$ for $s > 0$, then there is a constant $C_2 > 0$ in such a way that

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

This convergence rate is close to optimal in the "standard" minimax configuration (see, [27]). Moreover, on the application of [52] (Theorem 3.2), one obtains that $\mathcal{B}_{\infty}^{\theta/2}(\mathbf{H}) \cap \mathcal{W}^{2(1-\theta)}(\mathbf{H})$ is the "maxiset" associated with $\widehat{f}(\cdot)$ at the convergence rate of $(\ln T/T)^{\theta}$, i.e.,

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

4. The Regression Estimation

Let $\rho : \mathbb{R}^q \to \mathbb{R}$ be a measurable function. The regression function $\mathbf{m}(\cdot, \rho)$ is defined 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 assume 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 a square-integrable with respect to the σ-finite measure ν. We shall assume that there is a known constant and $C_{\mathbf{m}} > 0$ in such a way that

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

In this framework, we redefine the probability measure $\mathbb{P}_{\mathbf{X}}$ in (11) and assume that $f(\cdot)$ is a nonnegative measurable known function.

(**M.1**) We shall assume that there is a known constant $C_{\mathbf{m}} > 0$ in such a way that

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

(**M.2**) We shall suppose that there is a known constant $c_f > 0$ in such a way that

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

Regression Function Estimator

In this framework, our goal is to estimate $\mathbf{m}(\cdot, \rho)$ based on observed functional data $(\mathbf{X}_t, Y_t)_{\{0\leq t\leq T\}}$. The kernel estimator for the regression function of functional data has been suggested by [59]

$$\widehat{\mathbf{m}}_{n;h_T}(\mathbf{x}, \rho) := \frac{\displaystyle\int_0^T \rho(Y_t) K\left(\frac{d(\mathbf{x}, \mathbf{X}_t)}{h_T^d}\right) dt}{\displaystyle\int_0^T K\left(\frac{d(\mathbf{x}, \mathbf{X}_t)}{h_T^d}\right) dt}.$$

This estimator is similar to the one proposed by [59] in the discrete framework. By using the idea of [58], we propose the hard wavelet thresholding estimator $\widehat{\mathbf{m}}(\cdot, \rho)$, for all $\mathbf{x} \in \mathbf{S}$, by

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

where

$$\widehat{\eta}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt, \tag{22}$$

$$\widehat{\theta}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})dt. \tag{23}$$

where m_T is the integer satisfying

$$\frac{1}{2}\frac{T}{(\ln T)^2} \leq |\mathcal{J}_{m_T}| \leq \frac{T}{(\ln T)^2}, \tag{24}$$

and κ is a large-enough constant. The multivariate case for discrete- and continuous-time linear wavelet estimators was examined by [36,38].

Theorem 2. *Under the assumptions (**E.1**), (**E.2**) (**M.1**), (**M.2**), (**C.0**), and (**C.1**), combined with condition (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 is a constant $C_3 > 0$ such that*

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

for T large enough.

Assume that $\mathbf{m}(\cdot,\rho)$ and $f(\cdot)$ fulfill (**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 is a constant $C_3 > 0$ in such a way that

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

for T 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 is a constant $C_4 > 0$ in such a way that

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

This convergence rate is close to optimal in the "standard" minimax framework (see, [27]) up to an extra logarithmic term. According to our knowledge, Theorem 2 is the first one to examine an adaptive wavelet-based estimator for functional data in the context of nonparametric regression for an ergodic process. Ref. [49] studied the same estimator in a discrete time setting.

From the fact that the coefficients defined in Equations (22) and (23) depend on the unknown function $f(\cdot)$, it is possible to use

$$\widetilde{\eta}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\rho(Y_t)}{\widehat{f}(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt, \tag{26}$$

$$\widetilde{\theta}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\rho(Y_t)}{\widehat{f}(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})dt. \tag{27}$$

The wavelet hard thresholding estimator is defined by $\widetilde{\mathbf{m}}(\cdot,\rho)$, for all $\mathbf{x}\in\mathbf{S}$, by

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

Keep in mind the following elementary observation

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

We can infer, from the last equation, the following

$$\begin{aligned}\widetilde{\eta}_{k,\ell} &= \widehat{\eta}_{k,\ell}+\frac{1}{T}\int_0^T\frac{(f(\mathbf{X}_t)-\widehat{f}(\mathbf{X}_t))}{f(\mathbf{X}_t)\widehat{f}(\mathbf{X}_t)}\rho(Y_t)\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt,\\ \widetilde{\theta}_{k,\ell} &= \widehat{\theta}_{k,\ell}+\frac{1}{T}\int_0^T\frac{(f(\mathbf{X}_t)-\widehat{f}(\mathbf{X}_t))}{f(\mathbf{X}_t)\widehat{f}(\mathbf{X}_t)}\rho(Y_t)\psi_k(\mathbf{X}_t;\eta_{k,\ell})dt.\end{aligned}$$

Notice that

$$\begin{aligned}(\widetilde{\eta}_{k,\ell}-\widehat{\eta}_{k,\ell})^2 &= \left\{\frac{1}{T}\int_0^T\frac{(f(\mathbf{X}_t)-\widehat{f}(\mathbf{X}_t))}{f(\mathbf{X}_t)\widehat{f}(\mathbf{X}_t)}\rho(Y_t)\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt\right\}^2\\ &\leq \frac{1}{T}\int_0^T\left\{\frac{(f(\mathbf{X}_t)-\widehat{f}(\mathbf{X}_t))}{f(\mathbf{X}_t)\widehat{f}(\mathbf{X}_t)}\right\}^2dt\\ &\quad\times\frac{1}{T}\int_0^T\{\rho(Y_t)\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\}^2dt.\end{aligned}$$

This gives

$$\begin{aligned}\mathbb{E}(\widetilde{\eta}_{k,\ell}-\widehat{\eta}_{k,\ell})^2 &\leq \mathbb{E}\left\{\frac{1}{T}\int_0^T\left\{\frac{(f(\mathbf{X}_t)-\widehat{f}(\mathbf{X}_t))}{f(\mathbf{X}_t)\widehat{f}(\mathbf{X}_t)}\right\}^2dt\right.\\ &\quad\left.\times\frac{1}{T}\int_0^T\mathbb{E}(\rho(Y_t)^2\mid\mathbf{X}_t)\phi_k(\mathbf{X}_t;\zeta_{k,\ell})^2dt\right\}.\end{aligned}$$

Under the conditions of Theorems 1 and 2, one can find a positive $\mathfrak{C}$ constant such that

$$\mathbb{E}(\widetilde{\eta}_{k,\ell}-\widehat{\eta}_{k,\ell})^2 \leq \mathfrak{C}\left(\frac{\ln T}{T}\right)^{\theta}.$$

A combination of Theorem 1 with Theorem 2 gives the following corollary.

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

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

for T large enough.

Remark 2. *Let $\{\mathbf{X}_n, n\in\mathbb{Z}\}$ 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 shift transformation or time-evolution. The labeling of strong mixing in the preceding definition is more rigorous than what is commonly referred to as strong mixing (when using the terminology of measure-preserving dynamical systems), meaning that it satisfies the following conditions

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

for any two measurable sets A, B, see, for instance, [63]. Therefore, substantial mixing requires ergodicity, although the opposite is not necessarily true (see, for instance, Remark 2.6 on page 50 about Proposition 2.8 on page 51 in [64]).

Remark 3. *Ref. [43] 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*

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

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 [65]). It is known that any sequence $(\varepsilon_i)_{i\in\mathbb{Z}}$ of independent and identically distributed random variables is ergodic. Hence, according to Proposition 2.10 in [64], 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 $\vartheta(\cdot)$. Ref. [66] has constructed an example of a non-mixing ergodic continuous-time process. It is well known that the fractional Brownian motion $\{W_{\mathrm{t}}^H:t\geq0\}$ with parameter $H\in(0,1)$ has strictly stationary increments. Otherwise, the fractional Gaussian noise, defined for every $s>0$ by

$$\{G_{\mathrm{t}}^H:t\geq0\}:=\{W_{\mathrm{t}+s}^H-W_{\mathrm{t}}^H:t\geq0\},$$

is a strictly stationary-centered, long memory process when $H\in(\frac{1}{2},1)$ (for instance, see [67] p. 55, [68] p. 17), hence the condition of strong mixing is not satisfied. Let $\{G_{\mathrm{t}}:t\geq0\}$ be a strictly stationary-centered Gaussian process with correlation function

$$R(t)=\mathbb{E}[G_0G_{\mathrm{t}}].$$

Relying on the work of [69], Lemma 4.2, it follows that the process $\{G_{\mathrm{t}}:t\geq0\}$ is ergodic whenever

$$\lim_{\mathrm{t}\to\infty}R(t)=0,$$

which is the case for the process $\{G_{\mathrm{t}}^H:t\geq0\}$.

Remark 4. *In continuous time, sampling is frequently used to obtain data. Several discretization strategies, including deterministic and random sampling, have been proposed in the literature. The interested reader is referred to [70–74]. To simplify the concept, we consider the density estimator of $f(\cdot)$ based on $\{\mathbf{X}_t:t\in[0,T]\}$ and its sampled discrete sequence $\{\mathbf{X}(t_k):k=1,\ldots,n\}$. The estimator of the sampled density $f(\cdot)$ is*

$$f_n(\mathbf{x})=\frac{1}{nh_n^d}\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{X}_{t_j}}{h_n}\right).$$

As stated in [70], we can only recollect two designs: irregular sampling and random sampling.

- **Deterministic sampling.** *Consider the situation where* $(t_k)_{1\leq k\leq n}$ *is deterministically irregularly spaced with*

$$\inf_{1\leq k\leq n} |t_{j+1} - t_j| = \frac{1}{\tau},$$

for some $\tau > 0$. $\mathcal{G}_k := \sigma(X(t_k))$ *the* σ*-field generated by* $\{\mathbf{X}_s : 0 \leq s \leq t_k\}$. *Clearly,* $(\mathcal{G}_k)_{1\leq k\leq n}$ *in a family of increasing* σ*-fields.*
- ***Random sampling.*** *Assume that the instants* $(t_k)_{1\leq k\leq n}$ *in the interval* $[0, T]$ *form a sequence of uniform random variables independent of the process* $\{\mathbf{X}_t : t \in [0, T]\}$. *Define*

$$0 \leq \tau_1 < \cdots < \tau_n \leq T,$$

as the corresponding order statistics. Observe that $(\tau_k)_{1\leq k\leq n}$ *are the observation points for the procedure. Clearly, all of the distances between these sites are positive. As a consequence, taking* $\mathcal{G}_k := \sigma(X(t_k))$ *the* σ*-field generated by* $\{\mathbf{X}_s : 0 \leq s \leq \tau_k\}$, *it follows that* $(\mathcal{G}_k)_{1\leq k\leq n}$ *is a sequence of increasing* σ*-fields.*

As established in [75], the penalization approach for the choice of the mesh δ *of the observations provides an optimal rate of convergence; we leave this subject open for future investigation within the context of ergodic processes.*

Remark 5. *In a previous publication [36], we tackled the nonparametric estimation of the density and regression function in a finite-dimensional setting using an orthonormal wavelet basis. Our findings differ significantly from those presented in this publication. In [36], we demonstrated the strong uniform consistency characteristics of these estimators over compact subsets of* $\mathbb{R}^d$ *under a general ergodic condition on the underlying processes. In addition, we demonstrate the asymptotic normality of wavelet-based estimators. The Burkholder–Rosenthal inequality, a more complex technique than the exponential inequality used in the previous publication, was the central concept of this study. Significantly, the current study studies the mean integrated square error over compact subsets, which is fundamentally different from the conclusions of the prior publication.*

5. Applications

5.1. The Conditional Distribution

Our findings can be utilized to examine the conditional distribution $F(\mathbf{y} \mid \mathbf{x})$ for $\mathbf{y} \in \mathbb{R}^d$. To be more precise, let $\rho(\mathbf{y}) = \mathbb{1}\{\mathbf{y} \leq \mathbf{z}\}$. The wavelet's hard thresholding estimator of $F(\mathbf{y} \mid \mathbf{x})$, is defined, for all $\mathbf{x} \in \mathbf{S}$, by

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

where

$$\breve{\eta}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\mathbb{1}\{\mathbf{Y}_t \leq \mathbf{y}\}}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt, \tag{31}$$

$$\breve{\theta}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\mathbb{1}\{\mathbf{Y}_t \leq \mathbf{y}\}}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})dt. \tag{32}$$

A direct consequence of Theorem 2 is

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

5.2. The Conditional Quantile

Remark that whenever $F(\cdot \mid \mathbf{x})$ is strictly increasing and continuous in a neighborhood of $q_\alpha(\mathbf{x})$, the function $F(\cdot \mid \mathbf{x})$ has a unique quantile of order α at a point $q_\alpha(\mathbf{x})$, which is $F(q_\alpha(\mathbf{x}) \mid \mathbf{x}) = \alpha$. In this situation

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

which can be estimated by $\widehat{q}_{T,\alpha}(x) = \widehat{F}^{-1}(\alpha \mid \mathbf{x})$. Consequently, the fact

$$F(q_\alpha(\mathbf{x}) \mid x) = \alpha = \widehat{F}(\widehat{q}_{T,\alpha}(\mathbf{x}) \mid \mathbf{x})$$

and $\widehat{F}(\cdot \mid \mathbf{x})$ is continuous and strictly increasing, we then have

$$\forall \epsilon > 0, \exists \eta(\epsilon) > 0, \forall y, \quad \left|\widehat{F}(y \mid \mathbf{x}) - \widehat{F}(q_\alpha(\mathbf{x}) \mid \mathbf{x})\right| \leq \eta(\epsilon) \Rightarrow |y - q_\alpha(\mathbf{x})| \leq \epsilon,$$

implying that, $\forall \epsilon > 0, \exists \eta(\epsilon) > 0$

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

In addition, suppose that, for fixed $\mathbf{x}_0 \in \mathcal{C} \subset \mathbf{S}$, $F(y \mid \mathbf{x}_0)$ is differentiable at $q_\alpha(\mathbf{x}_0)$ with

$$\frac{\partial}{\partial y} F(y \mid x_0)\Big|_{y=q_\alpha(\mathbf{x}_0)} := g(q_\alpha(x_0) \mid x_0) > \nu > 0,$$

where ν is a real number, and $g(\cdot \mid \mathbf{x})$ is uniformly continuous for all $\mathbf{x} \in \mathcal{C}$. From Taylor's expansion of the function $F(\widehat{q}_{T,\alpha}(x) \mid x)$ in the neighborhood of $q_\alpha(\mathbf{x})$, we have

$$F(\widehat{q}_{T,\alpha}(\mathbf{x}) \mid x) - F(q_\alpha(\mathbf{x}) \mid \mathbf{x}) = (\widehat{q}_{T,\alpha}(x) - q_\alpha(\mathbf{x}))g\big(q^*_{T,\alpha}(\mathbf{x}) \mid \mathbf{x}\big)$$

where $q^*_{T,\alpha}(\mathbf{x})$ is a point between $q_\alpha(\mathbf{x})$ and $\widehat{q}_{T,\alpha}(\mathbf{x})$. Using the fact that $\widehat{q}_{T,\alpha}(\mathbf{x})$ converges a.s. toward $q_\alpha(\mathbf{x})$ as T goes to infinity, combined with the uniform continuity of $g(\cdot \mid \mathbf{x})$, allows us to write that

$$\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{q}_{T,\alpha}(\mathbf{x}) - q_\alpha(\mathbf{x})| \sup_{\mathbf{x}\in\mathcal{C}}|g(q_\alpha(\mathbf{x}) \mid \mathbf{x})| = O_{a.s.}\left(\sup_{y\in S}\sup_{\mathbf{x}\in\mathcal{C}}\left|\widehat{F}_T(y \mid \mathbf{x}) - F(y \mid \mathbf{x})\right|\right).$$

By the fact that $g(q_\alpha(\mathbf{x}) \mid \mathbf{x})$ is uniformly bounded from below, we can then claim that

$$\mathbb{E}\Big(\|\widehat{q}_{T,\alpha}(\mathbf{x}) - q_\alpha(\mathbf{x})\|^2\Big) \leq C\left(\frac{\ln T}{T}\right)^\theta. \tag{34}$$

Remark 6. *(Expectile regression). For $p \in (0,1)$, the choice given by $\boldsymbol{\psi}(T - \boldsymbol{\theta}) = (p - \mathbb{1}\{T - \boldsymbol{\theta} \leq 0\})|T - \boldsymbol{\theta}|$ leads to quantities called expectiles by [76]. Expectiles, as defined by [76], may be introduced either as a generalization of the mean or as an alternative to quantiles. Indeed, classical regression provides us with a high sensitivity to extreme values, allowing for more reactive risk management. Quantile regression, on the other hand, provides the ability to acquire exhaustive information on the effect of the explanatory variable on the response variable by examining its conditional distribution; refer to [77–79] for further details on expectiles in functional data settings.*

Remark 7. *(Conditional winsorized mean). As in [80], if we consider $\boldsymbol{\psi}(T - \boldsymbol{\theta}) = -k, T - \boldsymbol{\theta}, k$ if $T - \boldsymbol{\theta} < -k$, $|T - \boldsymbol{\theta}| \leq k$, or $T - \boldsymbol{\theta} > k$, then $\mathbf{m}(\mathbf{x}; \psi)$ will be the conditional winsorized mean. Notably, this parameter was not considered in the literature on nonparametric functional*

data analysis involving wavelet estimators. Our paper offers asymptotic results for the conditional winsorized mean when the covariates are functions.

5.3. Shannon's Entropy

The differential (or Shannon) entropy of $f(\cdot)$ is defined to be

$$H(f) \quad = \quad -\int_{\mathbf{S}} f_{\mathbf{X}}(\mathbf{x}) \log(f_{\mathbf{X}}(\mathbf{x}))\nu(d\mathbf{x}), \tag{35}$$

whenever this integral is meaningful. We will apply the convention $0\log(0) = 0$ since $u\log(u) \to 0$ as $u \to 0$. The notion of differential entropy was originally introduced in [81]. Since this epoch, the concept of entropy has been the topic of extensive theoretical and applied research. We are referencing [82] (Chapter 8) for a thorough overview of differential entropy and its applications and mathematical characteristics. Entropy concepts and principles play a fundamental role in many applications, such as quantization theory [83], statistical decision theory [84], and contingency table analysis [85]. Ref. [86] introduced the concept of convergence in entropy and showed that the latter convergence concept implies convergence in $\mathcal{L}_1$. This property indicates that entropy is a useful concept for measuring "*closeness in distribution*", and also heuristically justifies the usage of sample entropy as a test statistic when designing entropy-based tests of goodness-of-fit. This line of research has been pursued by [87–89] (including the references therein). The idea here is that many families of distributions are characterized by maximization of entropy subject to constraints (see [90,91]). Given $\widehat{f}_T(\cdot)$ in Equation (13), we estimate $H(f)$ using representation (35), by setting

$$H_T(f) = -\int_{\mathbf{A}_n} \widehat{f}_T(\mathbf{x}) \log\left(\widehat{f}_T(\mathbf{x})\right)\nu(d\mathbf{x}), \tag{36}$$

where

$$\mathbf{A}_n = \{\mathbf{x} \in \mathbf{S} : \widehat{f}_T(\mathbf{x}) \geq \gamma_n\},$$

and $\gamma_n \downarrow 0$ is a sequence of positive constants. To prove the strong consistency of $H_T(f)$, we shall consider another, but more appropriate and more computationally convenient, centering factor than the expectation $\mathbb{E}H_T(f)$, which is delicate to handle. This is given by

$$\widehat{\mathbb{E}}H_T(f) = -\int_{\mathbf{A}_n} \mathbb{E}\widehat{f}_T(\mathbf{x}) \log\left(\mathbb{E}\widehat{f}_T(\mathbf{x})\right)\nu(d\mathbf{x}).$$

We first decompose $H_T(f) - \widehat{\mathbb{E}}H_T(f)$, as in [92–95], into the sum of two components, by writing

$$\begin{aligned} & H_T(f) - \widehat{\mathbb{E}}H_T(f) \\ = & -\int_{\mathbf{A}_n} \widehat{f}_T(\mathbf{x}) \log\left(\widehat{f}_T(\mathbf{x})\right)\nu(d\mathbf{x}) \\ & + \int_{\mathbf{A}_n} \mathbb{E}\widehat{f}_T(\mathbf{x}) \log\left(\mathbb{E}\widehat{f}_T(\mathbf{x})\right)\nu(d\mathbf{x}) \\ = & -\int_{\mathbf{A}_n} \left\{\log \widehat{f}_T(\mathbf{x}) - \log \mathbb{E}\widehat{f}_T(\mathbf{x})\right\}\mathbb{E}\widehat{f}_T(\mathbf{x})\nu(d\mathbf{x}) \\ & - \int_{\mathbf{A}_n} \left\{\widehat{f}_T(\mathbf{x}) - \mathbb{E}\widehat{f}_T(\mathbf{x})\right\} \log \widehat{f}_T(\mathbf{x})\nu(d\mathbf{x}). \end{aligned} \tag{37}$$

We remark that for all $z > 0$,

$$|\log z| \leq \left|\frac{1}{z} - 1\right| + |z - 1|.$$

Thus, for any $\mathbf{x} \in \mathbf{A}_n$, we obtain

$$\begin{aligned} |\log \widehat{f}_T(\mathbf{x}) - \log \mathbb{E}\widehat{f}_T(\mathbf{x})| &= \left|\log \frac{\widehat{f}_T(\mathbf{x})}{\mathbb{E}\widehat{f}_T(\mathbf{x})}\right| \\ &\leq \left|\frac{\mathbb{E}\widehat{f}_T(\mathbf{x})}{\widehat{f}_T(\mathbf{x})} - 1\right| + \left|\frac{\widehat{f}_T(\mathbf{x})}{\mathbb{E}\widehat{f}_T(\mathbf{x})} - 1\right| \\ &= \frac{\left|\mathbb{E}\widehat{f}_T(\mathbf{x}) - \widehat{f}_T(\mathbf{x})\right|}{\widehat{f}_T(\mathbf{x})} + \frac{\left|\widehat{f}_T(\mathbf{x}) - \mathbb{E}\widehat{f}_T(\mathbf{x})\right|}{\mathbb{E}\widehat{f}_T(\mathbf{x})}. \end{aligned}$$

Making use of Theorem 1, gives

$$\mathbb{E}\Big((H_T(f) - \widehat{\mathbb{E}}H_T(f))^2\Big) \leq C\left(\frac{\ln T}{T}\right)^{\theta}, \tag{38}$$

for positive constant C. Note that this is the first estimation result for the entropy of functional continuous time series processes. A similar idea can be used to estimate other density functions, such as the extropy; refer to [96] for a definition.

5.4. The Curve Discrimination

The curve discrimination problem can be stated as follows. Let $\{\mathbf{X}_t\}_{0\leq t\leq T}$ denote a sample of curves, and each of them is known to pertain to one among $\mathbf{G}$ groups $\boldsymbol{\iota} = 1, \ldots, \mathbf{G}$. We denote the group of the curve $\mathbf{X}_t$ by $\mathcal{T}_t$. Suppose that each pair of variables $(\mathbf{X}_t, \mathcal{T}_t)$ has the same distribution as the pair $(\mathbf{X}, \mathcal{T})$. Given a new curve $\mathbf{x}$, the question is to determine its class membership. To do so, we will estimate the conditional probability for every $\boldsymbol{\iota} = 1, \ldots, \mathbf{G}$, as follows:

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

Following the proposal made in [97,98] permitting the estimation of these probabilities by

$$\widehat{p}_{\boldsymbol{\iota}}(\mathbf{x}) = \sum_{\ell\in\mathcal{I}_0} \breve{\breve{\eta}}_{0,\ell}\phi_0(\mathbf{x};\zeta_{0,\ell}) + \sum_{k=0}^{m_T}\sum_{\ell\in\mathcal{J}_k} \breve{\breve{\theta}}_{k,\ell}\mathbb{1}\left\{|\breve{\breve{\theta}}_{k,\ell}| \geq \kappa\sqrt{\frac{\ln T}{T}}\right\}\psi_k(\mathbf{x};\eta_{k;\ell}), \tag{39}$$

where

$$\breve{\breve{\eta}}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\mathbb{1}\{T_t = \boldsymbol{\iota}\}}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt, \tag{40}$$

$$\breve{\breve{\theta}}_{k,\ell} = \frac{1}{T}\int_0^T \frac{\mathbb{1}\{T_t = \boldsymbol{\iota}\}}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})dt. \tag{41}$$

As remarked by [97,98], for each $\boldsymbol{\iota}$ we make use of the notation

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

then we can write

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

An application of Theorem 2 is

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

6. Concluding Remarks

In this study, we investigated the nonparametric estimation of density and regression function using continuous functional stationary processes and wavelet bases for Hilbert spaces of functions. We described the mean square error integrated over compact subsets. The martingale method was used to determine the asymptotic properties of these estimators, which differs significantly from the mixing and independent settings. Ergodicity is the assumption of the process's dependence. Extending nonparametric functional concepts to local stationary processes is a relatively new area of research. It would be interesting to extend our work to the case of the functional local stationary process, but this would require nontrivial mathematics and is well outside the scope of this paper. The exact logarithm rates of convergence depend on the smoothness parameters θ of functions $f(\cdot)$ and $\mathbf{m}(\cdot,\rho)$ defined in the space $\mathbf{H}$. These results are typical of all nonparametric estimations and are consistent with the vast majority of academic references. Although θ is unknown, the smoothness metrics assumed here are more nuanced than the differentiability criterion of integer orders required for convolution kernel methods. To make the estimator more practical, techniques have been developed to select the optimal adaptive value of τ, such that the most commonly used methods of "Stein," the "rule of thumb," and "cross-validation", we refer to [99] and [27] for details on an asymptotically optimal empirical bandwidth selection rule. It would be worthwhile to investigate this subject in the future.

7. Proofs

This section is dedicated to proving our findings. The previously presented notations are used again in the sequel. In this research, we need an upper bound inequality for partial sums of unbounded martingale differences to derive the asymptotic results for the estimations of the density and regression functions based on strictly stationary and ergodic functional data. Here and in the following, "C" denotes a positive constant that may vary from line to line. The following lemmas express this inequality.

Lemma 1. *(***Burkholder–Rosenthal inequality***) Following Notation* **1** *in [100].*

Let $(X_i)_{i\geq 1}$ be a stationary martingale adapted to the filtration $(\mathcal{F}_i)_{i\geq 1}$, define $(d_i)_{i\geq 1}$ as 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,

$$\| \max_{1\leq j\leq n} |S_j|)\|_p \ll n^{1/p}\|d_1\|_p + \left\| \sum_{k=1}^{n} \mathbb{E}(d_k^2/\mathcal{F}_{k-1}) \right\|_{p/2}^{1/2}, \quad \text{for any} \quad p \geq 2; \tag{42}$$

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

Lemma 2 ([101])**.** *Let $\{Z_i, i \geq 1\}$ denote a sequence of martingale differences in such a way that*

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

then, for all $\epsilon > 0$ and all n large enough, we obtain

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

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

Lemma 3. *For each $k \in \{0, \ldots, m_T\}$ and each $\ell \in \mathcal{I}_k$, under conditions (**C.0**), (**C.1**), (**F.1**), and (**E.1**)(i), there is a constant $C > 0$ in such a way that*

$$\mathbb{E}\left(\left|\widehat{\alpha}_{k,\ell} - \alpha_{k,\ell}\right|^2\right) \leq C\left(\frac{\ln T}{T}\right). \tag{43}$$

Lemma 4. *For each $k \in \{0, \ldots, m_T\}$ and each $\ell \in \mathcal{J}_k$, and under conditions (**C.0**), (**C.1**), (**F.1**), (**E.1**), and (**E.2**), and condition (16), there is a constant $C > 0$ in such a way that*

$$\mathbb{E}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right|^4\right) = C\left(\frac{\ln T}{T}\right)^2, \quad a.s. \tag{44}$$

Lemma 5. *For each $k \in \{0, \ldots, m_T\}$ and each $\ell \in \mathcal{J}_k$, for $\kappa > 0$ sufficiently large and under conditions (**C.0**), (**C.1**), (**F.1**), (**E.1**), and (**E.2**), and assumption (16) there is a constant $C > 0$ in such a way that*

$$\mathbb{P}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln T}{T}}\right) \leq C\left(\frac{\ln T}{T}\right)^2. \tag{45}$$

Proof of Theorem 1. Keep in mind that the proof of Theorem 1 is a straightforward application is a direct application of [52] (Theorem 3.1), using Lemmas 3–5 alongside $c(n) = (\ln T/T)^{1/2}$, $\sigma_i = 1$, $r = 2$. We modified and expanded the approach used to prove Theorem 3.1 of [34], and to include the stationary ergodic process. □

Proof of Theorem 3. Consider the subsequent decomposition

$$\begin{aligned} \widehat{\alpha}_{k,\ell} - \alpha_{k,\ell} &= \widehat{\alpha}_{k,\ell} - \widetilde{\alpha}_{k,\ell} + \widetilde{\alpha}_{k,\ell} - \alpha_{k,\ell} \\ &= A_{k,\ell,1} + A_{k,\ell,2}, \end{aligned} \tag{46}$$

where

$$\widetilde{\alpha}_{k,\ell} = \frac{1}{T}\int_0^T \mathbb{E}[\phi_k(\mathbf{x}_t; \zeta_{k,\ell})|\mathcal{F}_{t-\delta}]dt.$$

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

$$\begin{aligned} \widetilde{\alpha}_{k,\ell} &= \frac{1}{T}\int_0^T \int_S \phi_k(\mathbf{x}; \zeta_{k,\ell}) f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})\nu(d\mathbf{x})dt \\ &= \int_S \phi_k(\mathbf{x}; \zeta_{k,\ell})\left(\frac{1}{T}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt\right)\nu(d\mathbf{x}) \\ &= \int_S \phi_k(\mathbf{x}; \zeta_{k,\ell})(f(\mathbf{x}) + o(1))\nu(d\mathbf{x}) \\ &= \int_S \phi_k(\mathbf{x}; \zeta_{k,\ell}) f(\mathbf{x})\nu(d\mathbf{x}) + o(1) \\ &= \alpha_{k,\ell} + o(1). \end{aligned}$$

We readily obtain

$$\widetilde{\alpha}_{k,\ell} = \alpha_{k,\ell}, \quad \text{as}, \quad T \to \infty, \tag{47}$$

implying that

$$A_{k,\ell,2} = o(1), \quad \text{a.s.} \tag{48}$$

Therefore, we infer that

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

We now consider the term $A_{k,\ell,1}$. We infer

$$\begin{aligned} A_{k,\ell,1} &= \widehat{\alpha}_{k,\ell} - \widetilde{\alpha}_{k,\ell} \\ &= \frac{1}{T}\int_0^T (\phi_k(\mathbf{x}_t;\zeta_{k,\ell}) - \mathbb{E}[\phi_k(\mathbf{x}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}])dt \\ &= \frac{1}{n}\sum_{i=1}^{n} \Phi_k(\zeta_{k,\ell}), \end{aligned}$$

where

$$\Phi_{i,k}(\zeta_{k,\ell}) = \frac{1}{\delta}\int_{T_{i-1}}^{T_i} (\phi_k(\mathbf{x}_t;\zeta_{k,\ell}) - \mathbb{E}[\phi_k(\mathbf{x}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}])dt.$$

Notice that, with respect to the sequence of $\sigma-$fields $(\mathcal{F}_i)_{0\leq k\leq n}$, $(\Phi_{i,k}(\zeta_{k,\ell}))_{0\leq k\leq n}$ is a sequence of martingale differences. It is evident, by Lemma 1, to infer that

$$\mathbb{E}\Big[|A_{k,\ell,1}|^2\Big] = \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{i,k}(\zeta_{k,\ell})\right|^2\right],$$

where

$$\begin{aligned} \left(\mathbb{E}\Big[\Big|\textstyle\sum_{i=1}^{n}\Phi_{i,k}(\zeta_{k,\ell})\Big|^2\Big]\right)^{\frac{1}{2}} &\leq n^{1/2}\big\|\Phi_{1,k}(\zeta_{k,\ell})\big\|_2 + \Big\|\textstyle\sum_{i=1}^{n}\mathbb{E}\Big[\Phi^2_{i,k}(\zeta_{k,\ell})|\mathcal{F}_{i-1}\Big]\Big\|_1^{1/2} \\ &= \Phi_{(1)} + \Phi_{(2)}. \end{aligned} \tag{49}$$

Observe that for all $0 \leq t \leq \delta$, the σ-field $\mathcal{F}_{t-\delta} = \mathcal{F}_0$ represents the trivial one. On the one hand, by using the classical decomposition in connection with the fact that $\mathcal{F}_0$ is the trivial $\sigma-$field, we have

$$\begin{aligned} \frac{1}{n}\Phi^2_{(1)} &= \| \Phi_{1,k}(\zeta_{k,\ell}) \|_2^2 \\ &= \mathbb{E}\left[\left|\tfrac{1}{\delta}\textstyle\int_0^\delta(\phi_k(\mathbf{x}_t;\zeta_{k,\ell}) - \mathbb{E}[\phi_k(\mathbf{x}_t;\zeta_{k,\ell})|\mathcal{F}_0])dt\right|^2\right] \\ &\leq \tfrac{1}{\delta^2}\textstyle\int_0^\delta \mathbb{E}\Big[\sum_{j=0}^{2} \mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid^j (\mathbb{E}[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid])^{2-j}\Big]dt \\ &= \tfrac{1}{\delta^2}\textstyle\int_0^\delta \Big(\sum_{j=0}^{2} C_2^j\mathbb{E}\big[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid^j\big].(\mathbb{E}[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid])^{2-j}\Big)dt \\ &= \tfrac{1}{\delta^2}\textstyle\int_0^\delta \Big(C_2^2\mathbb{E}\big[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid^2\big] + C_2^1(\mathbb{E}[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid])^2 + C_2^0\mathbb{E}\big[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid^2\big]\Big)dt. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}\Big[\big|\phi_k(\mathbf{x}_t;\zeta_{k,\ell})\big|^2\Big] &= \textstyle\int_{\mathbf{S}}\big|\phi_k(\mathbf{x};\zeta_{k,\ell})\big|^2 f(\mathbf{x})\nu(d\mathbf{x}) \\ &\leq C_f\textstyle\int_{\mathbf{S}}\big|\phi_k(\mathbf{x};\zeta_{k,\ell})\big|^2\nu(d\mathbf{x}) \\ &= C_f\textstyle\int_{\mathbf{S}}\Big|\sum_{j\in\mathcal{I}_k}\frac{1}{\sqrt{g_{j,k,\ell}}}e_j(\zeta_{k,\ell})e_j(\mathbf{x})\Big|^2\nu(d\mathbf{x}) \\ &= C_f\textstyle\int_{\mathbf{S}}\sum_{j\in\mathcal{I}_k}\frac{1}{g_{j,k,\ell}}\big|e_j(\zeta_{k,\ell})\big|^2\nu(d\mathbf{x}) \\ &\leq C_fC_1, \end{aligned} \tag{50}$$

where C_1 is a positive constant,

$$\mathbb{E}\Big[\mid \phi_k(\mathbf{x}_t;\zeta_{k,\ell}) \mid^2\Big] = O(1), \tag{51}$$

therefore,

$$\Phi_{(1)} = O(T^{1/2}). \tag{52}$$

Furthermore, we examine the second term of decomposition (49), and remark that

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

using the notable identity combined with Jensen's inequality, we obtain

$$\begin{aligned}\mathbb{E}\left[\Phi_{i,k}^2(\zeta_{k,\ell})\right] &= \mathbb{E}\left[\left(\left|\frac{1}{\delta}\int_{T_{i-1}}^{T_i}(\phi_k(\mathbf{x}_t;\zeta_{k,\ell})-\mathbb{E}[\phi_k(\mathbf{x}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}])dt\right|\right)^2\right]\\ &\leq \frac{1}{\delta^2}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\left(\left|(\phi_k(\mathbf{x}_t;\zeta_{k,\ell})-\mathbb{E}[\phi_k(\mathbf{x}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}])\right|\right)^2\right]dt\\ &\leq \frac{1}{\delta^2}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[2\mid\phi_k(\mathbf{x}_t;\zeta_{k,\ell})\mid^2+2\mathbb{E}\left[\mid\phi_k(\mathbf{x}_t;\zeta_{k,\ell})\mid^2|\mathcal{F}_{t-\delta}\right]\right]dt\\ &\leq \frac{1}{\delta^2}\int_{T_{i-1}}^{T_i}\left(2\mathbb{E}\left[\mid\phi_k(\mathbf{x}_t;\zeta_{k,\ell})\mid^2\right]+2\mathbb{E}\left[\mathbb{E}\left[\mid\phi_k(\mathbf{x}_t;\zeta_{k,\ell})\mid^2|\mathcal{F}_{t-\delta}\right]\right]\right)dt\\ &\leq \frac{4}{\delta^2}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\mid\phi_k(\mathbf{x}_t;\zeta_{k,\ell})\mid^2\right]dt,\end{aligned}$$

observe that, using (50), we have

$$\Phi_2 = O(T^{1/2}). \tag{53}$$

Therefore, combining Equations (52) and (53), we obtain

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{i,k}(\zeta_{k,\ell})\right|^2\right]\right)^{\frac{1}{2}} = O(T^{1/2}).$$

Hence,

$$\begin{aligned}\mathbb{E}\left[|A_{k,\ell,1}|^2\right] &= \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{i,k}(\zeta_{k,\ell})\right|^2\right]\\ &= \frac{\delta^2}{T^2}O(T)\\ &\leq C\left(\frac{\ln T}{T}\right).\end{aligned}$$

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

$$\mathbb{E}\left(\left|\widehat{\alpha}_{k,\ell}-\alpha_{k,\ell}\right|^2\right) \leq \frac{4C\delta}{T} \leq 4C\left(\frac{\ln T}{T}\right). \tag{54}$$

Hence the proof is complete. □

Proof of Lemma 4. Consider the following decomposition

$$\begin{aligned}\widehat{\beta}_{k,\ell}-\beta_{k,\ell} &= \widehat{\beta}_{k,\ell}-\widetilde{\beta}_{k,\ell}+\widetilde{\beta}_{k,\ell}-\beta_{k,\ell}\\ &= B_{k,\ell,1}+B_{k,\ell,2},\end{aligned} \tag{55}$$

where

$$\widetilde{\beta}_{k,\ell} = \frac{1}{T}\int_0^T \mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})|\mathcal{F}_{t-\delta}]dt.$$

Remark that, under conditions (**F.1**) and (**E.1**)(i) and making use of the fact that $\mathcal{E}$ is an orthonormal basis of $\mathbf{H}$, and reasoning in a similar way as in Equation (47), we infer that

$$\widetilde{\beta}_{k,\ell} = \beta_{k,\ell}, \quad \text{as}, \quad n \to \infty. \tag{56}$$

This, in turn, implies that

$$B_{k,\ell,2} = o(1), \quad \text{a.s.} \tag{57}$$

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

$$\mathbb{E}\left(\left|\widehat{\beta}_{k,\ell} - \beta_{k,\ell}\right|^4\right) \quad = \quad \frac{1}{n^4}\mathbb{E}\left(\left|\sum_{i=1}^n \Psi_{i,k,\ell}\right|^4\right), \tag{58}$$

where

$$\Psi_{i,k,\ell} = \frac{1}{\delta}\int_{T_{i-1}}^{T_i} (\psi_k(\mathbf{x}_t;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}])dt.$$

Remark that, with respect to the sequence of $\sigma-$fields $(\mathcal{F}_{i-1})_{0\le k\le n}$, $(\Psi_{i,k,\ell})_{0\le k\le n}$ is a sequence of martingale differences, an application of the Burkholder–Rosenthal inequality (see Lemma 1), gives

$$\begin{aligned}\left(\mathbb{E}\left(\left|\sum_{i=1}^n \Psi_{i,k,\ell}\right|^4\right)\right)^{1/4} &\le \left\|\max_{1\le j\le n}\left|\sum_{i=1}^j \Psi_{i,k,\ell}\right|\right\|_4 \\ &\ll n^{1/4}\|\Psi_{1,k,\ell}\|_4 + \left\|\sum_{i=1}^n \mathbb{E}(\Psi_{i,k,\ell}^2|\mathcal{F}_{i-2})\right\|_{4/2}^{1/2} \\ &= \Psi_{k,\ell}^{(1)} + \Psi_{k,\ell}^{(2)}.\end{aligned} \tag{59}$$

Consider the first term of Equation (59). We recall that for all $0 \le t \le \delta$, the σ-field $\mathcal{F}_{t-\delta} = \mathcal{F}_0$ represents the trivial one. Applying Jensen's inequality, we have

$$\begin{aligned}\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 &= \|\Psi_{1,k,\ell}\|_4^4 \\ &= \mathbb{E}\left(\left|\frac{1}{\delta}\int_0^\delta (\psi_k(\mathbf{x}_t;\eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_t;\eta_{k,\ell})|\mathcal{F}_0])dt\right|^4\right) \\ &\le \frac{1}{\delta^4}\int_0^\delta \mathbb{E}\Big[\left(|\psi_k(\mathbf{x}_t;\eta_{k,\ell})| + \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|\right]\right)^4\Big]dt.\end{aligned}$$

Using the Minkowski inequality, we obtain

$$\begin{aligned}&\int_0^\delta \mathbb{E}\Big[\left(|\psi_k(\mathbf{x}_t;\eta_{k,\ell})| + \mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|\right]\right)^4\Big]dt \\ &= \int_0^\delta \left(\mathbb{E}^{1/4}\Big[|\psi_k(\mathbf{x}_t;\eta_{k,\ell})|^4\Big] + \mathbb{E}^{1/4}\Big[\left(\mathbb{E}\left[|\psi_k(\mathbf{x}_1;\eta_{k,\ell})|\right]\right)^4\Big]\right)^4 dt \\ &= 16\int_0^\delta \mathbb{E}\Big[|\psi_k(\mathbf{x}_t;\eta_{k,\ell})|^4\Big]dt.\end{aligned} \tag{60}$$

This gives that

$$\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 \leq 16\int_0^{\delta}\left(\mathbb{E}^{1/4}\left[\left|\psi_k(\mathbf{x}_t;\eta_{k,\ell})\right|^4\right]\right)^4 dt.$$

By reasoning in a similar way as in Equation (50) and making use of conditions (**F.1**) and (**E.1**)(i), for all $0 \leq t \leq T$, we infer that

$$\mathbb{E}\left[\left|\psi_k(\mathbf{x}_t;\eta_{k,\ell})\right|^2\right] \leq C, \tag{61}$$

where C denotes a positive constant. In addition, by the Cauchy–Schwarz inequality in combination with conditions (**E.1**)(ii), (**E.2**) and condition (16), we infer

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})| &\leq \sup_{\mathbf{x}\in\mathbf{S}}\sum_{j\in\mathcal{J}_k}\frac{1}{\sqrt{h_{j,k,\ell}}}|e_j(\eta_{k,\ell})||e_j(\mathbf{x})| \\
&\leq \left(\sum_{j\in\mathcal{J}_k}\frac{1}{h_{j,k,\ell}}|e_j(\eta_{k,\ell})|^2\right)^{1/2}\left(\sup_{\mathbf{x}\in\mathbf{S}}\sum_{j\in\mathcal{J}_k}|e_j(\mathbf{x})|^2\right)^{1/2} \\
&\leq C_1^{1/2}C_2^{1/2}\sqrt{|\mathcal{J}_k|} \\
&\leq C_4\sqrt{|\mathcal{J}_{m_n}|} \\
&\leq C_4\sqrt{\frac{T}{\ln T}}.
\end{aligned} \tag{62}$$

We then obtain

$$\mathbb{E}\left[\left|\psi_k(\mathbf{x}_1;\eta_{k,\ell})\right|^4\right] = O\left(\frac{T}{\ln T}\right). \tag{63}$$

Recall that $T = \delta n$, we deduce that

$$\Psi_{k,\ell}^{(1)} = O\left(\frac{T^{1/2}}{(\ln T)^{1/4}}\right). \tag{64}$$

We now consider the upper bound of $\Psi_{k,\ell}^{(2)}$ in Equation (59). Remark that

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

Observe that for all $T_{i-1} \leq t \leq T_i$ for all $i = 1, \ldots, n$; we have $\mathcal{F}_{i-2} \subset \mathcal{F}_{t-\delta}$. Making use of Jensen and Minkowski's inequality, it follows that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\Big[\Psi_{i,k,\ell}^2 \mid \mathcal{F}_{i-2}\Big] \\
&= \sum_{i=1}^{n} \left(\mathbb{E}\left[\left(\frac{1}{\delta} \int_{T_{i-1}}^{T_i} (\psi_k(\mathbf{x}_t; \eta_{k,\ell}) - \mathbb{E}[\psi_k(\mathbf{x}_t; \eta_{k,\ell}) \mid \mathcal{F}_{t-\delta}])dt \right)^2 \mid \mathcal{F}_{i-2} \right] \right) \\
&\leq \frac{1}{\delta^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \left(\mathbb{E}^{1/2}\Big[\psi_k^2(\mathbf{x}_t; \eta_{k,\ell}) \mid \mathcal{F}_{i-2}\Big] + \mathbb{E}^{1/2}\Big[\mathbb{E}\Big[\psi_k^2(\mathbf{x}_t; \eta_{k,\ell}) \mid \mathcal{F}_{t-\delta}\Big] \mid \mathcal{F}_{i-2}\Big] \right)^2 dt \\
&= \frac{4}{\delta^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\Big[\psi_k^2(\mathbf{x}_t; \eta_{k,\ell}) \mid \mathcal{F}_{i-2}\Big] dt.
\end{aligned}$$

For all $T_{i-1} \leq t \leq T_i$, using the stationarity of the process $(\mathbf{X}_t)_{t \geq 0}$, we have

$$f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) = f_{T_{i-1}}^{\mathcal{F}_{i-2}}(\mathbf{x}).$$

Remark, under conditions (**F.1**), (**E.1**)(i), (**C.0**), and (**C.1**) and statement (61), that

$$\begin{aligned}
\sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\Big[(\psi_k(\mathbf{x}_t; \eta_{k,\ell}))^2 \mid \mathcal{F}_{i-2}\Big] dt &= n \int_{\mathbf{S}} \big|\psi_k(\mathbf{x}; \eta_{k,\ell})\big|^2 \left(\frac{1}{n} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) dt \right) \nu(d\mathbf{x}) \\
&\leq n \int_{\mathbf{S}} \big|\psi_k(\mathbf{x}; \eta_{k,\ell})\big|^2 \left(\delta \frac{1}{n} \sum_{i=1}^{n} f_{T_{i-1}}^{\mathcal{F}_{i-2}}(\mathbf{x}) \right) \nu(d\mathbf{x}) \\
&= n \int_{\mathbf{S}} \big|\psi_k(\mathbf{x}; \eta_{k,\ell})\big|^2 (f(\mathbf{x}) + o(1)) \nu(d\mathbf{x}) \\
&\leq n\Big(C_f + o(1)\Big) \int_{\mathbf{S}} \big|\psi_k(\mathbf{x}; \zeta_{k,\ell})\big|^2 \nu(d\mathbf{x}) \\
&\leq \frac{T C_f C_1}{\delta}.
\end{aligned} \tag{65}$$

It follows that

$$\Psi_{k,\ell}^{(2)} = O\Big(T^{1/2}\Big). \tag{66}$$

Combining statements (59), (64), and (66), we obtain

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

We conclude that

$$\mathbb{E}\left(\left| \widehat{\beta}_{k,\ell} - \beta_{k,\ell} \right|^4 \right) = O\left(\frac{1}{T^2 \ln T} \right) + O\left(\frac{1}{T^2} \right). \tag{67}$$

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

$$\mathbb{E}\left(\left| \widehat{\beta}_{k,\ell} - \beta_{k,\ell} \right|^4 \right) \leq C\left(\frac{\ln T}{T} \right)^2. \tag{68}$$

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

$$\begin{aligned}B_{k,\ell,1} &= \frac{1}{n}\sum_{i=1}^{n}\Psi_{i,k,\ell}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\delta}\int_{T_{i-1}}^{T_i}(\psi_k(\mathbf{x}_t;\eta_{k,\ell})-\mathbb{E}[\psi_k(\mathbf{x}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}])dt\right),\\ B_{k,\ell,2} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\delta}\int_{T_{i-1}}^{T_i}(\mathbb{E}[\psi_k(\mathbf{x}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}]-\beta_{k,\ell})dt\right).\end{aligned}$$

Making use of Equation (57), we achieve the desired result for the term $B_{k,\ell,2}$

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

We now remark

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

The application of Lemma 2 implies that

$$\begin{aligned}\left|\Psi_{i,k,\ell}\right| &= \left|\psi_k(\mathbf{x}_i;\eta_{k,\ell})-\mathbb{E}[\psi_k(\mathbf{x}_i;\eta_{k,\ell})\mid\mathcal{F}_{i-1}]\right|\\ &\leq 2\sup_{\mathbf{x}\in\mathbf{S}}\left|\psi_k(\mathbf{x};\eta_{k,\ell})\right|\\ &\leq C_3\sqrt{\tfrac{T}{\ln T}}\\ &\leq C_3\sqrt{T}.\end{aligned}\tag{69}$$

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

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right|\geq\frac{\kappa}{2}\sqrt{T\ln T}\right) &\leq 2\exp\left\{-\frac{\epsilon_T^2}{2nB^2}\right\}\\ &= 2\exp\left\{-\frac{\left(\frac{\delta\kappa}{2}\sqrt{T\ln T}\right)^2}{2(T/\delta)\left(C_3\sqrt{T}\right)^2}\right\}\\ &= 2\exp\left\{-\frac{\delta^3\kappa^2\ln T}{8C_3^2T}\right\}\\ &= 2\exp\left\{\ln T^{-\frac{\delta^3\kappa^2}{8C_3^2T}}\right\}\\ &= 2T^{-w(\kappa,T)},\end{aligned}\tag{70}$$

where

$$w(\kappa,T)=\frac{\delta^3\kappa^2}{8C_3^2T}.$$

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

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{\beta}_{k,\ell}-\beta_{k,\ell}\right|\geq\frac{\kappa}{2}\sqrt{\frac{\ln T}{T}}\right) &\leq C\frac{1}{T^2}+o(1)\\ &\leq C\left(\frac{\ln T}{T}\right)^2. \end{aligned}\tag{71}$$

The proof of (45) is achieved. □

Recall that

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

where

$$\widehat{\eta}_{k,\ell} = \frac{1}{T}\int_0^T\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})dt,\quad \widehat{\theta}_{k,\ell}=\frac{1}{T}\int_0^T\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})dt.$$

Lemma 6. *For each $k\in\{0,\ldots,m_T\}$ and each $\ell\in\mathcal{I}_k$ and under conditions (**E.1**)(i), (**M.1**), (**M.2**), (**C.0**), and (**C.1**), there is a constant $C>0$ in such a way that*

$$\mathbb{E}\left(\left|\widehat{\eta}_{k,\ell}-\eta_{k,\ell}\right|^2\right)\leq C\left(\frac{\ln T}{T}\right).\tag{72}$$

Lemma 7. *For each $k\in\{0,\ldots,m_T\}$ and each $\ell\in\mathcal{J}_k$, and under conditions (**E.1**), (E.2) (**M.1**), (**M.2**), (**C.0**), and (**C.1**), in combination with condition (24), there is a constant $C>0$ in such a way that*

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right)=C\left(\frac{\ln T}{T}\right)^2,\quad a.s.\tag{73}$$

Lemma 8. *For each $k\in\{0,\ldots,m_T\}$ and each $\ell\in\mathcal{J}_k$, for $\kappa>0$ sufficiently large, (**E.1**), (**E.2**) (**M.1**), (**M.2**), (**C.0**), and (**C.1**), in combination with condition (24), there is a constant $C>0$ in such a way that*

$$\mathbb{P}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|\geq\frac{\kappa}{2}\sqrt{\frac{\ln T}{T}}\right)\leq C\left(\frac{\ln T}{T}\right)^2.\tag{74}$$

Proof of Lemma 2. Remark that the proof of Theorem 2 is a consequence of [52] (Theorem 3.1) with $c(n)=(\ln T/T)^{1/2}$, $\sigma_i=1$, $r=2$ in connection with Lemmas 6–8. We generalized the method of the proof in [34] (Theorem 4.1). □

Proof of Lemma 6. Consider the subsequent decomposition

$$\begin{aligned}\widehat{\eta}_{k,\ell}-\eta_{k,\ell} &= \widehat{\eta}_{k,\ell}-\widetilde{\eta}_{k,\ell}+\widetilde{\eta}_{k,\ell}-\eta_{k,\ell}\\ &= A_{k,\ell,1}+A_{k,\ell,2},\end{aligned}\tag{75}$$

where

$$\begin{aligned}\widetilde{\eta}_{k,\ell} &= \frac{1}{T}\int_0^T\mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]dt\\ &= \frac{1}{T}\int_0^T\mathbb{E}\left[\frac{(\mathbf{m}(\mathbf{X}_t,\rho)+\epsilon_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]dt\\ &= \frac{1}{T}\int_0^T\mathbb{E}\left[\frac{\mathbf{m}(\mathbf{X}_t,\rho)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]dt+\frac{1}{T}\int_0^T\mathbb{E}\left[\frac{\epsilon_t}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]dt.\end{aligned}$$

For all $0 \leq t \leq T$; we have $\mathcal{F}_{t-\delta} \subset \mathcal{S}_{t-\delta}$ and from the independence between ϵ_t and $\mathbf{X}_t$, we have

$$\begin{aligned} \mathbb{E}[\epsilon_t|\mathcal{S}_{t-\delta}] &= \mathbb{E}[\epsilon_t|\mathbf{X}_t] \\ &= \mathbb{E}[\epsilon_t] \\ &= 0. \end{aligned} \tag{76}$$

Observe that

$$\begin{aligned} \mathbb{E}\left[\frac{\epsilon_t}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right] &= \mathbb{E}\left[\frac{\mathbb{E}[\epsilon_t|\mathcal{S}_{t-\delta}]}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[\epsilon_t|\mathbf{X}_t]}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[\epsilon_t]}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right] \\ &= 0. \end{aligned}$$

This implies that

$$\widetilde{\eta}_{k,\ell} = \frac{1}{T}\int_0^T \mathbb{E}\left[\frac{\mathbf{m}(\mathbf{X}_t,\rho)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]dt.$$

Using conditions (**M.1**), (**M.2**), (**C.0**), and (**C.1**), we obtain

$$\begin{aligned} \widetilde{\eta}_{k,\ell} &= \frac{1}{T}\int_0^T \int_{\mathcal{S}} \frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell}) f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})\nu(d\mathbf{x})dt \\ &= \int_{\mathbf{S}} \frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})\left(\frac{1}{T}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt\right)\nu(d\mathbf{x}) \\ &= \int_{\mathbf{S}} \frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})(f(\mathbf{x})+o(1))\nu(d\mathbf{x}) \\ &= \int_{\mathbf{S}} \frac{\mathbf{m}(\mathbf{x},\rho)}{f(\mathbf{x})}\phi_k(\mathbf{x};\zeta_{k,\ell})f(\mathbf{x})\nu(d\mathbf{x}) + o(1) \\ &= \eta_{k,\ell} + o(1). \end{aligned}$$

We readily obtain that

$$\widetilde{\eta}_{k,\ell} = \eta_{k,\ell}, \quad \text{as}, \quad T \to \infty, \tag{77}$$

implying that

$$A_{k,\ell,2} = o(1), \quad \text{a.s.} \tag{78}$$

Therefore, we infer that

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

Let us now concentrate on the term $A_{k,\ell,1}$ in Equation (75), we infer

$$\begin{aligned} A_{k,\ell,1} &= \widehat{\eta}_{k,\ell} - \widetilde{\eta}_{k,\ell} \\ &= \frac{1}{T}\int_0^T \left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt \\ &= \frac{1}{n}\sum_{i=1}^n \left(\frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt\right) \\ &= \frac{1}{n}\sum_{i=1}^n \Phi_{k,i}(\zeta_{k,\ell}), \end{aligned}$$

where

$$\Phi_{k,i}(\zeta_{k,\ell}) = \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt.$$

Remark that, with respect to the sequence of $\sigma-$fields $(\mathcal{F}_{i-1})_{0\leq k\leq m_T}$, $(\Phi_{k,i}(\zeta_{k,\ell}))_{0\leq k\leq m_T}$ is a sequence of martingale differences. It is obvious, going with the proof of Equation (54), to observe that

$$\mathbb{E}\left[|A_{k,\ell,1}|^2\right] = \frac{1}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{k,i}(\zeta_{k,\ell})\right|^2\right],$$

where

$$\begin{aligned}\left(\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{k,i}(\zeta_{k,\ell})\right|^2\right]\right)^{\frac{1}{2}} &\leq n^{1/2}\parallel\Phi_{k,1}(\zeta_{k,\ell})\parallel_2 + \parallel\sum_{i=1}^{n}\mathbb{E}\left[\Phi_{k,i}^2(\zeta_{k,\ell})|\mathcal{F}_{i-2}\right]\parallel_1^{1/2}\\ &= \Phi_{(1)}+\Phi_{(2)}. \end{aligned} \tag{79}$$

On the one hand, using Jensen and Minkowski's inequalities combined, we obtain

$$\begin{aligned}\frac{1}{n}\Phi_{(1)}^2 &= \parallel\Phi_{k,1}(\zeta_{k,\ell})\parallel_2^2\\ &= \mathbb{E}\left[\left|\frac{1}{\delta}\int_0^\delta\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt\right|^2\right]\\ &\leq \frac{1}{\delta^2}\int_0^\delta\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right|^2\right]dt\\ &\leq \frac{1}{\delta^2}\int_0^\delta\left(\mathbb{E}^{1/2}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^2\right] + \mathbb{E}^{1/2}\left[\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^2|\mathcal{F}_{t-\delta}\right]\right]\right)^2dt\\ &\leq \frac{4}{\delta^2}\int_0^\delta\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^2\right]dt.\end{aligned}$$

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

$$|\rho(Y_t)| \leq C_{\mathbf{m}} + |\epsilon_t|, \tag{80}$$

combined with the independence between $\mathbf{X}_t$ and ϵ_t, $\mathbb{E}[\epsilon_t^2] = 1$. Remark that, under condition (**E.1**)(i) and using the fact that $\mathcal{E}$ is an orthonormal basis of $\mathbf{H}$, we infer

$$\begin{aligned}\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^2\right] &\leq \left(C_{\mathbf{m}}^2+1\right)\frac{1}{C_f}\mathbb{E}\left[\left|\frac{1}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^2\right]\\ &\leq \left(C_{\mathbf{m}}^2+1\right)\int_{\mathbf{S}}\frac{1}{f(\mathbf{x})}|\phi_k(\mathbf{x};\zeta_{k,\ell})|^2f(\mathbf{x})\nu(d\mathbf{x})\\ &= \left(C_{\mathbf{m}}^2+1\right)\int_{\mathbf{S}}\left|\sum_{j\in\mathcal{I}_k}\frac{1}{\sqrt{g_{j,k,\ell}}}e_j(\zeta_{k,\ell})e_j(\mathbf{x})\right|^2\nu(d\mathbf{x})\\ &= \left(C_{\mathbf{m}}^2+1\right)\int_{\mathbf{S}}\sum_{j\in\mathcal{I}_k}\frac{1}{g_{j,k,\ell}}|e_j(\zeta_{k,\ell})|^2\nu(d\mathbf{x})\\ &\leq \left(C_{\mathbf{m}}^2+1\right)C_1 = O(1).\end{aligned} \tag{81}$$

Therefore, we have

$$\Phi_{(1)} = \left(\frac{\delta n (C_{\mathbf{m}}^2 + 1) C_1)}{\delta^2} \right)^{1/2} = O(T^{1/2}). \tag{82}$$

Furthermore, we examine the second term of decomposition (79) and continue as per the proof of (53) and considering (81), one obtains

$$\Phi_2 = O(T^{1/2}). \tag{83}$$

therefore, combining Equations (82) and (83), we obtain

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

Hence, there exists a positive constant C such that

$$\begin{aligned} \mathbb{E}\left[|A_{k,\ell,1}|^2 \right] &= \frac{1}{n^2} \mathbb{E}\left[\left| \sum_{i=1}^{n} \Phi_k(\mathbf{x}_i; \zeta_{k,\ell}) \right|^2 \right] \\ &= \frac{\delta^2}{T^2} O(T) \\ &\leq C \left(\frac{\ln T}{T} \right). \end{aligned} \tag{84}$$

Therefore, combining Equations (78) and (84) there exists a constant C such as

$$\mathbb{E}\left(\left| \widehat{\alpha}_{k,\ell} - \alpha_{k,\ell} \right|^2 \right) \leq C \left(\frac{\ln T}{T} \right). \tag{85}$$

Thus, the proof of Lemma 6 is completed. □

Proof of Lemma 7. Consider the subsequent decomposition

$$\begin{aligned} \widehat{\theta}_{k,\ell} - \theta_{k,\ell} &= \widehat{\theta}_{k,\ell} - \breve{\theta}_{k,\ell} + \breve{\theta}_{k,\ell} - \theta_{k,\ell} \\ &= B_{k,\ell,1} + B_{k,\ell,2}, \end{aligned} \tag{86}$$

where

$$\breve{\theta}_{k,\ell} = \frac{1}{T} \int_0^T \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)} \psi_k(\mathbf{X}_t; \eta_{k,\ell}) | \mathcal{F}_{t-\delta} \right] dt.$$

Remark that, under conditions (**M.1**), (**M.2**), and (**C.1**), Equation (76), and using the fact that $\mathcal{E}$ is an orthonormal basis of $\mathbf{H}$, by reasoning as in Equation (77), we obtain

$$\breve{\theta}_{k,\ell} = \theta_{k,\ell}, \quad \text{as}, \quad n \to \infty, \tag{87}$$

giving that

$$B_{k,\ell,2} = o(1), \quad \text{a.s.} \tag{88}$$

Therefore, we have

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

Hence we obtain

$$\mathbb{E}\left(\left| \widehat{\theta}_{k,\ell} - \theta_{k,\ell} \right|^4 \right) = \frac{1}{n^4} \mathbb{E}\left(\left| \sum_{i=1}^{n} \Psi_{i,k,\ell} \right|^4 \right), \tag{89}$$

where

$$\Psi_{i,k,\ell} = \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\frac{Y_t}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell}) - \mathbb{E}\left[\frac{Y_i}{f(\mathbf{x}_t)}\psi_k(\mathbf{x}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt.$$

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 $(\mathcal{F}_{i-1})_{0\leq k\leq n}$, applying the Burkholder–Rosenthal inequality (see Lemma 1), we obtain

$$\begin{aligned}\left(\mathbb{E}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right|^4\right)\right)^{1/4} &\leq \left\|\max_{1\leq j\leq n}\left|\sum_{i=1}^{j}\Psi_{i,k,\ell}\right|\right\|_4 \\ &\ll n^{1/4}\|\Psi_{1,k,\ell}\|_4 + \left\|\sum_{i=1}^{n}\mathbb{E}(\Psi_{i,k,\ell}^2|\mathcal{F}_{i-2})\right\|_{4/2}^{1/2} \\ &= \Psi_{k,\ell}^{(1)} + \Psi_{k,\ell}^{(2)}.\end{aligned} \tag{90}$$

Consider the first term of the Equation (90). Applying Jensen and Minkowski's inequalities, we have

$$\begin{aligned}\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 &= \|\Psi_{1,k,\ell}\|_4^4 \\ &= \mathbb{E}\left(\left|\frac{1}{\delta}\int_0^\delta\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell}) - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt\right|^4\right) \\ &\leq \frac{1}{\delta^4}\int_0^\delta\mathbb{E}\left[\left(\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\right| + \mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\right||\mathcal{F}_{t-\delta}\right]\right)^4\right] \\ &\leq \frac{1}{\delta^4}\int_0^\delta\left(\mathbb{E}^{1/4}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^4\right] + \mathbb{E}^{1/4}\left[\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^4|\mathcal{F}_{t-\delta}\right]\right]\right)^4 dt \\ &\leq \frac{16}{\delta^4}\int_0^\delta\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\phi_k(\mathbf{X}_t;\zeta_{k,\ell})\right|^4\right]dt.\end{aligned} \tag{91}$$

By proceeding as in Equation (81) and under the same conditions (**M.1**), (**M.2**), and (**E.1**)(i), for all $0\leq t\leq T$, we obtain

$$\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\right|^2\right] \leq C, \tag{92}$$

where C is a positive constant. Moreover, by the Cauchy–Schwarz inequality in connection with conditions (**E.1**)(ii) and (**E.2**) and assumption (24), we have

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})| &\leq \sup_{\mathbf{x}\in\mathbf{S}}\sum_{j\in\mathcal{J}_k}\frac{1}{\sqrt{h_{j,k,\ell}}}|e_j(\eta_{k,\ell})||e_j(\mathbf{x})| \\ &\leq \left(\sum_{j\in\mathcal{J}_k}\frac{1}{h_{j,k,\ell}}|e_j(\eta_{k,\ell})|^2\right)^{1/2}\left(\sup_{\mathbf{x}\in\mathbf{S}}\sum_{j\in\mathcal{J}_k}|e_j(\mathbf{x})|^2\right)^{1/2} \\ &\leq C_1^{1/2}C_2^{1/2}\sqrt{|\mathcal{J}_k|} \\ &\leq C_3\sqrt{|\mathcal{J}_{m_T}|} \\ &\leq C_3\sqrt{\frac{T}{(\ln T)^2}}.\end{aligned} \tag{93}$$

From statements (92) and (93), we deduce that

$$\frac{1}{n}\left(\Psi_{k,\ell}^{(1)}\right)^4 \leq C_4\left(\frac{T}{(\ln T)^2}\right).$$

It follows that

$$\Psi_{k,\ell}^{(1)} \leq \frac{C_4}{\delta}\left(\frac{T}{\ln T}\right)^{1/2}. \tag{94}$$

Let us now examine the upper bound of $\Psi_{k,\ell}^{(2)}$ of Equation (59). Remark that

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

Observe that for all $T_{i-1} \leq t \leq T_i$ for all $i = 1, \ldots, n$; we have

$$\mathcal{F}_{i-2} \subset \mathcal{F}_{t-\delta} \subset \mathcal{S}_{t-\delta}.$$

Making use of Jensen and Minkowski's inequality in the same manner as in Equation (91), it follows

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

From the independence between ϵ_t and $\mathbf{X}_t$, for all $T_{i-1} \leq t \leq T_i$, we have

$$\begin{aligned}\mathbb{E}\left[\epsilon_t^2|\mathcal{S}_{t-\delta}\right] &= \mathbb{E}\left[\epsilon_t^2|\mathbf{X}_t\right]\\ &= \mathbb{E}\left[\epsilon_t^2\right] = 1.\end{aligned} \tag{95}$$

Under conditions (**M.1**), (**M.2**), (**E.1**)(i), and (**C.1**) and Equations (61) and (95), we obtain

$$\begin{aligned}&\frac{4}{\delta^2}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\right)^2|\mathcal{F}_{i-2}\right]dt\\ &= \frac{4}{\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\left(\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_i;\eta_{k,\ell})\right)^2|\mathcal{S}_{t-\delta}\right]|\mathcal{F}_{i-2}\right]\\ &\leq \frac{4n\left(C_m^2+1\right)}{\delta^2 C_f}\int_{\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})|^2\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}f_t^{\mathcal{F}_{i-2}}(\mathbf{x})dt}{f(\mathbf{x})}\right)\nu(d\mathbf{x}).\end{aligned}$$

For all $T_{i-1} \leq t \leq T_i$, using the stationarity of the process $(\mathbf{X}_t)_{t\geq 0}$, we have

$$f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) = f_{T_{i-1}}^{\mathcal{F}_{i-2}}(\mathbf{x}).$$

We deduce

$$\begin{aligned}
&\frac{4}{\delta^2}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\right)^2|\mathcal{F}_{i-2}\right]dt\\
&\leq \frac{4\delta n\left(C_m^2+1\right)}{\delta^2 C_f}\int_{\mathbf{S}}\left|\psi_k(\mathbf{x};\eta_{k,\ell})\right|^2\left(\frac{\frac{1}{n}\sum_{i=1}^{n}f_{T_{i-1}}^{\mathcal{F}_{i-2}}(\mathbf{x})}{f(\mathbf{x})}\right)\nu(d\mathbf{x})\\
&= \frac{T\left(C_m^2+1\right)}{\delta^2 C_f}(1+o(1))\int_{\mathbf{S}}\left|\psi_k(\mathbf{x};\eta_{k,\ell})\right|^2\nu(d\mathbf{x})\\
&\leq C_5 T,
\end{aligned} \tag{96}$$

where

$$C_5 = \frac{\left(C_m^2+1\right)}{\delta^2 C_f}(1+o(1)).$$

It follows

$$\Psi_{k,\ell}^{(2)} = CT^{1/2}. \tag{97}$$

Combining Equations (59), (64), and (66), we obtain

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

combining this with Equation (89), we conclude

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right) = O\left(\left(\frac{\delta^4}{T^4}\right)\left(\frac{T}{\ln T}\right)^2\right)+O\left(\left(\frac{\delta^4}{T^4}\right)T^2\right). \tag{98}$$

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

$$\mathbb{E}\left(\left|\widehat{\theta}_{k,\ell}-\theta_{k,\ell}\right|^4\right) \leq C\left(\frac{1}{T}\right)^2 \leq C\left(\frac{\ln T}{T}\right)^2. \tag{99}$$

The proof is achieved. □

Proof of Lemma 8. Consider decomposition (86) 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

$$\begin{aligned}
B_{k,\ell,1} &= \frac{1}{n}\sum_{i=1}^{n}\Psi_{i,k,\ell}\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})-\mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_i)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}\right]\right)dt\right),
\end{aligned} \tag{100}$$

$$B_{k,\ell,2} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})|\mathcal{F}_{t-\delta}\right]-\theta_{k,\ell}\right)dt. \tag{101}$$

Statement (88) 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

$$\Psi_{i,k,\ell} = V_{i,k,\ell}+W_{i,k,\ell}, \tag{102}$$

where

$$V_{i,k,\ell} = \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{x}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t} - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t}|\mathcal{F}_{t-\delta}\right]\right)dt,$$

$$W_{i,k,\ell} = \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}^c_{\mathcal{A}_t} - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}^c_{\mathcal{A}_t}|\mathcal{F}_{t-\delta}\right]\right)dt,$$

and

$$\mathbb{1}_{\mathcal{A}_t} = \left\{|\epsilon_t| \geq c_*\sqrt{\ln T}\right\},$$

and c_* represents a constant that will be selected later. Remark that

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{\theta}_{k,\ell} - \breve{\theta}_{k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{\frac{\ln T}{T}}\right) &\leq \mathbb{P}\left(|B_{k,\ell,1}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln T}{T}}\right) + o(1)\\ &= \mathbb{P}\left(\left|\sum_{i=1}^{n}\Psi_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{T\ln T}\right) + o(1)\\ &= I_1 + I_2 + o(1),\end{aligned} \tag{103}$$

where

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

First, we aim to bound the term I_1 of Equation (103). The Markov inequality and the Cauchy–Schwarz inequality yield

$$\begin{aligned}I_1 &\leq \frac{2}{\kappa\sqrt{T\ln T}}\mathbb{E}\left(\left|\sum_{i=1}^{n}V_{i,k,\ell}\right|\right)\\ &\leq \frac{2}{\kappa\sqrt{T\ln T}}\sum_{i=1}^{n}\mathbb{E}\left(\left|V_{i,k,\ell}\right|\right).\end{aligned} \tag{104}$$

Observe that

$$\begin{aligned}&\mathbb{E}\left(\left|V_{i,k,\ell}\right|\right)\\ &\leq \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\mathbb{E}\left(\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t} - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t}|\mathcal{F}_{t-\delta}\right]\right|\right)dt\\ &\leq \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left(\mathbb{E}\left(\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t}\right|\right) + \mathbb{E}\left(\mathbb{E}\left[\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t}\right||\mathcal{F}_{t-\delta}\right]\right)\right)dt\\ &\leq \frac{2}{\delta}\int_{T_{i-1}}^{T_i}\mathbb{E}\left(\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}_{\mathcal{A}_t}\right|\right)dt\\ &\leq \frac{2}{\delta}\left(\mathbb{E}\left(\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\right|^2\right)\right)^{1/2}(\mathbb{P}(\mathcal{A}_t))^{1/2}.\end{aligned} \tag{105}$$

Using Equation (92) combined with an elementary Gaussian inequality and taking c_* to have

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

We obtain

$$\begin{aligned} I_1 &\leq \frac{2C}{\delta\kappa}\sqrt{\frac{T}{\ln T}}\exp\left\{-\frac{c_*^2\ln T}{4}\right\} \\ &\leq \frac{2C}{\delta\kappa}\frac{T^{-\left(\frac{c_*^2}{4}-1/2\right)}}{\sqrt{\ln T}} \\ &\leq C_\kappa\left(\frac{1}{T^2}\right), \end{aligned} \tag{106}$$

where $C_\kappa = \frac{2C}{\delta\kappa}$. Now, we are going to look into determining an upper bound for I_2 from decomposition (103). We begin by confirming Lemma's 2 conditions. Assume that conditions (**M.1**) and (**M.2**) are fulfilled in connection with Equation (93), we infer

$$\begin{aligned} |Y_t\mathbb{1}_{\mathcal{A}_t^c}| &\leq C_m + c_*\sqrt{\ln T} \\ &\leq C\sqrt{\ln T}, \end{aligned} \tag{107}$$

which implies

$$\begin{aligned} |W_{i,k,\ell}| &\leq \frac{1}{\delta}\int_{T_{i-1}}^{T_i}\left|\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}^c_{\mathcal{A}_t} - \mathbb{E}\left[\frac{\rho(Y_t)}{f(\mathbf{X}_t)}\psi_k(\mathbf{X}_t;\eta_{k,\ell})\mathbb{1}^c_{\mathcal{A}_t}|\mathcal{F}_{t-\delta}\right]\right|dt \\ &\leq \frac{2C\sqrt{\ln T}}{\delta C}\sup_{\mathbf{x}\in\mathbf{S}}|\psi_k(\mathbf{x};\eta_{k,\ell})| \\ &\leq \frac{2C}{\delta C_f}\sqrt{\ln T}\sqrt{\frac{T}{(\ln T)^2}} \\ &\leq C_3\sqrt{\frac{T}{\ln T}} \\ &\leq C_3\sqrt{T}, \end{aligned} \tag{108}$$

where $C_3 = \frac{2C}{\delta C_f}$, let $B = C_3\sqrt{T}$, then, for all $\epsilon_T = \frac{\kappa}{2}\sqrt{T\ln T}$ with a sufficiently large T, we have

$$\begin{aligned} I_2 = \mathbb{P}\left(\left|\sum_{i=1}^{n} W_{i,k,\ell}\right| \geq \frac{\kappa}{2}\sqrt{T\ln T}\right) &\leq 2\exp\left\{-\frac{\epsilon_T^2}{2TB^2}\right\} \\ &= 2\exp\left\{-\frac{\left(\frac{\kappa}{2}\sqrt{T\ln T}\right)^2}{2T\left(C_3\sqrt{T}\right)^2}\right\} \\ &= 2\exp\left\{-\frac{\kappa^2\ln T}{4C_3^2T}\right\} \\ &= 2\exp\left\{\ln T^{-\frac{\kappa^2}{4C_3^2T}}\right\} \\ &= 2T^{-w(\kappa,T)}, \end{aligned} \tag{109}$$

where

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

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

$$I_2 \leq C\left(\frac{1}{T^2}\right). \tag{110}$$

It follows from Equations (103), (106), and (110) that

$$\mathbb{P}\left(\left|\widehat{\theta}_{k,\ell}-\widetilde{\theta}_{k,\ell}\right|\geq\frac{\kappa}{2}\sqrt{\frac{\ln T}{T}}\right) \leq C\frac{1}{T^2}+o(1) \tag{111}$$

$$\leq C\left(\frac{\ln T}{T}\right)^2. \tag{112}$$

The proof of Equation (74) is achieved. □

8. Besov Spaces

In terms of wavelet coefficients, in [23], the Besov space was described as follows: $0<s<r$ is a real-valued smoothness parameter of $\varrho\in L_p(\mathbb{R}^d)$, then $\varrho\in\mathcal{B}^s_{p,q}(\mathbb{R}^d)$ equivalent to

(B.1) $J_{s,p,q}(\varrho)=\|P_{V_0}\varrho\|_{L_p}+\left(\sum_{j>0}\left(2^{js}\|P_{W_j}\varrho\|_{L_p}\right)^q\right)^{1/q}<\infty$,

(B.2) $J'_{s,p,q}(\varrho)=\|a_0\cdot\|_{l_p}+\left(\sum_{j>0}\left(2^{j(s+d(1/2-1/p))}\|b_j\cdot\|_{l_p}\right)^q\right)^{1/q}<\infty$

with

$$\|a_0\cdot\|_{l_p}=\left(\sum_{\mathbf{k}\in\mathbb{Z}^d}|a_{0,\mathbf{k}}|^p\right)^{1/p}$$

and

$$\|b_j\cdot\|_{l_p}=\left(\sum_{i=1}^{N}\sum_{\mathbf{k}\in\mathbb{Z}^d}|b_{i,j,\mathbf{k}}|^p\right)^{1/p}$$

and classical sup-norm modification for $q=\infty$. The Besov spaces are useful for describing the smoothness properties of functional estimation and approximation theory. They contain well-known statistical research spaces, such as the Hilbert–Sobolev space $\left(H^s_2=\mathcal{B}^s_{2,2}(\mathbb{R}^d)\right)$, Holder space $C^s=\mathcal{B}^s_{\infty,\infty}(\mathbb{R}^d)$ for $0<s\notin\mathbb{N}$, and others. We refer readers to [58] for additional descriptions of Besov space and its merits in approximation theory and statistics. Reasoning as in [102], suppose that $1\leq p\leq\infty, 1\leq q\leq\infty$. Let

$$(S_{\boldsymbol{\tau}}f)(\mathbf{x})=f(\mathbf{x}-\boldsymbol{\tau}).$$

For $0<s<1$, set

$$\gamma_{s,p,q}(f)=\left(\int_{\mathbb{R}^d}\left(\frac{\|S_{\boldsymbol{\tau}}f-f\|_{L_p}}{\|\boldsymbol{\tau}\|^s}\right)^q\frac{d\boldsymbol{\tau}}{\|\tau\|^d}\right)^{1/q}$$

$$\gamma_{s,p,\infty}=\sup_{\boldsymbol{\tau}\in\mathbb{R}^d}\frac{\|S_{\boldsymbol{\tau}}f-f\|_{L_p}}{\|\boldsymbol{\tau}\|^S}$$

For $s=1$, set

$$\gamma_{1,p,q}(f)=\left(\int_{\mathbb{R}^d}\left(\frac{\|S_{\tau}f+S_{-\boldsymbol{\tau}}f-2f\|_{L_p}}{\|\boldsymbol{\tau}\|}\right)^q\frac{d\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|^d}\right)^{1/q},$$

$$\gamma_{1,p,\infty}=\sup_{\boldsymbol{\tau}\in\mathbb{R}^d}\frac{\|S_{\boldsymbol{\tau}}f+S_{-\boldsymbol{\tau}}f-2f\|_{L_p}}{\|\boldsymbol{\tau}\|}.$$

For $0<s<1$ and $1\leq p,q\leq\infty$, define

$$\mathcal{B}^s_{p,q}(\mathbb{R}^d)=\{f\in L_p:\gamma_{s,p,q}<\infty\}.$$

For $s > 1$, put $s = [s]^- + \{s\}^+$ with $[s]^-$ an integer and $0 < \{s\}^+ \leq 1$. Define $\mathcal{B}^s_{p,q}(\mathbb{R}^d)$ to be the space of functions in $L_p\left(\mathbb{R}^d\right)$ such that $D^j f \in \mathcal{B}_{\{s\}^+,p,q}$ for all $|j| \leq [s]^-$. The norm is defined by

$$\|f\|_{\mathcal{B}^s_{p,q}(\mathbb{R}^d)} = \|f\|_{L_p} + \sum_{|j| \leq s\}^-} \gamma_{\{s\}^+,p,q}\left(D^j f\right).$$

Author Contributions: S.D. and S.B.: conceptualization, methodology, investigation, writing—original draft, and writing—review and Editing. 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 very grateful to the three referees whose comments helped in improving the quality of this work.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Bosq, D. Linear processes in function spaces . In *Lecture Notes in Statistics*; Springer: New York, NY, USA, 2000; Volume 149, pp. xiv+283 . doi: 10.1007/978-1-4612-1154-9. [CrossRef]
2. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2005; pp. xx+426.
3. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Springer Series in Statistics; Springer: New York, NY, USA, 2006; pp. xx+258.
4. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer Series in Statistics; Springer: New York, NY, USA, 2012; pp. xiv+422. [CrossRef]
5. Zhang, J.T. Analysis of variance for functional data. In *Monographs on Statistics and Applied Probability*; CRC Press: Boca Raton, FL, USA, 2014; Volume 127, pp. xxiv+386.
6. Shi, J.Q.; Choi, T. *Gaussian Process Regression Analysis for Functional Data*; CRC Press: Boca Raton, FL, USA, 2011; pp. xx+196.
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.L. 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.G. 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*; 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, G.; Cao, R.; Fraiman, R.; Genest, C.; Vieu, P. Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.* **2019**, *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 [Editorial]. *J. Multivariate Anal.* **2016**, *146*, 1–6. [CrossRef]
17. Almanjahie, I.M.; Bouzebda, S.; Chikr Elmezouar, Z.; Laksaci, A. The functional kNN 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**, *5*, 431–533. [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-N-N Local Linear Approach. *Mathematics* **2022**, *10*, 902. [CrossRef]
23. Meyer, Y. *Wavelets and Operators*; Cambridge University Press: Cambridge, UK, 1993; pp. 35–58. [CrossRef]
24. Daubechies, I. Ten lectures on wavelets. In *CBMS-NSF Regional Conference Series in Applied Mathematics*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 1992; Volume 61, pp. xx+357. [CrossRef]
25. Mallat, S. *A Wavelet Tour of Signal Processing*, 3rd ed.; Elsevier/Academic Press: Amsterdam, The Netherlands, 2009; pp. xxii+805.
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; pp. xiv+382. [CrossRef]
27. Härdle, W.; Kerkyacharian, G.; Picard, D.; Tsybakov, A. Wavelets, approximation, and statistical applications. In *Lecture Notes in Statistics*; Springer: New York, NY, USA, 1998; Volume 129, pp. xviii+265. [CrossRef]
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.; Prakasa Rao, B.L.S. 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. In *Asymptotics in Statistics and Probability, Papers in Honor of George Gregory Roussas*; De Gruyter: Berlin, Germany, 2018; pp. 321–330. [CrossRef]
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. Allaoui, S.; Bouzebda, S.; Liu, J. Multivariate wavelet estimators for weakly dependent processes: strong consistency rate. *Comm. Statist. Theory Methods* **2022**, 1–34. [CrossRef]
34. Chesneau, C.; Kachour, M.; Maillot, B. Nonparametric estimation for functional data by wavelet thresholding. *REVSTAT* **2013**, *11*, 211–230.
35. Laib, N.; Louani, D. Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *J. Multivar. Anal.* **2010**, *101*, 2266–2281. [CrossRef]
36. Bouzebda, S.; Didi, S.; El Hajj, L. Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: Asymptotic results. *Math. Methods Statist.* **2015**, *24*, 163–199. [CrossRef]
37. Bouzebda, S.; Didi, S. Additive regression model for stationary and ergodic continuous time processes. *Comm. Statist. Theory Methods* **2017**, *46*, 2454–2493. [CrossRef]
38. 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]
39. 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]
40. 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]
41. 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]
42. 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]
43. Leucht, A.; Neumann, M.H. Degenerate U- and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.* **2013**, *65*, 349–386. [CrossRef]
44. Aït-Sahalia, Y. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **2002**, *70*, 223–262. [CrossRef]
45. Jiang, G.J.; Knight, J.L. A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econom. Theory* **1997**, *13*, 615–645. [CrossRef]
46. Kutoyants, Y.A. Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **1998**, *1*, 131–155. [CrossRef]
47. Banon, G. Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **1978**, *16*, 380–395. [CrossRef]
48. Castellana, J.V.; Leadbetter, M.R. On smoothed probability density estimation for stationary processes. *Stoch. Process. Appl.* **1986**, *21*, 179–193. [CrossRef]
49. Didi, S.; Alharby, A.; Bouzebda, S. Wavelet Density and Regression Estimators for Functional Stationary and Ergodic Data: Discrete Time. *Mathematics* **2022**, *10*, 3433. [CrossRef]
50. Goh, S.S. Wavelet bases for Hilbert spaces of functions. *Complex Var. Elliptic Equ.* **2007**, *52*, 245–260. [CrossRef]
51. Kerkyacharian, G.; Picard, D. Density estimation in Besov spaces. *Statist. Probab. Lett.* **1992**, *13*, 15–24. [CrossRef]
52. Kerkyacharian, G.; Picard, D. Thresholding algorithms, maxisets and well-concentrated bases. *Test* **2000**, *9*, 283–344. [CrossRef]

53. Prakasa Rao, B.L.S. Nonparametric density estimation for functional data via wavelets. *Commun. Stat. Theory Methods* **2010**, *39*, 1608–1618. [CrossRef]
54. 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]
55. DeVore, R.A. Nonlinear approximation. In *Acta Numerica, 1998*; Cambridge University Press: Cambridge, UK, 1998; Volume 7, pp. 51–150. [CrossRef]
56. Autin, F. Point de vue Maxiset en Estimation non Paramétrique. Ph.D. Thesis , Université Paris, Paris, France, 2004.
57. 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]
58. Donoho, D.L.; Johnstone, I.M.; Kerkyacharian, G.; Picard, D. Density estimation by wavelet thresholding. *Ann. Statist.* **1996**, *24*, 508–539. [CrossRef]
59. Ferraty, F.; Vieu, P. Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.* **2004**, *16*, 111–125. The International Conference on Recent Trends and Directions in Nonparametric Statistics, [CrossRef]
60. Peškir, G. The uniform mean-square ergodic theorem for wide sense stationary processes. *Stoch. Anal. Appl.* **1998**, *16*, 697–720. [CrossRef]
61. Delecroix, M. Sur l'Estimation et la Prévision non Paramétrique des Processus Ergodiques Cycles économiques et Taches Solaires. Ph.D. Thesis, Université Lille, Lille, France, 1987.
62. Didi, S.; Louani, D. Asymptotic results for the regression function estimate on continuous time stationary and ergodic data. *Stat. Risk Model.* **2014**, *31*, 129–150. [CrossRef]
63. Rosenblatt, M. Uniform ergodicity and strong mixing. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1972**, *24*, 79–84. [CrossRef]
64. Bradley, R.C. *Introduction to Strong Mixing Conditions*; Kendrick Press: Heber City, UT, USA, 2007; Volume 3, pp. xii+597.
65. Neumann, M.H. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **2011**, *17*, 1268–1284. [CrossRef]
66. Didi, S. Quelques propriétés Asymptotiques en Estimation non Paramétrique de Fonctionnelles de Processus Stationnaires en Temps Continu. Ph.D. Thesis, Université Pierre et Marie Curie, Paris, France , 2014.
67. Beran, J. Statistics for Long-Memory Processes. In *Monographs on Statistics and Applied Probability*; Chapman and Hall: New York, NY, USA, 1994; Volume 61, pp. x+315.
68. Lu, Z. Analysis of Stationary and Non-Stationary Long Memory Processes: Estimation, Applications and Forecast. 2009. Available online: https://tel.archives-ouvertes.fr/tel-00422376/document (accessed on 10 October 2022).
69. Maslowski, B.; Pospíšil, J. Ergodicity and parameter estimates for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.* **2008**, *57*, 401–429. [CrossRef]
70. Masry, E. Probability density estimation from sampled data. *IEEE Trans. Inform. Theory* **1983**, *29*, 696–709. [CrossRef]
71. Prakasa Rao, B.L.S. Nonparametric density estimation for stochastic processes from sampled data. *Publ. Inst. Statist. Univ. Paris* **1990**, *35*, 51–83.
72. Prakasa Rao, B.L.S. Statistical inference for diffusion type processes. In *Kendall's Library of Statistics*; Edward Arnold: London, UK; Oxford University Press: New York, NY, USA, 1999; Volume 8, pp. xvi+349.
73. Bosq, D. *Nonparametric Statistics for Stochastic Processes*, 2nd ed.; Springer: New York, NY, USA, 1998; Volume 110, pp. xvi+210. [CrossRef]
74. Blanke, D.; Pumo, B. Optimal sampling for density estimation in continuous time. *J. Time Ser. Anal.* **2003**, *24*, 1–23. [CrossRef]
75. Comte, F.; Merlevède, F. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.* **2002**, *6*, 211–238. [CrossRef]
76. Newey, W.K.; Powell, J.L. Asymmetric least squares estimation and testing. *Econometrica* **1987**, *55*, 819–847. [CrossRef]
77. Mohammedi, M.; Bouzebda, S.; Laksaci, A. On the nonparametric estimation of the functional expectile regression. *C. R. Math. Acad. Sci. Paris* **2020**, *358*, 267–272. [CrossRef]
78. Mohammedi, M.; Bouzebda, S.; Laksaci, A. The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data. *J. Multivar. Anal.* **2021**, *181*, 104673. [CrossRef]
79. Litimein, O.; Laksaci, A.; Mechab, B.; Bouzebda, S. Local linear estimate of the functional expectile regression. *Statist. Probab. Lett.* **2023**, *192*, 109682. [CrossRef]
80. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Statist.* **1964**, *35*, 73–101. [CrossRef]
81. Shannon, C.E. A mathematical theory of communication. *Bell System Tech. J.* **1948**, *27*, 379–423. 623–656. [CrossRef]
82. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley Series in Telecommunications; John Wiley & Sons, Inc.: New York, NY, USA, 1991; pp. xxiv+542. [CrossRef]
83. Rényi, A. On the dimension and entropy of probability distributions. *Acta Math. Acad. Sci. Hungar.* **1959**, *10*, 193–215. [CrossRef]
84. Kullback, S. *Information Theory and Statistics*; John Wiley & Sons, Inc.: New York, NY, USA; Chapman & Hall, Ltd.: London, UK, 1959; pp. xvii+395.
85. Gokhale, D.V.; Kullback, S. The information in contingency tables. In *Statistics: Textbooks and Monographs*; Marcel Dekker, Inc.: New York, NY, USA, 1978; Volume 23, pp. x+365.
86. Csiszár, I. Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen. *Magy. Tud. Akad. Mat. Kutató Int. Közl.* **1962**, *7*, 137–158.

87. Vasicek, O. A test for normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B* **1976**, *38*, 54–59. [CrossRef]
88. Dudewicz, E.J.; van der Meulen, E.C. Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.* **1981**, *76*, 967–974. [CrossRef]
89. Ebrahimi, N.; Habibullah, M.; Soofi, E. Testing exponentiality based on Kullback-Leibler information. *J. Roy. Statist. Soc. Ser. B* **1992**, *54*, 739–748. [CrossRef]
90. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
91. Verdugo Lazo, A.C.G.; Rathie, P.N. On the entropy of continuous probability distributions. *IEEE Trans. Inform. Theory* **1978**, *24*, 120–122. [CrossRef]
92. Bouzebda, S.; Elhattab, I. A strong consistency of a nonparametric estimate of entropy under random censorship. *C. R. Math. Acad. Sci. Paris* **2009**, *347*, 821–826. [CrossRef]
93. Bouzebda, S.; Elhattab, I. Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy. *C. R. Math. Acad. Sci. Paris* **2010**, *348*, 317–321. [CrossRef]
94. Bouzebda, S.; Elhattab, I.; Nemouchi, B. On the uniform-in-bandwidth consistency of the general conditional *U*-statistics based on the copula representation. *J. Nonparametr. Stat.* **2021**, *33*, 321–358. [CrossRef]
95. Bouzebda, S.; El-hadjali, T. Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.* **2020**, *32*, 864–914. [CrossRef]
96. Lad, F.; Sanfilippo, G.; Agrò, G. Extropy: Complementary dual of entropy. *Statist. Sci.* **2015**, *30*, 40–58. [CrossRef]
97. Ferraty, F.; Vieu, P. The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **2002**, *17*, 545–564. [CrossRef]
98. 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]
99. Hall, P.; Penev, S. Cross-validation for choosing resolution level for nonlinear wavelet curve estimators. *Bernoulli* **2001**, *7*, 317–341. [CrossRef]
100. Burkholder, D.L. Distribution function inequalities for martingales. *Ann. Probab.* **1973**, *1*, 19–42. [CrossRef]
101. de la Peña, V.H.; Giné, E. *Decoupling*; Probability and its Applications (New York); Springer: New York, NY, USA, 1999; pp. xvi+392. [CrossRef]
102. Masry, E. Wavelet-based estimation of multivariate regression functions in Besov spaces. *J. Nonparametr. Statist.* **2000**, *12*, 283–308. [CrossRef]