



Research article

Single index regression for locally stationary functional time series

Breix Michael Agua^{1,2} and Salim Bouzebda^{1,*}

¹ Université de technologie de Compiègne, Laboratory of Applied Mathematics of Compiègne (LMAC), CS 60319 - 57 avenue de Landshut, Compiègne, France

² Department of Mathematics, Caraga State University, Butuan City, Philippines

* **Correspondence:** Email: salim.bouzebda@utc.fr.

Abstract: In this research, we formulated an asymptotic theory for single index regression applied to locally stationary functional time series. Our approach involved introducing estimators featuring a regression function that exhibited smooth temporal changes. We rigorously established the uniform convergence rates for kernel estimators, specifically the Nadaraya-Watson (NW) estimator for the regression function. Additionally, we provided a central limit theorem for the NW estimator. Finally, the theory was supported by a comprehensive simulation study to investigate the finite-sample performance of our proposed method.

Keywords: convergence rates; exponential inequality; kernel regression; functional time series; locally stationary process; single index model; functional data analysis

Mathematics Subject Classification: 60F05, 62F40, 60G15, 60K05, 60K15

1. Introduction

In recent decades, there has been a burgeoning interest in statistical issues related to the analysis of functional random variables, particularly those taking values in infinite-dimensional spaces. This surge is driven by the increasing availability of data collected on ever more refined temporal and spatial grids, common in fields such as meteorology, medicine, satellite imagery, and various other research domains. Consequently, the statistical modeling of these data, conceptualized as random functions, has engendered numerous challenging theoretical and computational research questions. Several monographs can be consulted for a comprehensive understanding of both theoretical and practical aspects of functional data analysis. Specifically, the researchers in [14] examined linear models for random variables within a Hilbert space, while the researchers in [75] provide insights into scalar-on-function and function-on-function linear models, functional principal component analysis, and parametric discriminant analysis. For those interested in nonparametric methods, particularly

kernel-type estimation for scalar-on-function nonlinear regression models, [43] offers an extensive exploration, extending these methodologies to classification and discrimination analysis. Additionally, the researchers in [55] discuss extending several pivotal statistical concepts—such as goodness-of-fit tests, portmanteau tests, and change detection—to the functional data framework. The researchers in [85] focus on the analysis of variance for functional data, and the researcher in [77] delves into regression analysis for Gaussian processes. The literature also encompasses semiparametric models, including but not limited to projection pursuit models [31], partial linear models [3], and functional sliced inverse regression [45]. The authors of [40] investigated functional expectile regression as a framework for modeling spatial financial risk, proposing a nonparametric estimator tailored to the FSIR structure. In [15], the researchers addressed the intricate task of estimating the regression function operator and its partial derivatives for stationary mixing random processes using local higher-order polynomial fitting, achieving a key result in establishing the joint asymptotic normality of the estimators. Moreover, [26] focused on the weak convergence of the conditional empirical process indexed by a suitable function class and the k -NN conditional U -processes in the context of functional explanatory variables, that is extended in [18]. A notable contribution of this work was the establishment of sharp, almost uniform consistency in the number of neighbors for the proposed estimator. Finally, estimators for the single-index conditional U -statistics operator, designed to accommodate the nonstationary nature of the data-generating process, were analyzed in the time series framework in [19] and extended to spatial data in [17], building on the foundational work of [27]. For more recent insights and surveys on functional data modeling and analysis, readers can refer to [2, 4, 23, 25, 28–30, 33, 47, 67].

Literature strongly advocates for regression models that incorporate dimension reduction techniques. Single-index models, widely employed for this purpose, assume that the influence of predictors on the response can be simplified to a single index. This index, representing a projection in a specified direction, is combined with a nonparametric link function. By doing so, these models reduce the predictors to a single-variable index while retaining crucial characteristics. Notably, the nonparametric link function operates solely on a one-dimensional index, mitigating issues associated with high dimensionality, often referred to as the curse of dimensionality. The single-index model extends the concept of linear regression by incorporating a link function equivalent to the identity function (see [12, 50, 54, 59, 80]). Advances in functional data analysis underscore the need for models addressing dimensionality effects (see [33, 48, 61] for recent surveys, and [2, 24] for related studies). Semiparametric approaches naturally emerge as suitable candidates for such models. In this context, [44] and [1] investigated the functional single-index model (FSIM). The researchers in [56] introduced functional single-index composite quantile regression, estimating the unknown slope function and link function using B -spline basis functions. The researchers in [70] proposed a compact functional single-index model with a coefficient function that is nonzero only in a subregion. The researchers in [86] focused on estimating a general functional single-index model, where the conditional distribution of the response depends on the functional predictor through a functional single-index structure. The researchers in [81] developed a new estimation procedure that combines functional principal component analysis of the functional predictors, B -spline modeling for parameters, and profile estimation of unknown parameters and functions in the model.

Additionally, [62, 63] investigated the estimation of the functional single-index regression model with missing responses at random for strongly mixing time series data. The researchers

in [42] introduced a functional single-index varying coefficient model, with the functional predictor forming the single-index part. Utilizing functional principal component analysis and basis function approximation, they obtained estimators for the slope function and coefficient functions, proposing an iterative estimation procedure. [71] developed an automatic and location-adaptive procedure for estimating regression in an FSIM based on k -Nearest Neighbors (k NN) ideas. Motivated by the analysis of imaging data, [58] proposed a novel functional varying-coefficient single-index model for regression analysis of functional response data on a set of covariates of interest. The researchers in [5] and [7] investigated a functional Hilbertian regressor for nonparametric estimation of the conditional cumulative distribution with a scalar response variable in the single-index structure. In particular, the authors of the last reference tackled the challenge of nonparametric estimation for the regression function within the FSIM under a random censoring framework for the i.i.d. data. In an alternative approach, [31] extended their methodology to the multi-index case, avoiding anchoring the true parameter to a prespecified sieve. They provided a detailed theoretical analysis of a direct kernel-based estimation scheme, establishing a polynomial convergence rate. For references on the subject, we refer to [6, 28, 30, 68, 74].

The common assumption of stationarity in time series modeling has led to the development of numerous models, techniques, and methodologies. However, this assumption is often inappropriate for spatio-temporal data, even after applying detrending and deseasonalization methods. Many key time series models exhibit nonstationarity, as observed in various physical phenomena and economic data, rendering traditional stationary approaches inadequate. To address this issue, [79] introduced the concept of the locally stationary random process, which approximates a nonstationary process by treating it as stationary over short time intervals. This notion of local stationarity has been further explored in the works of [35, 36, 69, 72, 76], among others. Notably, the seminal work by [35] provides a strong theoretical foundation for inference on locally stationary processes.

1.1. Paper contribution

In [64], the author examined the asymptotic properties of nonparametric regression for dependent functional data, focusing on stationary processes. Building on this foundation, our work extends the framework to accommodate nonstationary processes, leveraging more advanced techniques in a more realistic setting. Specifically, we introduce the concept of local stationarity to model time-dependent functional data in the single index setting. While [64] established the convergence rate for a Nadaraya-Watson-type estimator, we broaden the scope by deriving this rate not only for that estimator but also for a wider class of kernel estimators. Both studies consider semi-metric spaces for strongly mixing data; however, the approaches differ significantly: The researcher in [64] employed the norm, whereas we utilize the inner product. This distinction enhances the adaptability of our results, making them applicable to a broader range of scenarios, particularly in the single-index setting. Additionally, [57] explored nonstationary functional time series within a semi-metric space defined by a norm for mixing data, obtaining similar results. However, our use of an inner-product-based semi-metric generalizes their findings, particularly as [57] did not address the single-index direction θ . Our work further advances the field by providing detailed proofs of asymptotic normality, expanding upon the methodology in [64], and reinforcing the theoretical foundations of our approach. Our primary aim of this paper is to establish a comprehensive framework for the single-index model in a nonparametric setting, with a focus on regression involving functional covariates and the challenges posed by the

potential nonstationarity of functional time series. We conduct a rigorous theoretical analysis to address the complexities of this setting, including the unbounded nature of the functional space, which necessitates intricate and extensive proofs. Finally, the theoretical results are supported by a simulation study, demonstrating the finite-sample performance of our proposed method and underscoring its practical relevance and robustness.

1.2. Paper organization

The structure of the paper is as follows: In Section 2, we introduce the concept of local stationarity for functional time series, which take values in a semi-metric space equipped with the semi-metric $d_\theta(\cdot, \cdot)$, where θ is a single index from a Hilbert space \mathcal{H} . The novelty of this work lies in the incorporation of a single-index $\theta \in \mathcal{H}$, which serves as a filter and effectively represents the explanatory variables $X_{t,T}$ that influence the response variable $Y_{t,T}$. This section also covers the dependence structure of the functional time series considered in the study.

In Section 3, we present the derivation of uniform convergence rates for general kernel estimators, along with results on uniform convergence and asymptotic normality for the Nadaraya-Watson estimator of the regression function. In particular, we observe that the general kernel estimator converges uniformly to its mean at the rate $\sqrt{\frac{\log T}{Th\phi_\theta(h)}}$. For the Nadaraya-Watson estimator, we establish a convergence rate with two distinct components: One addressing the stochastic part and the other addressing the bias part. Following a decomposition similar to [64] and [57], we obtain our convergence result based on the proof structure of Theorem 3.1. However, we incorporate the impact of the single-index θ , reflected in the dependence structure in the second part of the proof of Proposition 3.1, which is essential for proving Theorem 3.1. The small-ball probability $\phi_\theta(h)$ provides insight into the concentration of random variables as governed by the semi-metric $d_\theta(\cdot, \cdot)$, highlighting the contribution of this work. This section concludes with the derivation of asymptotic normality for the Nadaraya-Watson estimator. The proof begins by showing that the bias term converges to zero, followed by establishing the asymptotic normality of the variance term. The argument is completed using Bernstein's blocking technique, along with key tools such as Davydov's Lemma, the Volkonski-Rosanolv inequality, and the Lindeberg-Feller theorem for finite normality, supplemented by appropriate truncation arguments. In Section 4, we present comprehensive simulation results to evaluate the finite-sample properties of the proposed approach. Concluding remarks and discussions on potential future research are provided in Section 5. For clarity and coherence, all proofs are collected in Appendix-A, with relevant technical results included in the Appendix-B.

1.3. Notations

Let $\{a_n\}$ and $\{b_n\}$ be arbitrary sequences of positive numbers. Throughout this paper, we adopt the notation $a_n \lesssim b_n$ to signify that there exists a constant $C > 0$, independent of n , such that $a_n \leq Cb_n$ for all n . When $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously, we write $a_n \sim b_n$, indicating that the two sequences are asymptotically comparable. In cases where $\frac{a_n}{b_n} \rightarrow 0$ as $n \rightarrow \infty$, we use the notation $a_n \ll b_n$. For any real numbers a and b , the expressions $a \vee b$ and $a \wedge b$ denote the maximum and minimum of a and b , respectively, i.e., $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Furthermore, we denote convergence in distribution by \xrightarrow{d} . Finally, for any real number x , $\lfloor x \rfloor$ represents the integer part

of x , also known as the floor function.

2. Settings

In this section, we present an advanced framework for analyzing locally stationary functional time series, incorporating the concept of a semi-metric $d_\theta(\cdot, \cdot)$ that depends on a single-index parameter $\theta \in \mathcal{H}$, where \mathcal{H} denotes a Hilbert space. This approach broadens the original notion of local stationarity as introduced by [35] while extending the developments on locally stationary functional series by [17, 19, 27, 57]. In this expanded framework, we examine not only the structural properties of local stationarity but also delve into the dependence structures inherent in the functional time series, providing a comprehensive view of the interplay between stationarity and dependence within this context.

2.1. Model

The semi-metric $d_\theta(\cdot, \cdot)$ associated to the single-index $\theta \in \mathcal{H}$, a Hilbert space, defined by:

$$d_\theta(u, v) := |\langle \theta, u - v \rangle|, \quad \forall u, v \in \mathcal{H}.$$

Let $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$ be random variables where $Y_{t,T}$ is real-valued and $X_{t,T}$ takes values in some semi-metric space \mathcal{H} with a semi-metric $d_\theta(\cdot, \cdot)$. In this study, we consider the following model:

$$Y_{t,T} = m\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) + \sigma\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) \varepsilon_t, \quad t = 1, \dots, T, \quad (2.1)$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed random variables that is independent of $\{X_{t,T}\}_{t=1}^T$, $m(\cdot, \cdot)$ is the regression function allowed to change smoothly over time and $\sigma(\cdot, \cdot)$ is the variance function. For notational convenience, we use $\varepsilon_{t,T}$ to denote $\sigma\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) \varepsilon_t$. For the identifiability issue, we assume that the regression function is differentiable and $\langle \theta, e_1 \rangle = 1$, where e_1 is the first element of the orthonormal basis of \mathcal{H} . Observe that $m\left(\frac{t}{T}, \langle \theta_1, x \rangle\right) = m\left(\frac{t}{T}, \langle \theta_2, x \rangle\right)$ implies that $\theta_1 \equiv \theta_2$. We also assume that $\{X_{t,T}\}$ is a locally stationary functional time series, and the regression function m is allowed to change smoothly over time.

2.2. Local stationarity

A functional time series $\{X_{t,T}\}_{t=1}^T$, where $T \rightarrow \infty$, is intuitively regarded as locally stationary if it exhibits approximate stationarity within localized time intervals. This concept implies that while the series may display non-stationary characteristics over its entirety, within any sufficiently small time window, its behavior can be approximated as stationary. For an in-depth discussion on the theoretical framework and broader applications of locally stationary time series, we refer to the works of [37] and [38]. Furthermore, investigations into the concept of local stationarity for time series in a Hilbert space setting are available in [82] and [8]. Local stationarity at each normalized time point u can be characterized by stochastically approximating the original process $\{X_{t,T}\}$ with a stationary functional time series $\{X_t^{(u)}\}$. The following formal definition captures this idea rigorously.

Definition 2.1. [82] The \mathcal{H} -valued stochastic process $\{X_{t,T}\}_{t=1}^T$ is locally stationary if for each rescaled time point $u \in [0, 1]$, there exists an associated \mathcal{H} -valued process $\{X_t^{(u)}\}_{t \in \mathbb{Z}}$ with the following properties:

(i) $\{X_t^{(u)}\}_{t \in \mathbb{Z}}$ is strictly stationary.

(ii) It holds that

$$d_\theta(X_{t,T}, X_t^{(u)}) \leq \left(\left| \frac{t}{T} - u \right| + \frac{1}{T} \right) U_{t,T}^{(u)} \text{ a.s.}, \quad (2.2)$$

for all $1 \leq t \leq T$, where $\{U_{t,T}^{(u)}\}$ is a process of positive variables satisfying $\mathbb{E}[(U_{t,T}^{(u)})^\rho] < C$ for some $\rho > 0$ and $C < \infty$ that are independent of u, t , and T .

We extend the concept of local stationarity for real-valued time series introduced by [35] and for local functional time series studied by [57] by introducing a new semi-metric $d_\theta(\cdot, \cdot)$ associated with a single index θ in a Hilbert space \mathcal{H} (see Definition 2.1). This semi-metric $d_\theta(\cdot, \cdot)$ is defined similarly to that in [63], where $\theta \in \Theta_{\mathcal{H}}$ and $\Theta_{\mathcal{H}}$ is a compact subset of \mathcal{H} . In their work, they applied the semi-metric to data exhibiting arithmetic strong mixing with identical distributions. Furthermore, our Definition 2.1 corresponds to Definition 2.1 of [57], where they used the semi-metric $d(u, v) = \|u - v\|$ in a Banach or Hilbert space \mathcal{H} with norm $\|\cdot\|$. Moreover, when \mathcal{H} is the Hilbert space $L_{\mathbb{R}}^2([0, 1])$ of all real-valued, square-integrable functions on the unit interval $[0, 1]$, our definition aligns with Definition 2.1 of [82]. In their study, they used the L_2 -norm for $f, g \in L_{\mathbb{R}}^2([0, 1])$, defined as

$$\|f\|_2 = \sqrt{\langle f, f \rangle}, \quad \langle f, g \rangle = \int_0^1 f(t)g(t) dt.$$

Observe that if we choose $\theta \in L_{\mathbb{R}}^2([0, 1])$ defined by

$$\theta(t) = \frac{f(t)g(t)}{f(t) - g(t)},$$

assuming $f(t) - g(t) > 0$ almost everywhere, then

$$\begin{aligned} \langle \theta, f - g \rangle &= \int_0^1 \frac{f(t)g(t)}{f(t) - g(t)} [f(t) - g(t)] dt \\ &= \int_0^1 f(t)g(t) dt = \langle f, g \rangle. \end{aligned}$$

Therefore,

$$d_\theta(f, g) = |\langle \theta, f - g \rangle| = |\langle f, g \rangle|.$$

This demonstrates that, in this particular case, the semi-metric $d_\theta(\cdot, \cdot)$ reduces to the absolute value of the inner product of f and g , thereby generalizing the L_2 -norm.

Remark 2.1. [82] generalizes the definition of local stationary processes, initially proposed by [34], to the functional setting in the frequency domain. This extension is made under the following assumptions:

(A1) (i) $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ is a weakly stationary white noise process taking values in \mathcal{H} with a spectral representation given by

$$\varepsilon_j = \int_{-\pi}^{\pi} e^{i\omega j} dZ_\omega,$$

where Z_ω is a 2π -periodic orthogonal increment process taking values in $\mathcal{H}_{\mathbb{C}} = L_{\mathbb{C}}^2([0, 1])$;

(ii) the functional process $X_{i,n}$ with $i = 1, \dots, n$ and $n \in \mathbb{N}$ is given by

$$X_{j,n} = \int_{-\pi}^{\pi} e^{i\omega j} \mathcal{A}_{j,\omega}^{(n)} dZ_{\omega} \quad \text{a.e. in } \mathcal{H},$$

with the transfer operator $\mathcal{A}_{j,\omega}^{(n)} \in \mathcal{B}_p$ and an orthogonal increment process Z_{ω} .

(A2) There exists $\mathcal{A} : [0, 1] \times [-\pi, \pi] \rightarrow S_p(\mathcal{H}_{\mathbb{C}})$ with $\mathcal{A}_{u,\cdot} \in \mathcal{B}_p$ and $\mathcal{A}_{u,\omega}$ being continuous in u such that for all $n \in \mathbb{N}$

$$\sup_{\omega, t} \|\mathcal{A}_{t,\omega}^{(n)} - \mathcal{A}_{t,\omega}^{(n)}\|_p = O\left(\frac{1}{n}\right).$$

They have proved in [82, Proposition 2.2] that:

Proposition 2.1. *Suppose that Assumptions (A1) and (A2) hold. Then, $\{X_{i,n}\}$ is a locally stationary process in \mathcal{H} .*

2.3. Mixing condition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let \mathcal{A} and \mathcal{B} be subfields of \mathcal{F} . Define

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|.$$

Moreover, for an array $\{Z_{t,T} : 1 \leq t \leq T\}$, define the coefficients

$$\alpha(k) = \sup_{t, T: 1 \leq t \leq T-k} \alpha(\sigma(Z_{s,T} : 1 \leq s \leq t), \sigma(Z_{s,T} : t+k \leq s \leq T)),$$

where $\sigma(Z)$ is the σ -field generated by Z . The array $\{Z_{t,T}\}$ is said to be α -mixing (or strongly mixing) if $\alpha(k) \rightarrow 0$ as $k \rightarrow \infty$.

Among the various mixing conditions explored in the literature, α -mixing is a relatively weak but widely applicable property, satisfied by a broad range of stochastic processes, including numerous time series models. The researchers in [49] and [84] established the conditions required for a linear process to exhibit α -mixing. Under minimal assumptions, the linear autoregressive and more general bilinear time series models demonstrate strong mixing properties with exponentially decaying mixing coefficients. Furthermore, the researchers in [9] provided valuable insights into the role of α -mixing (including geometric ergodicity) in identifying nonlinear time series models (see also [20–22]). Similarly, the researchers in [32] showed that functional autoregressive processes attain geometric ergodicity under specific conditions. Additionally, [65, 66] demonstrated that, with mild assumptions, both autoregressive conditional heteroscedastic processes and nonlinear additive autoregressive models with exogenous variables are stationary and α -mixing.

3. Main results

We provide the main results of this paper in this section by first considering general kernel estimators as represented in Section 3.1. Correspondingly, we derived the uniform convergence rate as seen in Proposition 3.1. Furthermore, some results on uniform convergence rate and asymptotic normality of the Nadaraya-Watson estimator for the regression function in Model (2.1) are provided in the succeeding subsections (see Theorems 3.1 and 3.2, respectively).

3.1. Kernel estimation for regression functions

As mentioned, we consider the following kernel estimator for $m(u, \langle \theta, x \rangle) = m_\theta(u, x)$ in Model (2.1):

$$\hat{m}_\theta(u, x) = \frac{\sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) Y_{t,T}}{\sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T}))}, \quad (3.1)$$

where $K_1(\cdot)$ and $K_2(\cdot)$ denote one-dimensional kernel functions. Here, for $j = 1, 2$, we used the notations $K_{j,h}(v) = K_j(v/h)$. Moreover, $h = h_T$ is a bandwidth satisfying $h \rightarrow 0$ as $T \rightarrow \infty$. The estimator defined in (3.1) differs from the traditional NW estimator, typically used in strictly stationary settings, by incorporating an additional kernel along the time dimension. Consequently, smoothing is applied not only in the direction of the covariates $X_{t,T}$ but also across time, accounting for variations in the regression function over time.

Remark 3.1. In the finite-dimensional framework, the researcher in [83] provided a detailed study of the following model:

$$Y_{t,T} = m\left(\frac{t}{T}, X_{t,T}\right) + \varepsilon_{t,T}, \quad \text{for } t = 1, \dots, T, \quad (3.2)$$

where $Y_{t,T}$ and $X_{t,T}$ are random variables of dimensions 1 and d , respectively, and the noise satisfies $\mathbb{E}[\varepsilon_{t,T} | X_{t,T}] = 0$. The NW estimator for the regression function in model (3.2) is expressed as:

$$\hat{m}(u, x) = \frac{\sum_{t=1}^T K_h(u - t/T) \prod_{j=1}^d K_h(x^j - X_{t,T}^j) Y_{t,T}}{\sum_{t=1}^T K_h(u - t/T) \prod_{j=1}^d K_h(x^j - X_{t,T}^j)}.$$

Furthermore, the researchers in [83] explored structured models where the regression function decomposes into time-varying additive components. In [83], it is demonstrated that a locally stationary sequence can be effectively approximated by decomposing it into a stationary time series component and a time-varying trend function. Specifically, this representation is expressed as:

$$Y_{t,T} = Y_t^* + \vartheta_1\left(\frac{t}{T}\right), \quad X_{t,T} = X_t^* + \vartheta_2\left(\frac{t}{T}\right),$$

where $\vartheta_1\left(\frac{t}{T}\right)$, $\vartheta_2\left(\frac{t}{T}\right)$, are unknown time-varying functions and (X_t^*, Y_t^*) are assumed to be strictly stationary. In contrast, the present paper focuses on a different framework: The functional single-index model. This setting introduces new challenges and requires distinct conditions to derive the asymptotic properties.

We first enumerate the model and kernel assumptions which are important in deriving our main results. Mentioned assumptions are made for Model (2.1) and the kernel functions therein. These assumptions are standard, and similar assumptions are made among others by [57, 63, 64, 82].

Assumption 3.1. (Model assumptions)

- (M1) The process $\{X_{t,T}\}$ is locally stationary, that is, $\{X_{t,T}\}$ satisfies Definition 2.1.
- (M2) Let $B_\theta(x, h) = \{y \in \mathcal{H} : d_\theta(x, y) \leq h\}$ denote the ball of radius h centered in $x \in \mathcal{H}$. We assume that there exist positive constants c_d , and C_d where $c_d < C_d$, such that for all $u \in [0, 1]$, all $x \in \mathcal{H}$, and all $h > 0$,

$$0 < c_d \phi_\theta(h) f_1(x) \leq \mathbb{P}(X_t^{(u)} \in B_\theta(x, h)) =: F_u(h; x, \theta) \leq C_d \phi_\theta(h) f_1(x),$$

where $\phi_{\theta,x}(h) = \mathbb{P}(X_t^{(u)} \in B_\theta(x, h)) > 0$, $\phi_\theta(h) \rightarrow 0$ as $h \rightarrow 0$, and $f_1(x)$ is a nonnegative functional in $x \in \mathcal{H}$. Moreover, there exist constants $C_\phi > 0$ and $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$,

$$\int_0^\varepsilon \phi_\theta(u) du > C_\phi \varepsilon \phi_\theta(\varepsilon). \quad (3.3)$$

- (M3) $\sup_{s,t,T} \sup_{s \neq t} \mathbb{P}((X_{s,T}, X_{t,T}) \in B_\theta(x, h) \times B_\theta(x, h)) \leq \psi_\theta(h) f_2(x)$, where $\psi_\theta(h) \rightarrow 0$ as $h \rightarrow 0$, and $f_2(x)$ is a nonnegative functional in $x \in \mathcal{H}$. We assume that the ratio $\psi_\theta(h)/\phi_\theta^2(h)$ is bounded.
- (M4) $m_\theta(u, x)$ is twice continuously partially differentiable with respect to u . We also assume that

$$\sup_{u \in [0,1]} |m_\theta(u_1, x) - m_\theta(u_2, y)| \leq c_m (d_\theta(x, y)^\beta + |u_1 - u_2|^\beta),$$

for all $x, y \in \mathcal{H}$ for some $c_m > 0$ and $\beta > 0$.

Assumption 3.1 formalizes the local stationarity property of the process $\{X_{t,T}\}$ as seen in condition (M1), while the distributional behavior of the rescaled random variable $X_t^{(u)}$ is described in the second condition (M2). Also, condition (M2) ensures that through the function $\phi_\theta(h)$, the behavior of the small-ball probability is controlled around zero. Furthermore, condition (M2) is consistent with the ones made by [64] and [46]. The former used this assumption for strongly mixing processes while the latter in the context of density estimation for functional data. For strongly mixing processes in the locally functional time series setting, one may see [57].

Observe that similar to what [64] employed in their paper, we can satisfy the Condition (3.3) for fractal-type processes by defining $\phi_\theta(\varepsilon) \sim \varepsilon^\tau$ as $\varepsilon \rightarrow 0$ for some $\tau > 0$, with some change of notation but of similar meaning. Considering a separable Hilbert space \mathcal{H} , it can be expected that as $h \rightarrow 0$, $\phi_\theta(h)$ diminishes to 0 as in [46]. Since $\phi_\theta(h)$ is defined similarly as $\phi(h)$ by previously mentioned authors and more specifically by [64], we exemplify its form by $\phi_\theta(\varepsilon) = \varepsilon^\delta \exp(-C/\varepsilon^a)$ with $\delta, a \geq 0$. The case when $\phi_\theta(\varepsilon) = \exp(-C/\varepsilon^2)$ refers to the Ornstein-Uhlenbeck and general diffusion processes. Also, $\phi_\theta(\varepsilon) = \varepsilon^\delta \exp(-C)$ with $\delta > 0$ corresponds to the fractal processes. Excellent references detailing fractal-type processes and on concepts relating to the small ball probability $F_u(h; x)$ include those of [13] and [43]. Extending condition (M2) gives us $(d_\theta(X_{s,T}, x), d_\theta(X_{t,T}, x))$ that describes the behavior of the joint distribution near the origin. Thus, condition (M3). Consistent with the assumptions made by [43, 57, 64], we have included (M3) and (M4) in this study. To deal with the regression function $m_\theta(u, x)$ with respect to u and x , we have condition (M4) to handle the smoothness and continuity properties, respectively.

We impose the following conditions on σ (Assumption 3.2) and for kernel functions (Assumption 3.3). These conditions are similar to those made and assumed by [19, 43, 57, 63, 64, 83].

Assumption 3.2. (Conditions on σ)

($\Sigma 1$) $\sigma : [0, 1] \times \mathcal{H} \rightarrow \mathbb{R}$ is bounded by some constant $C_\sigma < \infty$ from above and by some constant $c_\sigma > 0$ from below, that is, $0 < c_\sigma \leq \sigma(u, \langle \theta, x \rangle) \leq C_\sigma < \infty$ for all u and x .

($\Sigma 2$) σ is Lipschitz continuous with respect to u .

($\Sigma 3$) $\sup_{u \in [0, 1]} \sup_{y: d_\theta(x, y) \leq h} |\sigma(u, \langle \theta, x \rangle) - \sigma(u, \langle \theta, y \rangle)| = o(1)$ as $h \rightarrow 0$.

Assumption 3.3. (Kernel assumptions)

(KB1) The kernel $K_1(\cdot)$ is symmetric around zero, bounded, and has a compact support, that is, $K_1(v) = 0$ for all $|v| > C_1$ for some $C_1 < \infty$. Moreover, $\int K_1(z) dz = 1$ and $K_1(\cdot)$ is Lipschitz continuous, that is, $|K_1(v_1) - K_1(v_2)| \leq C_2 |v_1 - v_2|$ for some $C_2 < \infty$ and all $v_1, v_2 \in \mathbb{R}$.

(KB2) The kernel $K_2(\cdot)$ is nonnegative, bounded, and has support in $[0, 1]$ such that $0 < K_2(0)$ and $K_2(1) = 0$. Moreover, $K_2'(v) = dK_2(v)/dv$ exists on $[0, 1]$ and satisfies $C_1' \leq K_2'(v) \leq C_2'$ for two real constants $-\infty < C_1' < C_2' < 0$. Moreover, suppose that $K_2(\cdot)$ is a Lipschitz continuous function.

In line with the assumptions made by [57, 83], we define conditions ($\Sigma 1$) and ($\Sigma 2$). Along with the notational convention $\sigma(u, \langle \theta, x \rangle)$, we employ condition ($\Sigma 3$) to investigate the asymptotic properties of the variance of $\hat{m}(u, x)$ and establish its asymptotic normality. Evaluating $Th\phi_\theta(h) \text{Var}(\hat{g}_\theta^{(1)}(u, x))$ in Eq (A.9) is essential to demonstrate the existence of the variance $V_\theta(u, x) > 0$.

Moreover, the assumptions on the kernel functions $K_1(\cdot)$ and $K_2(\cdot)$ are standard in the literature and are satisfied by popular kernels such as the (asymmetric) triangle and quadratic kernels. Condition (KB1) ensures that $K_1(\cdot)$ is bounded and has compact support. Additionally, its symmetry property implies that $K_1(\cdot)$ can be any symmetric kernel, such as the box, triangle, quadratic, or Gaussian kernel. Condition (KB2) indicates that $K_2(\cdot)$ is a Type II kernel function as defined in [43]. For a comprehensive introduction and detailed discussion of these assumptions, please refer to [43]. In our study, Assumption 3.3 aids in determining the upper bounds of the kernel functions, as demonstrated in the initial steps of the proof of Proposition 3.1 and in certain parts of the proof of Theorem 3.2.

3.2. Uniform convergence rates for kernel estimators

Our goal here is to establish the asymptotic properties of the estimator given in (3.1). We first investigate the mentioned properties for the general kernel estimator provided below. For an array of one-dimensional random variables $\{W_{i,T}\}$, we define

$$\hat{\psi}_\theta(u, x) = \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{i,T})) W_{i,T}. \quad (3.4)$$

Several representations of (3.4) can be made using kernel estimators including the Nadaraya-Watson (NW) estimator. For the purposes of this study, we use the results with $W_{i,T} = 1$ and $W_{i,T} = \varepsilon_{i,T}$.

To show the desired claim, we will derive the uniform convergence rate of $\hat{\psi}_\theta(u, x) - \mathbb{E}[\hat{\psi}_\theta(u, x)]$. We assume the following (Assumptions 3.4 and 3.5) for the components of $\hat{\psi}_\theta(u, x)$ in Eq (3.4). Assumption 3.4 relates to the mixing assumptions attributed to the array of random variables $\{X_{i,T}, W_{i,T}\}$ while Assumption 3.5 concerns the regularity conditions on h and $\phi_\theta(h)$.

Assumption 3.4. (Mixing assumptions)

(E1) It holds that $\sup_{t,T} \sup_{x \in \mathcal{H}} \mathbb{E}[|W_{t,T}|^\zeta | X_{t,T} = x] \leq C$ for some $\zeta > 2$ and $C < \infty$.

(E2) The α -mixing coefficients of the array $\{X_{t,T}, W_{t,T}\}$ satisfy $\alpha(k) \leq Ak^{-\gamma}$ for some $A > 0$ and $\gamma > 3$. We also assume that $\delta + 1 < \gamma(1 - \frac{2}{\nu})$ for some $\nu > 2$ and $\delta > 1 - \frac{2}{\nu}$, and

$$h^{2(1 \wedge \beta) - 1} \left(\phi_\theta(h) \lambda_T + \sum_{k=\lambda_T}^{\infty} k^\delta (\alpha(k))^{1 - \frac{2}{\nu}} \right) \rightarrow 0, \quad (3.5)$$

as $T \rightarrow \infty$, where $\lambda_T = [(\phi_\theta(h))^{-(1 - \frac{2}{\nu})/\delta}]$.

To prove Theorem 3.2, which establishes the asymptotic normality of $\hat{m}_\theta(u, x)$, we need to apply condition (3.5) specified in (E2) of Assumption 3.4. This condition also aids in demonstrating the asymptotic negligibility of the bias of $\hat{m}_\theta(u, x)$. Similar assumptions for conditions (E1) and (E2) were made by [57]; however, we employ a slightly different version of condition (E2) in our work. Comparable conditions to Assumption 3.4 have also been utilized in other studies, such as [64, 83].

In the following section, we present the regularity conditions concerning h and $\phi_\theta(h)$, specified in Assumption 3.5.

Assumption 3.5. (Regularity assumptions)

As $T \rightarrow \infty$,

(R1) $\frac{(\log T)^{\frac{\gamma-1}{2} + \zeta_0(\gamma+1)}}{T^{\frac{\gamma-1}{2} - 1 - \frac{\gamma+1}{\zeta}} h^{\frac{\gamma-1}{2} + 1} (\phi_\theta(h))^{\frac{\gamma-1}{2}}} \rightarrow 0$ for some $\zeta_0 > 0$, and

(R2) $Th^3, Th\phi_\theta(h) \rightarrow \infty$,

where ζ and γ are positive constants that appear in Assumption 3.4.

To obtain the convergence rate of the general estimator $\hat{\psi}_\theta(u, x)$, we use an exponential inequality for α -mixing sequence in Lemma B.3, and impose the regularity assumption (R1). Consequently, the same is done for the Nadaraya-Watson estimator $\hat{m}_\theta(u, x)$. Considering the same values for γ and ζ in Assumption 3.4, condition (R1) holds since $\gamma > \frac{2 + 3\zeta}{\zeta - 2}$ and that the $\lim_{\zeta \rightarrow \infty} \frac{2 + 3\zeta}{\zeta - 2} = 3$. Moreover, we use the second regularity condition (R2) in dealing with the bias and in computing the convergence rate of the general estimator. Also, it can be noted that (R2) holds by considering $h \leq CT^{-\zeta}$ (see [83]), and $\phi_\theta(h) \sim h^\tau, \tau > 0$.

The succeeding result (Proposition 3.1) suggests a generalization of the work of [83] on the uniform convergence results to a functional time series. Comparably, we provide a more general result than those obtained by [63], wherein the convergence is obtained for identically distributed random variables, and extended the work of [57] that uses the measure $d(u, v) = \|u - v\|$. As a recall, we use the measure $d_\theta(u, v) = |\langle \theta, u - v \rangle|$.

Proposition 3.1. *Assume that Assumptions 3.1 (M1), (M2), 3.3, 3.4, and 3.5 are satisfied. Then, the following result holds for any $x \in \mathcal{H}$:*

$$\sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in [0,1]} |\hat{\psi}_\theta(u, x) - \mathbb{E}[\hat{\psi}_\theta(u, x)]| = O_{\mathbb{P}} \left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} \right).$$

Using Proposition 3.1, we established that our general estimator $\hat{\psi}_\theta(u, x)$ achieves a uniform convergence rate of $\sqrt{\frac{\log T}{Th\phi_\theta(h)}}$. This result is comparable to the convergence rates for general estimators in nonstationary functional data settings reported by [57] (see Proposition 3.1 therein). The key distinction in our approach is the use of $\phi_\theta(h)$ instead of $\phi(h)$, which enables us to incorporate the single index θ into the semi-metric $d_\theta(\cdot, \cdot)$ utilized in our study.

Moreover, in the context of strictly stationary functional time series, [43] derived a pointwise convergence rate for the nonparametric regression function that aligns with the rate we obtained in Proposition 3.1.

Building on these results, we will now determine the uniform convergence rate of the kernel estimator $\hat{m}_\theta(u, x)$ using Theorem 3.1 below.

Theorem 3.1. *Suppose that Assumptions 3.1–3.3, and 3.5 are satisfied and that Assumption 3.4 is satisfied with $W_{1,T} = 1$ and $W_{t,T} = \varepsilon_{t,T}$. Then, the following result holds for any $x \in \mathcal{H}$:*

$$\sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in [C_1 h, 1 - C_1 h]} |\hat{m}_\theta(u, x) - m_\theta(u, x)| = O_{\mathbb{P}} \left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} + h^{2 \wedge \beta} \right).$$

In our proof, the stochastic component is of order $O_{\mathbb{P}} \left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} \right)$, which relies heavily on the results obtained in Proposition 3.1. The bias component is shown to be of order $O_{\mathbb{P}} \left(h^{2 \wedge \beta} \right)$. Similar to the convergence results obtained by [57], and in contrast to [83] (see Theorem 4.2 therein), we do not encounter a bias term arising from the approximation error of $X_{t,T}$ by $X_t^{(u)}$. Under our assumptions, this approximation error is $O \left(T^{-1} h^{(1 \wedge \beta) - 1} \phi_\theta^{-1}(h) \right)$, which is negligible compared to $h^{2 \wedge \beta}$. Furthermore, by introducing a single-index θ in our framework, we generalized the pointwise convergence results from [43] and extended the findings for strictly stationary functional time series in [64]. This generalization is achieved through Theorem 3.1 presented above. To illustrate our result, refer to Remark 3.2 below. While [57] derived a convergence rate of order $O_{\mathbb{P}} \left(\left(\frac{\log T}{T} \right)^{\frac{\beta}{2\beta + \tau + 1}} \right)$ for fractal-type processes $\{X_t^{(u)}\}$ with $\beta \leq 2$, we have obtained a comparable convergence rate in our setting. According to [43] (see page 208), the bandwidth selection scheme corresponding to regression estimation is $h \sim \left(\frac{\log T}{T} \right)^{\frac{1}{2\beta + \tau}}$. This observation leads to the following remark.

Remark 3.2. For a fractal-type process $\{X_t^{(u)}\}$, the right-hand side of (3.1) with $\beta \leq 2$ is optimized by choosing $h \sim \left(\frac{\log T}{T} \right)^{\frac{1}{2\beta + \tau}}$, and the optimized rate is

$$\sup_{x \in \mathcal{S}} \sup_{\theta \in \Theta} \sup_{u \in [C_1 h, 1 - C_1 h]} |\hat{m}_\theta(u, x) - m_\theta(u, x)| = O_{\mathbb{P}} \left(\left(\frac{\log T}{T} \right)^{\frac{\beta}{2\beta + \tau}} \right).$$

3.3. Asymptotic normality for kernel estimators

We provide the central limit theorem for our Nadaraya-Watson estimator $\hat{m}_\theta(u, x)$ in this section. We need Assumption 3.6 to establish the asymptotic normality of the NW estimator $\hat{m}_\theta(u, x)$. We are

aided by the Bernstein's big-block and small-block procedure to accomplish the proof of Theorem 3.2 along with the assumptions considered therein. To simplify the proof, we set $K_2(\cdot)$ as the asymmetrical triangle kernel. That is, $K_2(x) = (1 - x)\mathbb{1}_{(x \in [0,1])}$.

Assumption 3.6. There exists a sequence of positive integers $\{v_T\}$ satisfying $v_T \rightarrow \infty$, $v_T = o(\sqrt{Th\phi_\theta(h)})$ and $\sqrt{\frac{T}{h\phi_\theta(h)}}\alpha(v_T) \rightarrow 0$ as $T \rightarrow \infty$.

Observe that

$$\begin{aligned}\hat{m}_\theta(u, x) - m_\theta(u, x) &= \frac{1}{\hat{m}_\theta^{(1)}(u, x)} \left(\hat{g}_\theta^{(1)}(u, x) + \hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x)\hat{m}_\theta^{(1)}(u, x) \right) \\ &= \frac{1}{\hat{m}_\theta^{(1)}(u, x)} \left(\hat{g}_\theta^{(1)}(u, x) + \hat{g}_\theta^B(u, x) \right),\end{aligned}$$

where

$$\begin{aligned}\hat{m}_\theta^{(1)}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})), \\ \hat{g}_\theta^{(1)}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})) \varepsilon_{t,T}, \\ \hat{g}_\theta^{(2)}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})) m_\theta \left(\frac{t}{T}, X_{t,T} \right).\end{aligned}$$

Under the same assumption in Theorem 3.1, we can show that

$$\text{Var}(\hat{g}_\theta^B(u, x)) = o\left(\frac{1}{Th\phi_\theta(h)}\right), \text{ and } \frac{1}{\hat{m}_\theta^{(1)}(u, x)} = O_{\mathbb{P}}(1).$$

See the proof of Theorem 3.2 for details. Then, we have

$$\hat{m}_\theta(u, x) - m_\theta(u, x) = \frac{\hat{g}_\theta^{(1)}(u, x)}{\hat{m}_\theta^{(1)}(u, x)} + B_{T,\theta}(u, x) + o_{\mathbb{P}}\left(\sqrt{\frac{1}{Th\phi_\theta(h)}}\right),$$

where $B_{T,\theta}(u, x) = \frac{\mathbb{E}[\hat{g}_\theta^B(u, x)]}{\mathbb{E}[\hat{m}_\theta^{(1)}(u, x)]}$ is the ‘‘bias’’ term and $\frac{\hat{g}_\theta^{(1)}(u, x)}{\hat{m}_\theta^{(1)}(u, x)}$ is the ‘‘variance’’ term.

Theorem 3.2. Assume that Assumptions 3.1–3.3, 3.5, and 3.6 are satisfied and that Assumption 3.4 is satisfied for both $W_{1,T} = 1$ and $W_{i,T} = \varepsilon_{i,T}$. Then as $T \rightarrow \infty$, the following result holds for any $x \in \mathcal{H}$:

$$\sqrt{Th\phi_\theta(h)}(\hat{m}_\theta(u, x) - m_\theta(u, x) - B_{T,\theta}(u, x)) \xrightarrow{d} N(0, V_\theta(u, x)),$$

where $B_{T,\theta}(u, x) = O(h^{2\wedge\beta})$ and

$$V_\theta(u, x) = \lim_{T \rightarrow \infty} Th\phi_\theta(h) \frac{\text{Var}(\hat{g}_\theta^{(1)}(u, x))}{\mathbb{E}[\hat{m}_\theta^{(1)}(u, x)]} > 0.$$

Remark 3.3. Theorem 3.2 is an extension of the results in [64, 83], and [57] to a locally stationary functional time series with a semi metric $d_\theta(\cdot, \cdot)$ associated with a single index θ from a Hilbert space \mathcal{H} . It is noteworthy that we use a decomposition containing the expressions $B_{T,\theta}(u, x)$ and $V_\theta(u, x)$ that are very similar to those in [57]. Moreover, we use a similar proof procedure to that of [64] and of [57] and obtain very similar results. It is important to note that the asymptotic negligibility of the bias part is achieved by requiring

$$Th^{1+2(2\wedge\beta)}\phi_\theta(h) \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$

This is satisfied whenever we have $h = T^{-\xi}$ and $\phi_\theta(h) = h^c$ for

$$0 < c < 1 - \frac{1}{\xi} \quad \text{and} \quad \frac{1}{(1+c) + 2(2 \wedge \beta)} < \xi < \frac{1}{1-c}.$$

Remark 3.4. The single functional index $\theta \in \mathcal{H}$ is typically unknown and must be estimated in practical applications. This challenge has been addressed in the literature on single functional regression models, where estimation approaches using cross-validation or maximum-likelihood methods have been explored, as in [1] and references therein. An alternative approach, adopted in this section, involves selecting $\theta(t)$ among the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}(X'))\langle X', \cdot \rangle_{\mathcal{H}}]$, where $X(t)$ is a diffusion-type process on a real interval $[a, b]$ and $X'(t)$ denotes its first derivative (see [5, 7, 40], for example). Given a training sample \mathcal{L} , the covariance operator can be estimated using its empirical version $(1/|\mathcal{L}|) \sum_{i \in \mathcal{L}} (X'_i - \mathbb{E}X')^t (X'_i - \mathbb{E}X')$. Subsequently, a discretized version of the eigenfunctions $\theta_i(t)$ can be obtained via principal component analysis.

3.4. The bandwidth selection criterion

Several methods have been established in the literature for bandwidth selection criteria. For instance, [78] lists various approaches for choosing the smoothing parameter h , which is crucially important in kernel density estimation for both univariate and multivariate data. In the univariate case, a natural method is the researcher's subjective choice, involving the plotting of multiple curves and selecting an estimate that aligns with prior beliefs about the density. Other methods involve referencing a particular standard distribution. Specifically, when using a Gaussian kernel, the optimal bandwidth is given by $h_{\text{opt}} = 1.06 \sigma n^{-1/5}$, where σ can be the sample standard deviation or a more robust estimator. To avoid oversmoothing in multimodal populations, one may use the interquartile range R and set $h_{\text{opt}} = 0.79 R n^{-1/5}$. Corresponding suggested optimal values of the smoothing parameter for multivariate data can also be found in Chapter 4 of [78]. Additionally, we refer to [51, 53, 73] for bandwidth selection rules concerning nonparametric kernel estimators, including the Nadaraya-Watson regression estimators. A suitable selection of the smoothing parameter h that works for both finite and infinite-dimensional cases must be chosen. We define our local cross-validation criterion as follows:

$$CV_{\theta,x}(h_T) = \frac{1}{T} \sum_{s=1}^T \left[Y_{s,T} - \hat{m}_\theta^{[s]} \left(\frac{s}{T}, X_{s,T} \right) \right]^2 \mathcal{W}(X_{s,T}), \quad (3.6)$$

where $\hat{m}_\theta^{[s]}(\cdot)$ represents the leave-one-out estimator of $\hat{m}_\theta(\cdot)$, based on the sample $(X_{1,T}, Y_{1,T}), \dots, (X_{T,T}, Y_{T,T})$ excluding the pair $(X_{s,T}, Y_{s,T})$. Our goal is to minimize the criterion (3.6) by selecting a bandwidth $\hat{h}_T \in [a_T, b_T]$ among $h \in [a_T, b_T]$.

Following the idea introduced by [10] and used by [16], we replace the global weights $\tilde{\mathcal{W}}(X_{s,T})$ with local weights $W(x, X_{s,T})$, which are independent of T . Thus, $CV_{\theta,x}(h_T)$ in (3.6) becomes

$$CV_{\theta,x}(h_T) = \frac{1}{T} \sum_{s=1}^T \left[Y_{s,T} - \hat{m}_{\theta}^{[s]} \left(\frac{s}{T}, X_{s,T} \right) \right]^2 W(x, X_{s,T}).$$

In practice, for $i \in \{1, \dots, T\}$, one may utilize uniform global weights $\tilde{\mathcal{W}}(X_{i,T}) = 1$ and define the local weights as

$$W(x, X_{i,T}) = \begin{cases} 1 & \text{if } d_{\theta}(x, X_{i,T}) \leq h, \\ 0 & \text{otherwise.} \end{cases}$$

4. Numerical experiment

In this section, we present the results of a numerical simulation study designed to illustrate the finite-sample behavior of the proposed estimator within the context of a single-index regression model for functional time series. The estimator for $m_{\theta}(u, x)$ is defined as

$$\hat{m}_{\theta}(u, x) = \frac{\sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_{\theta}(x, X_{t,T})) Y_{t,T}}{\sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_{\theta}(x, X_{t,T}))},$$

where $K_{1,h}(\cdot)$ and $K_{2,h}(\cdot)$ are kernel functions with bandwidth h , and $d_{\theta}(\cdot, \cdot)$ is the semi-metric associated with the single index θ . We consider the Hilbert space \mathcal{F} of square-integrable functions on $[0, 1]$:

$$\mathcal{F} = \left\{ f : [0, 1] \rightarrow \mathbb{R} \mid \int_0^1 f^2(t) dt < \infty \right\},$$

equipped with the inner product

$$\langle f, g \rangle = \int_0^1 f(t)g(t) dt,$$

for all $f, g \in \mathcal{F}$, and the associated L^2 -norm

$$\|f\| = \left(\int_0^1 f^2(t) dt \right)^{1/2}.$$

Let $\mathcal{L} : \mathcal{F} \rightarrow \mathcal{F}$ be a linear operator. An operator \mathcal{L} is said to be compact if it maps bounded sets to relatively compact sets. Since \mathcal{F} is a separable Hilbert space, any compact operator \mathcal{L} admits a singular value decomposition with a sequence of nonnegative singular values $\{s_n(\mathcal{L})\}_{n \in \mathbb{N}}$ decreasing to zero. The operator \mathcal{L} can thus be represented as

$$\mathcal{L}f = \sum_{n=1}^{\infty} s_n(\mathcal{L}) \langle f, \psi_n \rangle \phi_n, \quad \text{for all } f \in \mathcal{F},$$

where $\{\phi_n\}$ and $\{\psi_n\}$ are orthonormal sequences in \mathcal{F} . For $p \in [1, \infty]$, the Schatten p -class $S_p(\mathcal{F})$ consists of compact operators \mathcal{L} for which the Schatten p -norm $\|\mathcal{L}\|_p$ is finite:

$$\|\mathcal{L}\|_p = \begin{cases} \left(\sum_{n=1}^{\infty} s_n(\mathcal{L})^p \right)^{1/p}, & \text{if } 1 \leq p < \infty, \\ \sup_{\|x\| \leq 1} \|\mathcal{L}x\|, & \text{if } p = \infty. \end{cases} \quad (4.1)$$

Simulation steps:

Step 1: Generating the functional sample $X_{t,T}$. We generate the functional time series $X_{t,T}$ as a functional autoregressive process of order 1 (FAR(1)), following the method outlined in [82]:

$$X_{t,T}(\tau) = B_{t/T}(X_{t-1,T})(\tau) + \epsilon_t(\tau), \quad \tau \in [0, 1], \quad t = 1, \dots, T, \quad (4.2)$$

where $B_{t/T}$ is a time-varying linear operator, and $\epsilon_t(\tau)$ are innovation functions. The innovations ϵ_t are constructed using Fourier basis functions $\{\psi_j\}_{j \in \mathbb{N}}$, specifically sine and cosine functions, with coefficients $\langle \epsilon_t, \psi_j \rangle$ that are independent Gaussian random variables with mean zero and variance $\sigma_j^2 = [\pi(j - 1.5)]^{-2}$:

$$\epsilon_t = \sum_{j=1}^{\infty} \langle \epsilon_t, \psi_j \rangle \psi_j, \quad \text{with } \langle \epsilon_t, \psi_j \rangle \sim N(0, \sigma_j^2).$$

In practice, we truncate the infinite series at a finite number J to obtain an approximate representation:

$$X_{t,T} = \sum_{j=1}^J \langle X_{t,T}, \psi_j \rangle \psi_j.$$

Substituting into Eq (4.2) and exploiting the linearity of $B_{t/T}$, we derive the finite-dimensional recursion:

$$\mathbf{X}_t^{(T)} = \mathbf{B}_{t/T} \mathbf{X}_{t-1}^{(T)} + \boldsymbol{\epsilon}_t, \quad t = 1, \dots, T,$$

where $\mathbf{X}_t^{(T)} = (\langle X_{t,T}, \psi_1 \rangle, \dots, \langle X_{t,T}, \psi_J \rangle)^\top$, $\boldsymbol{\epsilon}_t = (\langle \epsilon_t, \psi_1 \rangle, \dots, \langle \epsilon_t, \psi_J \rangle)^\top$, and $\mathbf{B}_{t/T}$ is a $J \times J$ matrix with entries $b_{ij} = \langle B_{t/T} \psi_i, \psi_j \rangle$. To construct $\mathbf{B}_{t/T}$, we generate a $J \times J$ matrix \mathbf{A}_u with entries a_{ij} being independent Gaussian random variables with variance

$$\sigma_{ij}^2 = ui^{-2c} + (1 - u)e^{-i-j},$$

where $u = t/T$, $c = 3$, and $i, j = 1, \dots, J$. The operator $B_{t/T}$ is then represented by normalizing \mathbf{A}_u using the operator norm:

$$\mathbf{B}_{t/T} = \frac{\eta \mathbf{A}_u}{\|\mathbf{A}_u\|_\infty},$$

where $\eta = 0.4$. In this step, we have adopted the parameter values for η and c as proposed by [82]. The parameter c plays a crucial role in modulating the decay rate of the variance for the entries a_{ij} , thereby significantly influencing the overall structure of the functional data, as illustrated in Figure 1. Moreover, the parameter η primarily affects the scale of $X_{t,T}$, allowing us to adjust the influence of $B_{t,T}$

on the updates to $X_{t,T}$ at each time step. By tuning η , we can regulate the extent of each update to $X_{t,T}$, with smaller values of η serving to mitigate potential over-amplification in these updates. For our simulations, we have adopted these parameter values, not only to align with the cited work, but also to enhance computational efficiency. Readers interested in further analysis might consider examining how variations in η and c impact the dynamics of the model. Additionally, we utilize the following Fourier basis functions for the terms $X_{t,T}(\tau)$ and $\epsilon_t(\tau)$ for each $t = 1, \dots, T$ and $\tau \in [0, 1]$:

$$\psi_k(\tau) = \sqrt{2} \sin(\pi k \tau), \quad \text{for odd } k \leq J,$$

$$\psi_k(\tau) = \sqrt{2} \cos(\pi k \tau), \quad \text{for even } k \leq J.$$

These basis functions help define the structural form of each functional term, enhancing the representational fidelity of the model.

Step 2: Generating the single index $\theta(\tau)$. We generate the single index function $\theta(\tau)$ as a linear combination of basis functions, following [63]:

$$\theta(\tau) = \frac{1}{\sqrt{3}}\phi_1(\tau) + \frac{1}{\sqrt{3}}\phi_2(\tau) + \frac{1}{\sqrt{6}}\phi_3(\tau) + \frac{1}{\sqrt{6}}\phi_4(\tau), \quad \tau \in [0, 1],$$

where the basis functions are defined as

$$\phi_1(\tau) = \sqrt{2} \sin(\pi\tau),$$

$$\phi_2(\tau) = \sqrt{2} \cos(\pi\tau),$$

$$\phi_3(\tau) = \sqrt{2} \sin(3\pi\tau),$$

$$\phi_4(\tau) = \sqrt{2} \cos(3\pi\tau).$$

Step 3: Generating the scalar response variable $Y_{t,T}$. We generate the scalar response variable $Y_{t,T}$ according to the model

$$Y_{t,T} = m\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) + \epsilon_{t,T}, \quad t = 1, \dots, T,$$

where $\epsilon_{t,T}$ are independent standard normal random variables. The regression function $m_\theta(u, x)$ is specified as

$$m_\theta(u, x) = m(u, \langle \theta, x \rangle) = 2.5 \sin(2\pi u) \cdot \cos(\pi \langle \theta, x \rangle).$$

Step 4: Selecting the Bandwidth h . We select the bandwidth h using the cross-validation criterion, minimizing the cross-validation score

$$CV_{\theta,x}(h) = \frac{1}{T} \sum_{s=1}^T \left[Y_{s,T} - \hat{m}_\theta^{[s]} \left(\frac{s}{T}, X_{s,T} \right) \right]^2,$$

where $\hat{m}_\theta^{[s]}(\cdot)$ is the leave-one-out estimator of $\hat{m}_\theta(\cdot)$, computed without the s -th observation.

Visualization. Figure 1 displays samples of 100 curves $X_{t,T}$ over the interval $[0, 1]$ for different numbers of basis functions $J = 5, 15, 25,$ and 45 . Each panel illustrates realizations of the time-varying functional process $X_{t,T}$ for $t = 1, \dots, 100$.

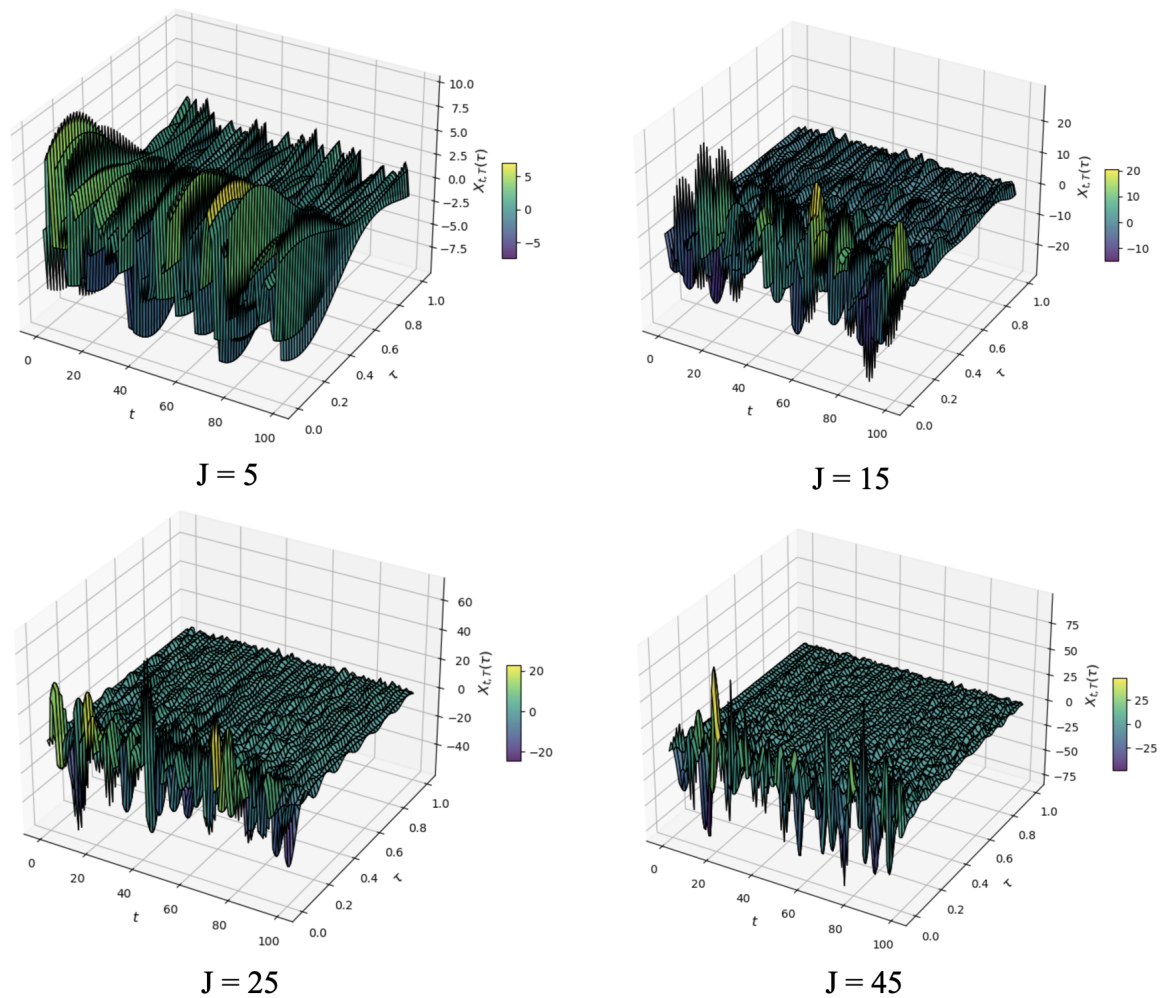


Figure 1. Realizations of the curves $X_{t,100}(\tau)$, $\tau \in (0, 1)$ and $t = 1, \dots, 100$ for $FAR(1)$.

4.1. Point-wise mean square convergence

To demonstrate the pointwise convergence of the estimator $\hat{m}_\theta(u, x)$ to the true regression function $m_\theta(u, x)$, we conducted simulations with $T = 1000$ observations and evaluated the estimator at various rescaled time points $u = 0.25, 0.50,$ and 0.95 . We selected $x = X_{3,T}$ as the covariate point of interest.

Recall that the Nadaraya-Watson estimator is defined as

$$\hat{m}_\theta(u, x) = \frac{\sum_{t=1}^T K_1\left(\frac{u - \frac{t}{T}}{h}\right) K_2\left(\frac{d_\theta(x, X_{t,T})}{h}\right) Y_{t,T}}{\sum_{t=1}^T K_1\left(\frac{u - \frac{t}{T}}{h}\right) K_2\left(\frac{d_\theta(x, X_{t,T})}{h}\right)},$$

where $h > 0$ is the bandwidth parameter, $d_\theta(\cdot, \cdot)$ is the semi-metric associated with the single index θ , and $K_1(\cdot)$ and $K_2(\cdot)$ are kernel functions. In our simulations, we employed the uniform kernel for $K_1(\cdot)$ and the Gaussian kernel for $K_2(\cdot)$. Specifically, the uniform kernel $K_1(\cdot)$ is defined as

$$K_1(w) = \begin{cases} 1, & \text{if } |w| \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

which provides equal weighting to observations within the bandwidth and zero weight outside. The Gaussian kernel $K_2(\cdot)$ is given by

$$K_2(w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right),$$

as described in [43]. Notably, we utilized an asymmetric version of $K_2(\cdot)$ by considering only non-negative values of w , reflecting the non-negativity of the semi-metric $d_\theta(x, X_{t,T})$. It is worth noting that while our simulation study employed the uniform and gaussian kernels for $K_1(\cdot)$ and $K_2(\cdot)$, respectively, one may also explore utilizing other kernel functions satisfying Assumption 3.3. A comparison with regards to the estimator's mean squared errors (MSEs) across various choices of kernel functions may be investigated. The MSE values computed from the simulations using the uniform and gaussian kernels are summarized in Table 1, illustrating the estimator's performance at the specified time points. These results confirm the theoretical convergence properties of $\hat{m}_\theta(u, x)$ as established in our asymptotic analysis.

Table 1. Mean squared error of the regression estimator $\hat{m}_\theta(u, x)$ where $X_{t,T}$ is simulated using the FAR(1) model across values of J and u (figures in parentheses indicate the computed best value of the bandwidth h using the cross-validation criterion).

	Number of basis functions J			
	$J = 5$	$J = 15$	$J = 25$	$J = 45$
$u = 0.25$	3.252051 (0.434172)	3.072263 (0.485943)	3.133377 (0.456048)	3.345938 (0.428430)
$u = 0.50$	0.003168 (0.426549)	0.003110 (0.427440)	0.003141 (0.462879)	0.003114 (0.416057)
$u = 0.95$	0.309052 (0.454069)	0.303020 (0.426549)	0.273662 (0.464067)	0.283947 (0.477826)

The analysis explores the impact of varying the number of basis functions, $J = 5, 15, 25$, and 45 , along with the various rescaled time points. For each of the results obtained (for both the MSEs and the optimal bandwidths), we set $T = 1000$, 100 replications, and 100 candidate optimal bandwidths to choose from, ranging between 0.01 to 0.99. The MSE values observed at each rescaled time u

display minimal variation across the different values of J , suggesting that the choice of basis functions within this range has negligible influence on the results. The table further suggests that for the rescaled time points $u = 0.25$, and $u = 0.95$, we obtain relatively higher MSEs compared to that at $u = 0.50$. These findings suggest that for the subsequent analysis of the estimator $\hat{m}_\theta(u, x)$'s performance as T increases, we focus on $u = 0.50$, using a bandwidth of $h = 0.426549$ and $J = 5$ basis functions. We have demonstrated the performance of the mentioned estimator at varying values of $T = 100, 500, 1000$, and 1500 . Although not displayed in Figure 2, we have computed the mean values of the MSEs which are $0.0260923, 0.00506145, 0.00264353$, and 0.00182452 for $T = 100, 500, 1000$, and 1500 , respectively. Hence, our assertion concerning decreasing approximation errors induced by our estimator as T increases. As is common in inferential contexts, larger sample sizes generally yield better performance. Simple inspection of the results in Figure 2 shows that larger sample sizes T lead to smaller MSEs. More precisely, Figure 2 shows a steady decrease in approximation error as T grows, supporting the inference that as T becomes larger and larger, the MSE values become smaller and smaller. These empirical findings are thus in alignment with the theoretical results outlined in Theorem 3.1.

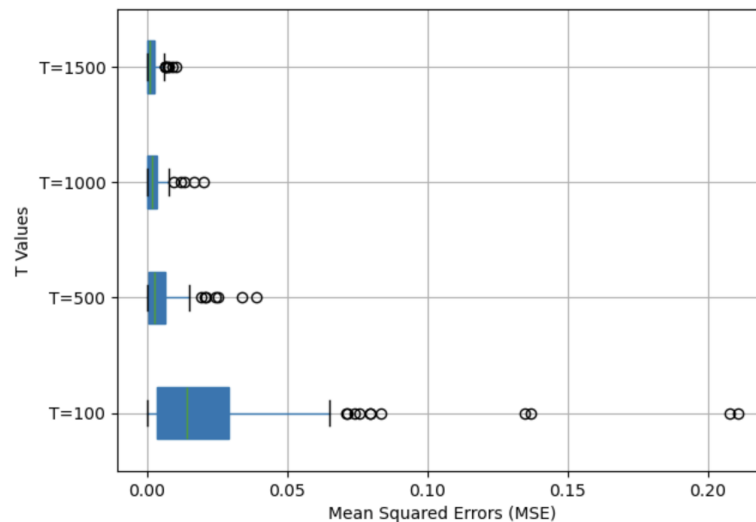


Figure 2. Mean squared error for the regression estimator $\hat{m}_\theta(u, x)$ where $X_{i,T}$ is simulated using the FAR(1) model.

4.2. Asymptotic normality

As imposed in the previous subsection, we also consider the rescaled time point $u = 0.50$, with a bandwidth of $h = 0.426549$, $T = 1500$, and $J = 5$ basis functions to illustrate the asymptotic normality of our estimator. In this setting, we employ a quadratic kernel for $K_1(\cdot)$ and an asymptotic uniform kernel for $K_2(\cdot)$. The uniform kernel is defined analogously to that introduced in the previous subsection, but with support on the interval $[0, 1]$. The quadratic kernel is defined as $\frac{3}{4}(1 - u^2)\mathbb{1}_{[-1,1]}(u)$ (see [43]). While selecting a bandwidth h tailored to these kernels would be preferable, for the purpose of our simulations—specifically, to illustrate asymptotic normality—we consider an approximate optimal bandwidth of $h = 0.426549$. Our results suggest that this choice yields similar conclusions, provided that the assumptions on $K_1(\cdot)$ and $K_2(\cdot)$ stated in Assumption 3.3 hold. Therefore, our primary

aim here is to demonstrate that the theoretical results remain valid irrespective of the particular choices of $K_1(\cdot)$ and $K_2(\cdot)$, as long as they satisfy Assumption 3.3 and the other required conditions. Thus, we proceed with the simulation by first recalling the following quantities $\hat{m}_\theta(u, x)$, $m_\theta(u, x)$, $B_{T,\theta}(u, x)$, and $\hat{g}_\theta^B(u, x)$ in Assumption 3.6 and suppose that

$$S_\theta(u, x) =: \sqrt{Th\phi_\theta(h)}(\hat{m}(u, x) - m_\theta(u, x) - B_{T,\theta}(u, x)),$$

where

$$B_{T,\theta}(u, x) = \frac{\mathbb{E}[\hat{g}_\theta^B(u, x)]}{\mathbb{E}[\hat{m}_\theta^{(1)}(u, x)]}.$$

Here, with $B_{T,\theta}(u, x) = O(h^{2\wedge\beta})$ as per results obtained, we will demonstrate that $S_\theta(u, x)$ can be approximated by the normal distribution. That is,

$$S_\theta(u, x) \xrightarrow{d} N(0, V_\theta(u, x)),$$

where

$$V_\theta(u, x) = \lim_{T \rightarrow \infty} Th\phi_\theta(h) \frac{\text{Var}(\hat{g}_\theta^{(1)}(u, x))}{\mathbb{E}[\hat{m}_\theta^{(1)}(u, x)]} > 0.$$

To estimate $\text{Var}(\hat{g}_\theta^{(1)}(u, x))$ and $\phi_\theta(h)$ by $\hat{\sigma}(u, x)$ and $\hat{\phi}_\theta(h)$, respectively, please see [63]. That is, we have

$$\begin{aligned} \hat{\phi}_\theta(h) &= \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{\{|\langle \theta, x - X_{t,T} \rangle| < h\}}(X_{t,T}), \\ \text{and } \hat{\sigma}(u, x) &= \frac{\sum_{t=1}^T Y_{t,T}^2 K_1\left(\frac{u - \frac{t}{T}}{h}\right) K_2\left(\frac{d_\theta(x, X_{t,T})}{h}\right)}{\sum_{t=1}^T K_1\left(\frac{u - \frac{t}{T}}{h}\right) K_2\left(\frac{d_\theta(x, X_{t,T})}{h}\right)} - (\hat{m}_\theta(u, x))^2. \end{aligned}$$

We confirm our theoretical results by taking 100 copies of the random variable $S_\theta(u, x)$ and creating the corresponding histogram and Q - Q plot. See Figure 3 for an illustration. The histogram illustrates that our data can be approximated by a normal distribution (the distribution we theorized) with its shape forming like a bell. The presented Q - Q plot supports our claim since the points closely align with the diagonal line. This indicates that our simulated statistics approximate a normal distribution.

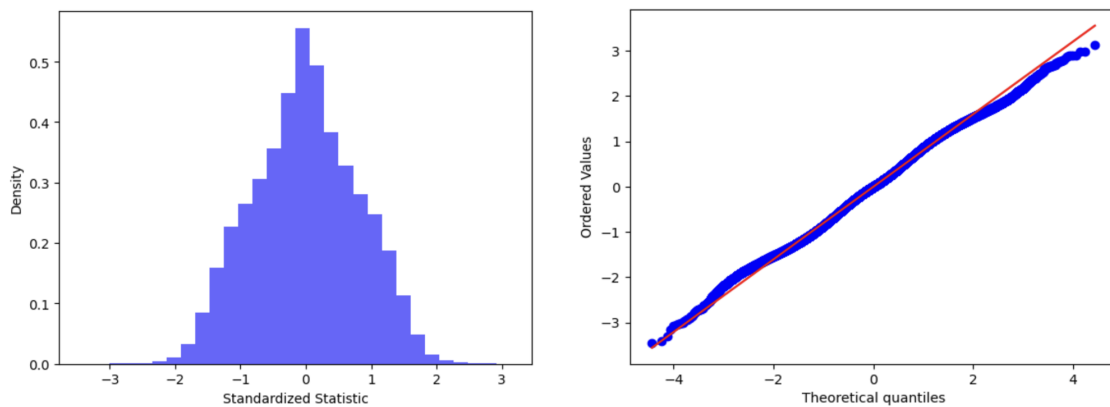


Figure 3. Histogram and Q - Q plot of $S_\theta(u, x)$ when the explanatory variable is the FAR(1) model.

5. Concluding remarks

In this paper, we develop an asymptotic theory for single-index nonparametric regression models applied to locally stationary functional time series under α -mixing dependent observations. We begin by defining a semi-metric $d_\theta(\cdot, \cdot)$ associated with a single index θ in a Hilbert space \mathcal{H} , where $d_\theta(u, v) := |\langle \theta, u - v \rangle|$ for all $u, v \in \mathcal{H}$. We consider and investigate the model

$$Y_{t,T} = m\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) + \sigma\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) \varepsilon_t, \quad t = 1, \dots, T,$$

where $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$ are random variables, $Y_{t,T}$ is real-valued, and $X_{t,T}$ takes values in the semi-metric space \mathcal{H} equipped with the semi-metric $d_\theta(\cdot, \cdot)$. Moreover, we construct an estimator for the nonparametric regression operator within this model and derive its asymptotic properties under mild conditions. Specifically, we obtain uniform convergence rates for both the general kernel estimator and the NW estimator of the regression function. Our results demonstrate that, under sufficient conditions, the general kernel estimator $\hat{\psi}_\theta(u, x)$ converges to its mean $\mathbb{E}[\hat{\psi}_\theta(u, x)]$ at the rate

$$O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}}\right),$$

and the NW estimator $\hat{m}_\theta(u, x)$ converges to $m_\theta(u, x)$ at the rate

$$O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} + h^{2\wedge\beta}\right).$$

The convergence rate of the NW estimator comprises two components: The first term relates to the variability of the estimate and depends on the concentration of the random variables $X_{t,T}$ as characterized by the small-ball probability $\phi_\theta(h)$, while the second term pertains to the bias of the estimate, which is influenced by the smoothness condition imposed on the operator $m_\theta(u, x)$. Specifically, the bias term depends on the parameter β and results from the application of the Lipschitz

assumption. To achieve a more efficient estimator for $m_\theta(u, x)$, it is crucial to minimize the dispersion of the functional data, thereby increasing the concentration of the random variables and maximizing the small-ball probability $\phi_\theta(h)$. A higher value of $\phi_\theta(h)$ leads to a faster convergence rate.

Consistent with the approaches of [57] and [83], one may explore the uniform convergence rate of $\hat{m}_\theta(u, x)$ over the domain $(1 - C_1 h, 1] \times \{x\}$ for forecasting purposes. This can be achieved by employing boundary-corrected kernels or one-sided kernels, assuming they are compactly supported and Lipschitz continuous to satisfy the present theoretical framework. It is important to note that achieving a high concentration value $\phi_\theta(h)$ is directly linked to the structure of the underlying space, which can be optimized by defining an appropriate semi-metric, such as the $d_\theta(\cdot, \cdot)$ introduced above. This optimization is further enhanced by selecting $k(\theta) = \arg \min_{k \in \{1, 2, \dots, N_{\theta, T}\}} \|\theta - \theta_k\|$ for θ , as demonstrated in the proof of Proposition 3.1. Consequently, the choice of θ plays a pivotal role in controlling $\phi_\theta(h)$. Additionally, employing an estimator $\theta(t) \in \mathcal{H}$ may provide more effective estimates and yield more reliable forecasting results. A pertinent reference for such an estimator is found in [63] (page 672).

Finally, we refer to Section 4 to illustrate the finite-sample behavior of the estimator. Our simulation study supports the pointwise convergence of $\hat{m}_\theta(u, x)$ to $m_\theta(u, x)$, as demonstrated using a first-order functional autoregressive process $X_{i,T}$. The asymptotic tightness of the estimator is evidenced by Figure 2, where the box plots shrink as T increases, indicating that the mean squared error (MSE) becomes asymptotically negligible for large T , which is consistent with our theoretical findings. Additionally, Figure 3 corroborates Theorem 3.2 by demonstrating the asymptotic normality of the estimator. To provide methodological recommendations for using the proposed estimators, it would be beneficial to conduct extensive Monte Carlo experiments comparing our procedures with other alternatives in the literature. However, this is beyond the scope of the present paper.

Author contributions

Breix Michael Agua and Salim Bouzebda: Conceptualization, methodology, investigation, writing—original draft, writing—review & editing. All authors of this article have been contributed equally. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

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

Acknowledgments

Mr. Agua's research is supported by the Department of Science and Technology - Science Education Institute (DOST-SEI) of the Philippine Government in partnership with Campus France through the PhilFrance-DOST Scholarship grant, which are greatly acknowledged. The authors extend their sincere gratitude to the Editor-in-Chief, the Associate Editor, and the three referees for their invaluable feedback. Their insightful comments have greatly refined and focused the original work, resulting in a markedly improved presentation.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Salim Bouzebda is the Guest Editor of special issue “Advances in Statistical Inference and Stochastic Processes: Theory and Applications” for AIMS Mathematics. Salim Bouzebda was not involved in the editorial review and the decision to publish this article.

References

1. A. Ait-Saïdi, F. Ferraty, R. Kassa, P. Vieu, Cross-validated estimations in the single-functional index model, *Statistics*, **42** (2008), 475–494. <https://doi.org/10.1080/02331880801980377>
2. I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Nonparametric estimation of expectile regression in functional dependent data, *J. Nonparametr. Stat.*, **34** (2022), 250–281. <https://doi.org/10.1080/10485252.2022.2027412>
3. G. Aneiros, P. Vieu, Partial linear modelling with multi-functional covariates, *Comput. Stat.*, **30** (2015), 647–671. <https://doi.org/10.1007/s00180-015-0568-8>
4. G. Aneiros-Pérez, P. Vieu, Automatic estimation procedure in partial linear model with functional data, *Statist. Papers*, **52** (2011), 751–771. <https://doi.org/10.1007/s00362-009-0280-2>
5. S. Attaoui, N. Ling, Asymptotic results of a nonparametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications, *Metrika*, **79** (2016), 485–511. <https://doi.org/10.1007/s00184-015-0564-6>
6. S. Attaoui, A. Laksaci, E. O. Said, A note on the conditional density estimate in the single functional index model, *Statist. Probab. Lett.*, **81** (2011), 45–53. <https://doi.org/10.1016/j.spl.2010.09.017>
7. S. Attaoui, B. Bentata, S. Bouzebda, A. Laksaci, The strong consistency and asymptotic normality of the kernel estimator type in functional single index model in presence of censored data, *AIMS Mathematics*, **9** (2024), 7340–7371. <https://doi.org/10.3934/math.2024356>
8. A. Aue, A. van Delft, Testing for stationarity of functional time series in the frequency domain, *Ann. Statist.*, **48** (2020), 2505–2547. <https://doi.org/10.1214/19-AOS1895>
9. B. Auestad, D. Tjøstheim, Identification of nonlinear time series: first order characterization and order determination, *Biometrika*, **77** (1990), 669–687. <https://doi.org/10.2307/2337091>
10. K. Benhenni, F. Ferraty, M. Rachdi, P. Vieu, Local smoothing regression with functional data, *Comput. Statist.*, **22** (2007), 353–369. <https://doi.org/10.1007/s00180-007-0045-0>
11. N. E. Berrahou, S. Bouzebda, L. Douge, Functional uniform-in-bandwidth moderate deviation principle for the local empirical processes involving functional data, *Math. Meth. Stat.*, **33** (2024), 26–69. <https://doi.org/10.3103/S1066530724700030>
12. S. Bhattacharjee, H. G. Müller, Single index Fréchet regression, *Ann. Statist.*, **51** (2023), 1770–1798. <https://doi.org/10.1214/23-AOS2307>
13. V. I. Bogachev, *Gaussian measures*, In: *Mathematical surveys and monographs*, American Mathematical Society, **62** (1998).

14. D. Bosq, *Linear processes in function spaces: Theory and applications*, New York: Springer, 2000. <https://doi.org/10.1007/978-1-4612-1154-9>
15. O. Bouanani, S. Bouzebda, Limit theorems for local polynomial estimation of regression for functional dependent data, *AIMS Mathematics*, **9** (2024), 23651–23691. <https://doi.org/10.3934/math.20241150>
16. S. Bouzebda, General tests of conditional independence based on empirical processes indexed by functions, *Jpn. J. Stat. Data Sci.*, **6** (2023), 115–177. <https://doi.org/10.1007/s42081-023-00193-3>
17. S. Bouzebda, Limit theorems in the nonparametric conditional single-index U -processes for locally stationary functional random fields under stochastic sampling design, *Mathematics*, **12** (2024), 1996. <https://doi.org/10.3390/math12131996>
18. S. Bouzebda, Uniform in number of neighbor consistency and weak convergence of k -Nearest neighbor single index conditional processes and k -Nearest neighbor single index conditional U -processes involving functional mixing data, *Symmetry*, **16** (2024), 1576. <https://doi.org/10.3390/sym16121576>
19. S. Bouzebda, Weak convergence of the conditional single index U -statistics for locally stationary functional time series, *AIMS Mathematics*, **9** (2024), 14807–14898. <https://doi.org/10.3934/math.2024720>
20. S. Bouzebda, S. Didi, Additive regression model for stationary and ergodic continuous time processes, *Comm. Statist. Theory Methods*, **46** (2017), 2454–2493. <https://doi.org/10.1080/03610926.2015.1048882>
21. S. Bouzebda, S. Didi, Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: asymptotic results, *Comm. Statist. Theory Methods*, **46** (2017), 1367–1406. <https://doi.org/10.1080/03610926.2015.1019144>
22. S. Bouzebda, S. Didi, Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes, *Rev. Mat. Complut.*, **34** (2021), 811–852. <https://doi.org/10.1007/s13163-020-00368-6>
23. S. Bouzebda, B. Nemouchi, Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U -statistics involving functional data, *J. Nonparametr. Stat.*, **32** (2020), 452–509. <https://doi.org/10.1080/10485252.2020.1759597>
24. S. Bouzebda, B. Nemouchi, Weak-convergence of empirical conditional processes and conditional U -processes involving functional mixing data, *Stat. Inference Stoch. Process.*, **26** (2023), 33–88. <https://doi.org/10.1007/s11203-022-09276-6>
25. S. Bouzebda, A. Nezzal, Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U -statistics involving functional data, *Jpn. J. Stat. Data Sci.*, **5** (2022), 431–533. <https://doi.org/10.1007/s42081-022-00161-3>
26. S. Bouzebda, A. Nezzal, Uniform in number of neighbors consistency and weak convergence of k NN empirical conditional processes and k NN conditional U -processes involving functional mixing data, *AIMS Mathematics*, **9** (2024), 4427–4550. <https://doi.org/10.3934/math.2024218>
27. S. Bouzebda, I. Soukarieh, Nonparametric conditional U -processes for locally stationary functional random fields under stochastic sampling design, *Mathematics*, **11** (2023), 16. <https://doi.org/10.3390/math11010016>

28. S. Bouzebda, A. Laksaci, M. Mohammedi, Single index regression model for functional quasi-associated time series data, *REVSTAT*, **20** (2022), 605–631. <https://doi.org/10.57805/revstat.v20i5.391>
29. S. Bouzebda, A. Nezzal, T. Zari, Uniform consistency for functional conditional U -statistics using delta-sequences, *Mathematics*, **11** (2023), 161. <https://doi.org/10.3390/math11010161>
30. S. Bouzebda, A. Laksaci, M. Mohammedi, The k -nearest neighbors method in single index regression model for functional quasi-associated time series data, *Rev. Mat. Complut.*, **36** (2023), 361–391. <https://doi.org/10.1007/s13163-022-00436-z>
31. D. Chen, P. Hall, H. G. Müller, Single and multiple index functional regression models with nonparametric link, *Ann. Statist.*, **39** (2011), 1720–1747. <http://dx.doi.org/10.1214/11-AOS882>
32. R. Chen, R. S. Tsay, Functional-coefficient autoregressive models, *J. Amer. Statist. Assoc.*, **88** (1993), 298–308. <https://doi.org/10.1080/01621459.1993.10594322>
33. A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Statist. Plann. Inference*, **147** (2014), 1–23. <https://doi.org/10.1016/j.jspi.2013.04.002>
34. R. Dahlhaus, On the kullback-leibler information divergence of locally stationary processes, *Stochastic Process. Appl.*, **62** (1996), 139–168. [https://doi.org/10.1016/0304-4149\(95\)00090-9](https://doi.org/10.1016/0304-4149(95)00090-9)
35. R. Dahlhaus, Fitting time series models to nonstationary processes, *Ann. Statist.*, **25** (1997), 1–37. <https://doi.org/10.1214/aos/1034276620>
36. R. Dahlhaus, W. Polonik, Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes, *Ann. Statist.*, **34** (2006), 2790–2824. <https://doi.org/10.1214/009053606000000867>
37. R. Dahlhaus, S. S. Rao, Statistical inference for time-varying ARCH processes, *Ann. Statist.*, **34** (2006), 1075–1114. <https://doi.org/10.1214/009053606000000227>
38. R. Dahlhaus, S. Richter, W. B. Wu, Towards a general theory for nonlinear locally stationary processes, *Bernoulli*, **25** (2019), 1013–1044. <https://doi.org/10.3150/17-BEJ1011>
39. Y. A. Davydov, Convergence of distributions generated by stationary stochastic processes, *Theory Probab. Appl.*, **13** (1968), 691–696. <https://doi.org/10.1137/1113086>
40. Z. C. Elmezouar, F. Alshahrani, I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Strong consistency rate in functional single index expectile model for spatial data, *AIMS Mathematics*, **9** (2024), 5550–5581. <https://doi.org/10.3934/math.2024269>
41. J. Fan, Q. Yao, *Nonlinear time series: Nonparametric and parametric methods*, New York: Springer, 2003. <https://doi.org/10.1007/978-0-387-69395-8>
42. S. Feng, P. Tian, Y. Hu, G. Li, Estimation in functional single-index varying coefficient model, *J. Statist. Plann. Inference*, **214** (2021), 62–75. <https://doi.org/10.1016/j.jspi.2021.01.003>
43. F. Ferraty, P. Vieu, *Nonparametric functional data analysis: Theory and practice*, New York: Springer, 2006. <https://doi.org/10.1007/0-387-36620-2>
44. F. Ferraty, A. Peuch, P. Vieu, Modèle à indice fonctionnel simple, *C. R. Acad. Sci. Paris, Ser. I*, **336** (2003), 1025–1028. [https://doi.org/10.1016/S1631-073X\(03\)00239-5](https://doi.org/10.1016/S1631-073X(03)00239-5)
45. L. Ferré, A. F. Yao, Smoothed functional inverse regression, *Statist. Sinica*, **15** (2005), 665–683.

46. T. Gasser, P. Hall, B. Presnell, Nonparametric estimation of the mode of a distribution of random curves, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60** (1998), 681–691. <https://doi.org/10.1111/1467-9868.00148>
47. G. Geenens, Curse of dimensionality and related issues in nonparametric functional regression, *Statist. Surveys*, **5** (2011), 30–43. <http://dx.doi.org/10.1214/09-SS049>
48. A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.*, **146** (2016), 1–6. <https://doi.org/10.1016/j.jmva.2015.12.001>
49. V. V. Gorodeckii, On the strong mixing property for linearly generated sequences, *Theory Probab. Appl.*, **22** (1978), 411–413. <https://doi.org/10.1137/1122049>
50. L. Gu, L. Yang, Oracally efficient estimation for single-index link function with simultaneous confidence band, *Electron. J. Statist.*, **9** (2015), 1540–1561.
51. P. Hall, Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, **67** (1984), 175–196. <https://doi.org/10.1007/BF00535267>
52. P. Hall, C. C. Heyde, *Martingale limit theory and its application*, Academic Press, 1980. <https://doi.org/10.1016/C2013-0-10818-5>
53. W. Hardle, J. S. Marron, Optimal bandwidth selection in nonparametric regression function estimation, *Ann. Statist.*, (1985), 1465–1481. <https://doi.org/10.1214/aos/1176349748>
54. W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, *Ann. Statist.*, **21** (1993), 157–178. <https://doi.org/10.1214/aos/1176349020>
55. L. Horváth, P. Kokoszka, *Inference for functional data with applications*, New York: Springer, 2012. <https://doi.org/10.1007/978-1-4614-3655-3>
56. Z. Jiang, Z. Huang, J. Zhang, Functional single-index composite quantile regression, *Metrika*, **86** (2023), 595–603. <https://doi.org/10.1007/s00184-022-00887-w>
57. D. Kurisu, Nonparametric regression for locally stationary functional time series, *Electron. J. Stat.*, **16** (2022), 3973–3995. <https://doi.org/10.1214/22-EJS2041>
58. J. Li, C. Huang, H. Zhu, A functional varying-coefficient single-index model for functional response data, *J. Amer. Statist. Assoc.*, **112** (2017), 1169–1181. <https://doi.org/10.1080/01621459.2016.1195742>
59. H. Liang, X. Liu, R. Li, C. L. Tsai, Estimation and testing for partially linear single-index models, *Ann. Statist.*, **38** (2010), 3811–3836. <https://doi.org/10.1214/10-AOS835>
60. E. Liebscher, Strong convergence of sums of α -mixing random variables with applications to density estimation. *Stoch. Process. Appl.*, **65** (1996), 69–80. [https://doi.org/10.1016/S0304-4149\(96\)00096-8](https://doi.org/10.1016/S0304-4149(96)00096-8)
61. N. Ling, P. Vieu, Nonparametric modelling for functional data: selected survey and tracks for future, *Statistics*, **52** (2018), 934–949. <https://doi.org/10.1080/02331888.2018.1487120>
62. N. Ling, L. Cheng, P. Vieu, Single functional index model under responses MAR and dependent observations, In: *Functional and high-dimensional statistics and related fields*, Cham: Springer, 2020, 161–168. https://doi.org/10.1007/978-3-030-47756-1_22

63. N. Ling, L. Cheng, P. Vieu, H. Ding, Missing responses at random in functional single index model for time series data, *Stat. Papers*, **63** (2022), 665–692. <https://doi.org/10.1007/s00362-021-01251-2>
64. E. Masry, Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stoch. Process. Appl.*, **115** (2005), 155–177. <https://doi.org/10.1016/j.spa.2004.07.006>
65. E. Masry, D. Tjøstheim, Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality, *Econometric Theory*, **11** (1995), 258–289.
66. E. Masry, D. Tjøstheim, Additive nonlinear ARX time series and projection estimates, *Econometric Theory*, **13** (1997), 214–252. <https://doi.org/10.1017/S0266466600005739>
67. M. Mohammadi, S. Bouzebda, A. Laksaci, The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data, *J. Multivariate Anal.*, **181** (2021), 104673. <https://doi.org/10.1016/j.jmva.2020.104673>
68. M. Mohammadi, S. Bouzebda, A. Laksaci, O. Bouanani, Asymptotic normality of the k-NN single index regression estimator for functional weak dependence data, *Comm. Statist. Theory Methods*, **53** (2024), 3143–3168. <https://doi.org/10.1080/03610926.2022.2150823>
69. M. H. Neumann, R. von Sachs, Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, *Ann. Statist.*, **25** (1997), 38–76. <https://doi.org/10.1214/aos/1034276621>
70. Y. Nie, L. Wang, J. Cao, Estimating functional single index models with compact support, *Environmetrics*, **34** (2023), e2784. <https://doi.org/10.1002/env.2784>
71. S. Novo, G. Aneiros, P. Vieu, Automatic and location-adaptive estimation in functional single-index regression, *J. Nonparametr. Stat.*, **31** (2019), 364–392. <https://doi.org/10.1080/10485252.2019.1567726>
72. M. B. Priestley, Evolutionary spectra and non-stationary processes, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **27** (1965), 204–229. <https://doi.org/10.1111/j.2517-6161.1965.tb01488.x>
73. M. Rachdi, P. Vieu, Nonparametric regression for functional data: Automatic smoothing parameter selection, *J. Statist. Plann. Inference*, **137** (2007), 2784–2801. <https://doi.org/10.1016/j.jspi.2006.10.001>
74. M. Rachdi, M. Alahiane, I. Ouassou, P. Vieu, Generalized functional partially linear single-index models, In: *Functional and high-dimensional statistics and related fields*, Cham: Springer, 2020, 221–228. https://doi.org/10.1007/978-3-030-47756-1_29
75. J. O. Ramsay, B. W. Silverman, *Functional data analysis*, New York: Springer, 2010. <https://doi.org/10.1007/b98888>
76. K. Sakiyama, M. Taniguchi, Discriminant analysis for locally stationary processes, *J. Multivariate Anal.*, **90** (2004), 282–300. <https://doi.org/10.1016/j.jmva.2003.08.002>
77. J. Q. Shi, T. Choi, *Gaussian process regression analysis for functional data*, New York: Chapman and Hall/CRC, 2011. <https://doi.org/10.1201/b11038>
78. B. W. Silverman, *Density estimation for statistics and data analysis*, New York: Routledge, 2018. <https://doi.org/10.1201/9781315140919>

79. R. A. Silverman, Locally stationary random processes, *IRE Trans. Inform. Theory*, **3** (1957), 182–187. <https://doi.org/10.1109/TIT.1957.1057413>
80. W. Stute, L. X. Zhu, Nonparametric checks for single-index models, *Ann. Statist.*, **33** (2005), 1048–1083. <https://doi.org/10.1214/009053605000000020>
81. Q. Tang, L. Kong, D. Rupper, R. J. Karunamuni, Partial functional partially linear single-index models, *Statist. Sinica*, **31** (2021), 107–133.
82. A. van Delft, M. Eichler, Locally stationary functional time series, *Electron. J. Statist.*, **12** (2018), 107–170. <https://doi.org/10.1214/17-EJS1384>
83. M. Vogt, Nonparametric regression for locally stationary time series, *Ann. Statist.*, **40** (2012), 2601–2633. <https://doi.org/10.1214/12-AOS1043>
84. C. S. Withers, Conditions for linear processes to be strong-mixing, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, **57** (1981), 477–480. <https://doi.org/10.1007/BF01025869>
85. J. T. Zhang, *Analysis of variance for functional data*, 1st Eds., New York: Chapman and Hall/CRC, 2013. <https://doi.org/10.1201/b15005>
86. H. Zhu, R. Zhang, Y. Liu, H. Ding, Robust estimation for a general functional single index model via quantile regression, *J. Korean Stat. Soc.*, **51** (2022), 1041–1070. <https://doi.org/10.1007/s42952-022-00174-4>

A. Appendix-A

In this section, we present detailed proofs of the conjectures under investigation. We begin by proving Proposition 3.1. The proof starts by decomposing $\hat{\psi}_\theta(u, x)$ into two terms: $\hat{\psi}_\theta^{(1)}(u, x)$ and $\hat{\psi}_\theta^{(2)}(u, x)$.

Proof of Proposition 3.1. We first define $B = [0, 1]$, $\alpha_T = \sqrt{\frac{\log T}{Th\phi_\theta(h)}}$, and $\tau_T = \rho_T T^{1/\zeta}$ with $\rho_T = (\log T)^{\zeta_0}$ for some $\zeta_0 > 0$. Now, define

$$\hat{\psi}_\theta^{(1)}(u, x) = \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)},$$

$$\hat{\psi}_\theta^{(2)}(u, x) = \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) W_{t,T} \mathbb{1}_{(|W_{t,T}| > \tau_T)}.$$

Clearly,

$$\hat{\psi}_\theta(u, x) = \left(\hat{\psi}_\theta^{(1)}(u, x) - \mathbb{E}[\hat{\psi}_\theta^{(1)}(u, x)]\right) + \left(\hat{\psi}_\theta^{(2)}(u, x) - \mathbb{E}[\hat{\psi}_\theta^{(2)}(u, x)]\right).$$

From this, we outline the proof of Proposition 3.1 into two steps as follows:

- (i) $\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} \left| \hat{\psi}_\theta^{(2)}(u, x) - \mathbb{E}[\hat{\psi}_\theta^{(2)}(u, x)] \right| = O_{\mathbb{P}}(\alpha_T)$, and
- (ii) $\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} \left| \hat{\psi}_\theta^{(1)}(u, x) - \mathbb{E}[\hat{\psi}_\theta^{(1)}(u, x)] \right| = O_{\mathbb{P}}(\alpha_T)$.

Step (i): We first tackle $\hat{\psi}_\theta^{(2)}(u, x) - \mathbb{E}[\hat{\psi}_\theta^{(2)}(u, x)]$. Observe that

$$\begin{aligned}
& \mathbb{P}\left(\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |\hat{\psi}_\theta^{(2)}(u, x)| > \alpha_T\right) \\
&= \mathbb{P}\left(\left\{\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |\hat{\psi}_\theta^{(2)}(u, x)| > \alpha_T\right\}\right. \\
&\quad \left. \cup \left\{\sup_{\theta \in \Theta} \sup_{x \in S} \bigcup_{t=1}^T |W_{t,T}| > \tau_T\right\} \cap \left\{\left\{\sup_{\theta \in \Theta} \sup_{x \in S} \bigcup_{t=1}^T |W_{t,T}| > \tau_T\right\}^c\right\}\right) \\
&\leq \mathbb{P}\left(\left\{\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |\hat{\psi}_\theta^{(2)}(u, x)| > \alpha_T\right\} \cap \left\{\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} \bigcup_{t=1}^T |W_{t,T}| > \tau_T\right\}\right) \\
&\quad + \mathbb{P}\left(\left\{\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |\hat{\psi}_\theta^{(2)}(u, x)| > \alpha_T\right\} \cap \left\{\left\{\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} \bigcup_{t=1}^T |W_{t,T}| > \tau_T\right\}^c\right\}\right) \\
&\leq \mathbb{P}\left(\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |W_{t,T}| > \tau_T\right) + \mathbb{P}(\emptyset) \text{ for some } t = 1, 2, \dots, T \\
&\leq \tau_T^{-\zeta} \sum_{t=1}^T \mathbb{E}\left[\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |W_{t,T}|^\zeta\right] \leq \tau_T^{-\zeta} T = \rho_T^{-\zeta} \rightarrow 0 \text{ as } T \rightarrow \infty.
\end{aligned}$$

Now, using Assumptions 3.1 (M1) and (M4), and Assumption 3.3 (KB2), we obtain

$$\begin{aligned}
K_{2,h}(d_\theta(x, X_{t,T})) &= K_{2,h}(d_\theta(x, X_{t,T})) + (K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)})) - K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)}))) \\
&\leq |K_{2,h}(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)}))| + K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)})) \\
&\leq h^{-1} |d_\theta(x, X_{t,T}) - d_\theta(x, X_{t,T}^{(t/T)})| + K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)})) \\
&\leq h^{-1} |d_\theta(X_{t,T}, X_{t,T}^{(t/T)})| + K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)})) \\
&\leq (Th)^{-1} U_{t,T}^{(t/T)} + K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)})), \tag{A.1}
\end{aligned}$$

where by using the elementary properties of summations and by the definition of $d_\theta(\cdot, \cdot)$, we get

$$\begin{aligned}
& d_\theta(x, X_{t,T}) - d_\theta(x, X_{t,T}^{(t/T)}) \\
&= |\langle \theta, x - X_{t,T} \rangle| - |\langle \theta, x - X_{t,T}^{(t/T)} \rangle| \\
&= \left| \int_a^b \theta(s)(x - X_{t,T})(s) ds \right| - \left| \int_a^b \theta(s)(x - X_{t,T}^{(t/T)})(s) ds \right| \\
&= \left| \int_a^b \theta(s)(x(s) - X_{t,T}(s)) ds \right| - \left| \int_a^b \theta(s)(x(s) - X_{t,T}^{(t/T)}(s)) ds \right| \\
&= \left| \int_a^b \theta(s)x(s) ds - \int_a^b \theta(s)X_{t,T}(s) ds \right| - \left| \int_a^b \theta(s)x(s) ds - \int_a^b \theta(s)X_{t,T}^{(t/T)}(s) ds \right| \\
&\leq \left| \int_a^b \theta(s)x(s) ds \right| + \left| \int_a^b \theta(s)X_{t,T}(s) ds \right| - \left| \int_a^b \theta(s)x(s) ds \right| - \left| \int_a^b \theta(s)X_{t,T}^{(t/T)}(s) ds \right| \\
&= \left| \int_a^b \theta(s)X_{t,T}(s) ds \right| - \left| \int_a^b \theta(s)X_{t,T}^{(t/T)}(s) ds \right|
\end{aligned}$$

$$\begin{aligned}
&= \left| \int_a^b (\theta(s)X_{t,T}(s)ds - \theta(s)X_{t,T}^{(t/T)}(s))ds \right| \\
&= \left| \int_a^b \theta(s)(X_{t,T} - X_{t,T}^{(t/T)}(s))ds \right| \\
&= |\langle \theta, X_{t,T} - X_{t,T}^{(t/T)} \rangle| \\
&= d_\theta(X_{t,T}, X_{t,T}^{(t/T)}).
\end{aligned}$$

Hence, combining the above results, we have

$$\begin{aligned}
\mathbb{E} \left[K_{2,h}(d_\theta(x, X_{t,T})) | W_{t,T} | \mathbb{1}_{(|W_{t,T}| > \tau_T)} \right] &\lesssim \tau_T^{-(\zeta-1)} \mathbb{E} [K_{2,h}(d_\theta(x, X_{t,T})) | W_{t,T} |^\zeta] \\
&\lesssim \tau_T^{-(\zeta-1)} \mathbb{E} [K_{2,h}(d_\theta(x, X_{t,T}))] \\
&\lesssim \tau_T^{-(\zeta-1)} \mathbb{E} \left[(Th)^{-1} U_{t,T}^{(t/T)} + K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)})) \right] \\
&\lesssim \frac{1}{Th\tau_T^{\zeta-1}} \mathbb{E} [U_{t,T}^{(t/T)}] + \frac{1}{\tau_T^{\zeta-1}} \mathbb{E} [K_{2,h}(d_\theta(x, X_{t,T}^{(t/T)}))] \\
&\lesssim \frac{1}{Th\tau_T^{\zeta-1}} + \frac{1}{\tau_T^{\zeta-1}} \mathbb{E} \left[\mathbb{1}_{(d_\theta(x, X_{t,T}^{(t/T)}) \leq h)} \right] \\
&\lesssim \frac{1}{Th\tau_T^{\zeta-1}} + \frac{1}{\tau_T^{\zeta-1}} F_{t/T}(h; x, \theta) \\
&\lesssim \frac{1}{\tau_T^{\zeta-1}} \phi_\theta(h).
\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}
\frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \mathbb{E} \left[K_{2,h}(d_\theta(x, X_{t,T})) | W_{t,T} | \mathbb{1}_{(|W_{t,T}| > \tau_T)} \right] &\lesssim \frac{1}{\tau_T^{\zeta-1}} \phi_\theta(h) \times \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \\
&= \frac{1}{\tau_T^{\zeta-1}} \left(\frac{1}{Th} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \right).
\end{aligned}$$

By Lemma B.2, we have

$$\mathbb{E} \left[|\hat{\psi}_\theta^{(2)}(u, x)| \right] \lesssim \frac{1}{\tau_T^{\zeta-1}} \times \left(O\left(\frac{1}{Th^2}\right) + o(h) \right) \lesssim \frac{1}{\tau_T^{\zeta-1}} \lesssim \alpha_T.$$

Inferring from the last result, we get

$$\sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in [0,1]} \left| \hat{\psi}_\theta^{(2)}(u, x) - \mathbb{E} \left[\hat{\psi}_\theta^{(2)}(u, x) \right] \right| = O_{\mathbb{P}}(\alpha_T).$$

Step (ii): We are left to show that

$$\sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in [0,1]} \left| \hat{\psi}_\theta^{(1)}(u, x) - \mathbb{E} \left[\hat{\psi}_\theta^{(1)}(u, x) \right] \right| = O_{\mathbb{P}}(\alpha_T).$$

To achieve the desired result, assume that $S_{\mathcal{H}}$ and $\Theta_{\mathcal{H}}$ are compact subsets of \mathcal{H} . Suppose also that $N_{S,T}$, and $N_{\theta,T}$ are the minimal number of balls in \mathcal{H} with radius h necessary to cover $S_{\mathcal{H}}$ and $\Theta_{\mathcal{H}}$ with centers $x_1, \dots, x_{N_{S,T}}$, and $\theta_1, \dots, \theta_{N_{\theta,T}}$, respectively such that $N_{S,T} \cdot N_{\theta,T} \leq C \cdot \frac{1}{\alpha_T}$ balls

$$B_{\theta}(x, h) = \{y \in \mathcal{H} : d_{\theta}(x, y) \leq h\}.$$

More discussion on the covering numbers may be found in [11]. We further suppose that with $N_{B_{i,T}} \leq C \frac{1}{h\alpha_T}$ balls $B_{i,T}$, \mathbf{B} is covered. Here,

$$B_{i,T} = \{u \in \mathbb{R} : |u - u_i| \leq \alpha_T h\},$$

where u_i is the midpoint of $B_{i,T}$. Now, assume that for $(w, v) \in \mathbb{R}^2$,

$$K^*(w, v) = C \mathbb{1}_{(|w| \leq 2C_1)} K_2(v).$$

With a sufficiently large T and for $u \in B_{i,T}$, we obtain

$$\begin{aligned} \left| K_{1,h}\left(u - \frac{t}{T}\right) - K_{1,h}\left(u_i - \frac{t}{T}\right) \right| K_{2,h}(d_{\theta}(x, X_{t,T})) &\leq C_1 \left| \left(u - \frac{t}{T}\right) - \left(u_i - \frac{t}{T}\right) \right| K_{2,h}(d_{\theta}(x, X_{t,T})) \\ &\leq C_1 \left| \left(u_i - \frac{t}{T}\right) + \left(u_i - \frac{t}{T}\right) \right| K_{2,h}(d_{\theta}(x, X_{t,T})) \\ &\leq 2C_1 \left| u_i - \frac{t}{T} \right| K_{2,h}(d_{\theta}(x, X_{t,T})) \\ &\leq C \mathbb{1}_{\left(\left|u_i - \frac{t}{T}\right| \leq 2C_1\right)} K_{2,h}(d_{\theta}(x, X_{t,T})) \\ &\leq \alpha_T K_h^*\left(u_i - \frac{t}{T}, d_{\theta}(x, X_{t,T})\right), \end{aligned}$$

where $K_h^*(v) = K^*(v/h)$. We now define $\bar{\psi}_{\theta}^{(1)}(u_i, x)$ as

$$\bar{\psi}_{\theta}^{(1)}(u, x) = \frac{1}{Th\phi_{\theta}(h)} \sum_{t=1}^T K_h^*\left(u - \frac{t}{T}, d_{\theta}(x, X_{t,T})\right) |W_{t,T}| \mathbb{1}_{(|W_{t,T}| \leq \tau_T)}.$$

Now, we write $t(x) = \arg \min_{t \in \{1, 2, \dots, N_{S,T}\}} \|x - x_t\|$, and $k(\theta) = \arg \min_{k \in \{1, 2, \dots, N_{\theta,T}\}} \|\theta - \theta_k\|$, and use a similar decomposition in [63] which is given as follows

$$\begin{aligned} \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{\theta}^{(1)}(u, x) - \mathbb{E}[\hat{\psi}_{\theta}^{(1)}(u, x)] \right| &\leq \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{\theta}^{(1)}(u, x) - \hat{\psi}_{\theta}^{(1)}(u, x_{t(x)}) \right| \\ &\quad + \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{\theta}^{(1)}(u, x_{t(x)}) - \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right| \\ &\quad + \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E}[\hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)})] \right| \\ &\quad + \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E}[\hat{\psi}_{\theta}^{(1)}(u, x_{t(x)})] \right| \\ &\quad + \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in S_{\mathcal{H}}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{\theta}^{(1)}(u, x_{t(x)}) - \mathbb{E}[\hat{\psi}_{\theta}^{(1)}(u, x)] \right| \\ &=: Q_{1,\theta} + Q_{2,\theta} + Q_{3,\theta} + Q_{4,\theta} + Q_{5,\theta}. \end{aligned}$$

We first deal with the term $Q_{3,\theta}$. Since for sufficiently large M , and

$$\mathbb{E} \left[|\bar{\psi}_\theta^{(1)}(u, x)| \right] \leq M < \infty,$$

for any $x \in \mathcal{H}$, we get

$$\begin{aligned} & \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right] \right| \\ & \leq \sup_{\theta \in \Theta} \sup_{x \in S} \left| \hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| \\ & \quad + \sup_{\theta \in \Theta} \sup_{x \in S} \alpha_T \left(\left| \bar{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right| + \mathbb{E} \left[\left| \bar{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right| \right] \right) \\ & \leq \sup_{\theta \in \Theta} \sup_{x \in S} \left| \hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| \\ & \quad + \sup_{\theta \in \Theta} \sup_{x \in S} \left| \bar{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\bar{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| + 2M\alpha_T. \end{aligned}$$

Thus, we have

$$\begin{aligned} & \mathbb{P} \left(\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right] \right| > 4M\alpha_T \right) \\ & \leq \mathbb{P} \left(N_{S,T} N_{\theta,T} N_{B_{i,T}} \max_{k(\theta) \in \{1, 2, \dots, N_{\theta,T}\}} \max_{t(x) \in \{1, 2, \dots, N_{S,T}\}} \right. \\ & \quad \left. \max_{1 \leq i \leq N_{B_{i,T}}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| > 4M\alpha_T \right) \\ & \leq N_{S,T} N_{\theta,T} N_{B_{i,T}} \max_{1 \leq i \leq N_{B_{i,T}}} \mathbb{P} \left(\left| \hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| > 4M\alpha_T \right) \\ & \leq Q_{3,1,T,\theta} + Q_{3,2,T,\theta}, \end{aligned}$$

where

$$Q_{3,1,T,\theta} = N_{S,T} N_{\theta,T} N_{B_{i,T}} \max_{1 \leq i \leq N_{B_{i,T}}} \mathbb{P} \left(\left| \hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| > M\alpha_T \right),$$

and

$$Q_{3,2,T,\theta} = N_{S,T} N_{\theta,T} N_{B_{i,T}} \max_{1 \leq i \leq N_{B_{i,T}}} \mathbb{P} \left(\left| \bar{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) - \mathbb{E} \left[\bar{\psi}_{k(\theta)}^{(1)}(u_i, x_{t(x)}) \right] \right| > M\alpha_T \right).$$

Before proceeding with the further parts of the proof, we first lay down some important notations. For $t = 1, 2, \dots, T$, and $\mathbb{1}_A(\cdot)$ is an indicator function of a set A , we write

$$\begin{aligned} \Delta_t(u, x; \theta) &= K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, x_t)), \\ \nabla_t &= \frac{1}{\phi_\theta(h)} \mathbb{1}_{\{B_\theta(x,h) \cup B_\theta(x_t,h)\}}(x_t), \text{ and} \\ \Gamma_t &= \frac{1}{\phi_\theta(h)} \mathbb{1}_{\{B_\theta(x_t,h) \cup B_{k(\theta)}(x_t,h)\}}(x_t). \end{aligned}$$

We now go back to showing the bound of $Q_{3,1,T,\theta}$, and $Q_{3,2,T,\theta}$. Observe that they can be analyzed in an almost similar manner. Thus, we focus our attention to $Q_{3,1,T,\theta}$. Now, for $t = 1, \dots, T$,

$$\begin{aligned} & \Delta_t(u_i, x_{t(x)}; k(\theta))W_{t,T}\mathbb{1}_{(|W_{t,T}| \leq \tau_T)} - \mathbb{E} \left[\Delta_t(u_i, x_{t(x)}; k(\theta))W_{t,T}\mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \right] \\ &= K_{1,h} \left(u_i - \frac{t}{T} \right) K_{2,h}(d_{k(\theta)}(x_{t(x)}, x_t)) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \\ & \quad - \mathbb{E} \left[K_{1,h} \left(u_i - \frac{t}{T} \right) K_{2,h}(d_{k(\theta)}(x_{t(x)}, x_t)) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \right] \\ &= K_{1,h} \left(u_i - \frac{t}{T} \right) \left\{ K_{2,h}(d_{k(\theta)}(x_{t(x)}, x_t)) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \right. \\ & \quad \left. - \mathbb{E} \left[K_{2,h}(d_{k(\theta)}(x_{t(x)}, x_t)) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \right] \right\} \\ &=: Z_{t,T,1,\theta}(u_i, x_{t(x)}). \end{aligned}$$

Note that for each $(u, x_{t(x)})$, the array $\{Z_{t,T,1,\theta}(u_i, x_{t(x)})\}$ is α -mixing with mixing coefficients $\alpha_{Z,T,k(\theta)}(k)$ satisfying

$$\alpha_{Z,T,k(\theta)}(k) \leq \alpha(k).$$

We set $\varepsilon = M\alpha_T Th\phi_\theta(h)$, $b_T = C\tau_T$ for a sufficiently large constant $C > 0$, and $S_T = \frac{1}{\alpha_T \tau_T}$, and apply Lemma B.3. Moreover, with a constant C' independent of (u, x) , Theorem 2 of [64] can be extended to show that

$$\sigma_{S_T,T}^2 \leq C' S_T h\phi_\theta(h).$$

Therefore, for any $(u, x_{t(x)})$ and sufficiently large T , we obtain

$$\begin{aligned} \mathbb{P} \left(\left| \sum_{t=1}^T Z_{t,T,1,k(\theta)}(u_i, x_{t(x)}) \right| \geq \varepsilon \right) &\leq 4 \exp \left(- \frac{\varepsilon^2}{64\sigma_{S_T,T}^2 \frac{T}{S_T} + \frac{8}{3}\varepsilon b_T S_T} \right) + 4 \frac{T}{S_T} \alpha(S_T) \\ &\leq 4 \exp \left(- \frac{M^2 \alpha_T^2 T^2 h^2 \phi_\theta^2(h)}{64C' S_T h\phi_\theta(h) \frac{T}{S_T} + \frac{8}{3} M \alpha_T Th\phi_\theta(h) b_T S_T} \right) + 4 \frac{T}{S_T} \alpha(S_T) \\ &\leq 4 \exp \left(- \frac{Th\phi_\theta(h)(M^2 \alpha_T^2 Th\phi_\theta(h))}{Th\phi_\theta(h)(64C' + \frac{8}{3}CM)} \right) + 4 \frac{T}{S_T} \alpha(S_T) \\ &\leq 4 \exp \left(- \frac{M \left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} \right)^2 Th\phi_\theta(h)}{64 \frac{C'}{M} + \frac{8}{3}C} \right) + 4 \frac{T}{S_T} \alpha(S_T) \\ &\lesssim \exp \left(\log T \left(- \frac{M}{64 \frac{C'}{M} + \frac{8}{3}C} \right) \right) + TS_T^{-\gamma-1} \\ &= T^{-\frac{M}{64 \frac{C'}{M} + \frac{8}{3}C}} + T\alpha_T^{\gamma+1} \tau_T^{\gamma+1} \\ &\leq T^{-\frac{M}{64 + \frac{8}{3}C}} + T\alpha_T^{\gamma+1} \tau_T^{\gamma+1}. \end{aligned}$$

The last inequality holds by picking a very large $M > C'$. We can then show that

$$Q_{3,1,T,\theta} = R_{3,1,T,\theta}^{(1)} + R_{3,1,T,\theta}^{(2)} \rightarrow 0.$$

That is,

$$\begin{aligned}
 R_{3,1,T,\theta}^{(1)} &= N_{S,T} \cdot N_{\theta,T} \cdot N_{B_i,T} \cdot T^{-\frac{M}{64+\frac{8}{3}C}} \\
 &= \frac{C}{\alpha_T} \cdot \frac{1}{h\alpha_T} \cdot T^{-\frac{M}{64+\frac{8}{3}C}} \\
 &\lesssim \frac{1}{h\alpha_T^2} \cdot T^{-\frac{M}{64+\frac{8}{3}C}} \\
 &= \frac{1}{h\left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}}\right)^2} \cdot T^{-\frac{M}{64+\frac{8}{3}C}} \\
 &= \frac{T\phi_\theta(h)}{\log(T)} \cdot T^{-\frac{M}{64+\frac{8}{3}C}} \\
 &= \frac{\phi_\theta(h)}{\log(T)} \cdot T^{-\frac{M}{64+\frac{8}{3}C}+1} \\
 &= o(1),
 \end{aligned}$$

for sufficiently large $M > 0$ (i.e., $M > 64 + \frac{8}{3}C$), and the fact that $\phi_\theta(h) \rightarrow 0$ as $h \rightarrow 0$, and $\log(T) \rightarrow \infty$ as $T \rightarrow \infty$. On the other hand, we have

$$\begin{aligned}
 R_{3,1,T,\theta}^{(2)} &= N_{S,T} \cdot N_{\theta,T} \cdot N_{B_i,T} \cdot T\alpha_T^{\gamma+1} \tau_T^{\gamma+1} \\
 &= \frac{C}{\alpha_T} \cdot \frac{1}{h\alpha_T} \cdot T\alpha_T^{\gamma+1} \tau_T^{\gamma+1} \\
 &\lesssim \frac{1}{h} T\alpha_T^{\gamma+1-2} \tau_T^{\gamma+1} \\
 &= \frac{1}{h} \left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} \right)^{\gamma-1} \rho_T^{\gamma+1} T \cdot T^{\frac{\gamma+1}{\zeta}} \\
 &= \frac{1}{h} \left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}} \right)^{\gamma-1} ((\log T)^{\zeta_0})^{\gamma+1} T \cdot T^{\frac{\gamma+1}{\zeta}} \\
 &= \frac{1}{h} \cdot \frac{(\log T)^{\frac{\gamma-1}{2}}}{T^{\frac{\gamma-1}{2}} h^{\frac{\gamma-1}{2}} (\phi_\theta(h))^{\frac{\gamma-1}{2}}} (\log T)^{\zeta_0(\gamma+1)} T^{\frac{\gamma+1}{\zeta}+1} \\
 &= \frac{(\log T)^{\frac{\gamma-1}{2}+\zeta_0(\gamma+1)}}{T^{\frac{\gamma-1}{2}-\frac{\gamma+1}{\zeta}-1} h^{\frac{\gamma-1}{2}+1} (\phi_\theta(h))^{\frac{\gamma-1}{2}}} \\
 &= o(1),
 \end{aligned}$$

using the first regularity condition (R1) in Assumption 3.5. The desired result is achieved by imposing

$$\frac{\gamma-1}{2} > \frac{\gamma+1}{\zeta} - 1.$$

That is, $\gamma > 3$ just as Assumption 3.5 requires. This means that

$$\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{\mathcal{H}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right] \right| \rightarrow 0.$$

We now treat the term $Q_{1,\theta}$. Let

$$i(u) = \arg \min_{i \in \{1, 2, \dots, N_{B_i, T}\}} |u - u_i|.$$

We observe that asymptotically

$$\mathbb{E}(\hat{\psi}_\theta^{(1)}(u, x) - \hat{\psi}_\theta^{(1)}(u, x_{t(x)})) = 0.$$

Observe that

$$\begin{aligned} Q_{1,\theta} &= \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} |\hat{\psi}_\theta^{(1)}(u, x) \hat{\psi}_\theta^{(1)}(u, x_{t(x)}) - \mathbb{E}(\hat{\psi}_\theta^{(1)}(u, x) - \hat{\psi}_\theta^{(1)}(u, x_{t(x)}))| \\ &\leq \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} \frac{1}{Th\phi_\theta(h)} \\ &\quad \times \left| \sum_{t=1}^T W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \{ \Delta_t(u, x; \theta) - \Delta_t(u, x_{t(x)}; \theta) \} - \mathbb{E}(\hat{\psi}_\theta^{(1)}(u, x) - \hat{\psi}_\theta^{(1)}(u, x_{t(x)})) \right| \\ &\leq \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} \frac{C}{Th} \\ &\quad \times \left| \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) \cdot \{ W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \cdot \nabla_t - \mathbb{E} [W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \cdot \nabla_t] \} \right| \\ &= I_{Q_{1,\theta}}. \end{aligned}$$

We now deal $I_{Q_{1,\theta}}$. Then, we infer that

$$\begin{aligned} &\mathbb{P} \left(\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} \left| \sum_{t=1}^T Z_{t,T,1,\theta} \right| > \varepsilon \right) \\ &\leq \mathbb{P} \left(N_{S,T} N_{\theta,T} N_{B_{i,T}} \max_{k(\theta) \in \{1, 2, \dots, N_{\theta,T}\}} \max_{t(x) \in \{1, 2, \dots, N_{S,T}\}} \max_{1 \leq i \leq N_{B_{i,T}}} \left| \sum_{t=1}^T Z_{t,T,1,\theta} \right| > \varepsilon \right) \\ &\leq N_{S,T} N_{\theta,T} N_{B_{i,T}} \max_{1 \leq i \leq N_{B_{i,T}}} \cdot \mathbb{P} \left(\left| \sum_{t=1}^T Z_{t,T,1,\theta} \right| > \varepsilon \right), \end{aligned}$$

where

$$Z_{t,T,1,\theta} = K_{1,h} \left(u - \frac{t}{T} \right) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \cdot \nabla_t - \mathbb{E} \left[K_{1,h} \left(u - \frac{t}{T} \right) W_{t,T} \mathbb{1}_{(|W_{t,T}| \leq \tau_T)} \cdot \nabla_t \right].$$

Observe that $Z_{t,T,1,\theta}$ is an alpha-mixing sequence. Therefore, by selecting the same parameter values for ε , b_T , S_T , and $\sigma_{S_T, T}^2$ as those used for $Z_{t,T,1,k(\theta)}$, and utilizing the corresponding constants C , C' , and M defined therein, we can apply a similar proof strategy as employed for $Q_{3,\theta}$. Specifically, by invoking Lemma B.3, we have

$$\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} |\hat{\psi}_\theta^{(1)}(u, x) - \hat{\psi}_\theta^{(1)}(u, x_{t(x)})| \rightarrow 0.$$

It can also be shown that

$$\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in B_{i,T}} |\hat{\psi}_\theta^{(1)}(u, x_{t(x)}) - \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)})| \rightarrow 0,$$

by employing proof techniques analogous to those used for $Q_{1,\theta}$ and $Q_{3,\theta}$, we obtain the desired result for the α -mixing variable

$$Z_{t,T,2,\theta} = K_{1,h} \left(u - \frac{t}{T} \right) W_{t,T} \mathbb{1}_{\{|W_{t,T}| \leq \tau_T\}} \cdot \Gamma_t - \mathbb{E} \left[K_{1,h} \left(u - \frac{t}{T} \right) W_{t,T} \mathbb{1}_{\{|W_{t,T}| \leq \tau_T\}} \cdot \Gamma_t \right].$$

The parameters are chosen as $\varepsilon = M\alpha_T Th\phi_\theta(h)$ (for a sufficiently large $M > C'$), $b_T = C\tau_T$ (with $C > 0$ sufficiently large), and

$$S_T = \frac{1}{\alpha_T \tau_T}.$$

Here, C' is a constant independent of (u, x) , and the variance satisfies

$$\sigma_{S_T, T}^2 \leq C' S_T h \phi_\theta(h).$$

It remains to establish the convergence of $Q_{4,\theta}$ and $Q_{5,\theta}$. Note that $Q_{4,\theta}$ can be addressed similarly to $Q_{2,\theta}$ since

$$\begin{aligned} \mathbb{E}[Q_{2,\theta}] &\geq \sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in B_{i,T}} \mathbb{E} \left[\left| \hat{\psi}_\theta^{(1)}(u, x_{t(x)}) - \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right| \right] \\ &\geq \sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in B_{i,T}} \left| \mathbb{E} \left[\hat{\psi}_\theta^{(1)}(u, x_{t(x)}) \right] - \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right] \right| \\ &= \sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in B_{i,T}} \left| \left\{ \mathbb{E} \left[\hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right] - \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) \right\} + \left\{ \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_\theta^{(1)}(u, x_{t(x)}) \right] \right\} \right| \\ &\geq \sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_\theta^{(1)}(u, x_{t(x)}) \right] \right| = Q_{4,\theta}. \end{aligned}$$

Thus, we have

$$Q_{4,\theta} = \sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_{k(\theta)}^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_\theta^{(1)}(u, x_{t(x)}) \right] \right| \rightarrow 0.$$

Finally,

$$Q_{5,\theta} = \sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in B_{i,T}} \left| \hat{\psi}_\theta^{(1)}(u, x_{t(x)}) - \mathbb{E} \left[\hat{\psi}_\theta^{(1)}(u, x) \right] \right| \rightarrow 0,$$

since $Q_{5,\theta} \leq \mathbb{E}[Q_{1,\theta}]$ by proving in a similar fashion as $Q_{4,\theta}$. The preceding results together with the result in step (i) complete the proof of Proposition 3.1. \square

Proof of Theorem 3.1. To prove Theorem 3.1, we use a similar decomposition used by [57]. That is, we have

$$\hat{m}_\theta(u, x) - m_\theta(u, x) = \frac{1}{\hat{m}_\theta^{(1)}(u, x)} \left(\hat{g}_\theta^{(1)}(u, x) + \hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x) \hat{m}_\theta^{(1)}(u, x) \right),$$

where

$$\begin{aligned} \hat{m}_\theta^{(1)}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h}(d_\theta(x, X_{t,T})), \\ \hat{g}_\theta^{(1)}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T}, \end{aligned}$$

$$\text{and } \hat{g}_\theta^{(2)}(u, x) = \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) m_\theta\left(\frac{t}{T}, X_{t,T}\right).$$

The proof is completed by showing the following four results:

$$(i) \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} |\hat{g}_\theta^{(1)}(u, x)| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}}\right),$$

$$(ii) \text{ with } \hat{g}_\theta^B(u, x) = \hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x)\hat{m}_\theta^{(1)}(u, x),$$

$$\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [0,1]} \left| \hat{g}_\theta^B(u, x) - \mathbb{E}[\hat{g}_\theta^B(u, x)] \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi_\theta(h)}}\right),$$

(iii) using $\hat{g}_\theta^B(u, x)$ in (ii),

$$\begin{aligned} \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [C_1h, 1-C_1h]} \left| \mathbb{E}[\hat{g}_\theta^B(u, x)] \right| &= \sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [C_1h, 1-C_1h]} \left| \mathbb{E}[\hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x)\hat{m}_\theta^{(1)}(u, x)] \right| \\ &= O(h^2) + O(h^\beta), \end{aligned}$$

$$(iv) \text{ and } \frac{1}{\inf_{\theta \in \Theta} \inf_{x \in S} \inf_{u \in [C_1h, 1-C_1h]} \hat{m}_\theta^{(1)}(u, x)} = O_{\mathbb{P}}(1).$$

Using Proposition 3.1, the first two can be shown by picking $\varepsilon_{i,T}$, and $m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x)$ for $W_{i,T}$, respectively. We now prove (iv). Thus, we have

$$\begin{aligned} \hat{m}_\theta^{(1)}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \{K_{2,h}(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_t^{(t/T)}))\} \\ &\quad + \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_t^{(t/T)})) \\ &=: \bar{m}_\theta^{(1)}(u, x) + \tilde{m}_\theta^{(1)}(u, x). \end{aligned} \tag{A.2}$$

Putting $W_{i,T} = 1$, we get

$$\sup_{\theta \in \Theta} \sup_{x \in S} \sup_{u \in [C_1h, 1-C_1h]} \left| \hat{m}_\theta^{(1)}(u, x) - \mathbb{E}[\hat{m}_\theta^{(1)}(u, x)] \right| = o_{\mathbb{P}}(1),$$

uniformly in u using Proposition 3.1. Furthermore, since $K_2(\cdot)$ is Lipschitz, and by using Definition 2.1 and Lemma B.2, we have

$$\mathbb{E}\left[|\bar{m}_\theta^{(1)}(u, x)|\right] = \mathbb{E}\left[\frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \{K_{2,h}(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_t^{(t/T)}))\}\right]$$

$$\begin{aligned}
&\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \frac{1}{Th} \mathbb{E}\left[U_{t,T}^{(t/T)}\right] \\
&\lesssim \frac{1}{T^2 h^2 \phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \\
&= \frac{1}{Th\phi_\theta(h)} \cdot \frac{1}{Th} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \\
&= o(1),
\end{aligned}$$

where, by the regularity assumption (R2) in Assumption 3.5, we have $\frac{1}{Th\phi_\theta(h)} \rightarrow 0$ as $T \rightarrow \infty$. Consequently, using (A.2), we have

$$\begin{aligned}
\hat{m}_\theta^{(1)}(u, x) &= \hat{m}_\theta^{(1)}(u, x) + \left\{ \mathbb{E}[\hat{m}_\theta^{(1)}(u, x)] - \mathbb{E}[\hat{m}_\theta^{(1)}(u, x)] \right\} \\
&= \left\{ \hat{m}_\theta^{(1)}(u, x) - \mathbb{E}[\hat{m}_\theta^{(1)}(u, x)] \right\} + \left\{ \mathbb{E}[\tilde{m}_\theta^{(1)}(u, x)] + \mathbb{E}[\tilde{m}_\theta^{(1)}(u, x)] \right\} \\
&= o_{\mathbb{P}}(1) + \mathbb{E}[\tilde{m}_\theta^{(1)}(u, x)] + o(1),
\end{aligned} \tag{A.3}$$

uniformly in u . Now, using Assumption 3.1,

$$\begin{aligned}
\mathbb{E}[\tilde{m}_\theta^{(1)}(u, x)] &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}(d_\theta(x, X_t^{(t/T)}))\right] \\
&= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \int_0^h K_{2,h}(y) dF_{t/T}(y; x, \theta) \\
&\gtrsim \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \phi_\theta(h) f_1(x) \sim f_1(x) > 0,
\end{aligned}$$

uniformly in u . Therefore,

$$\frac{1}{\inf_{x \in S} \inf_{\mathcal{H}} \inf_{\theta \in \Theta} \inf_{u \in [C_1 h, 1 - C_1 h]} \hat{m}_\theta^{(1)}(u, x)} = \frac{1}{\inf_{x \in S} \inf_{\mathcal{H}} \inf_{\theta \in \Theta} \inf_{u \in [C_1 h, 1 - C_1 h]} \left\{ O_{\mathbb{P}}(1) + \mathbb{E}[\tilde{m}_\theta^{(1)}(u, x)] + o(1) \right\}} \\
= O_{\mathbb{P}}(1).$$

Hence, the assertion. For the proof of (iii), with support $[0, q]$ for some $q > 1$, we suppose that $K_0 : [0, 1] \rightarrow \mathbb{R}$ is a Lipschitz continuous function such that $K_0(x) = 1$ for all $x \in [0, 1]$. Notice that

$$\mathbb{E}\left[\hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x) \hat{m}_\theta^{(1)}(u, x)\right] = \sum_{i=1}^4 P_{i,\theta}(u, x),$$

where

$$P_{i,\theta}(u, x) = \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) p_{i,\theta}(u, x),$$

such that

$$\begin{aligned}
 p_{1,\theta}(u, x) &= \mathbb{E} \left[K_{0,h}(d_\theta(x, X_{t,T})) \left\{ K_2(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_t^{(t/T)})) \right\} \right. \\
 &\quad \left. \times \left\{ m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right\} \right], \\
 p_{2,\theta}(u, x) &= \mathbb{E} \left[K_{0,h}(d_\theta(x, X_{t,T})) K_{2,h}(d_\theta(x, X_t^{(t/T)})) \left\{ m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta\left(\frac{t}{T}, X_t^{(t/T)}\right) \right\} \right], \\
 p_{3,\theta}(u, x) &= \mathbb{E} \left[\left\{ K_{0,h}(d_\theta(x, X_{t,T})) - K_{0,h}(d_\theta(x, X_t^{(t/T)})) \right\} K_{2,h}(d_\theta(x, X_t^{(t/T)})) \right. \\
 &\quad \left. \times \left\{ m_\theta\left(\frac{t}{T}, X_t^{(t/T)}\right) - m_\theta(u, x) \right\} \right], \\
 \text{and } p_{4,\theta}(u, x) &= \mathbb{E} \left[K_{2,h}(d_\theta(x, X_t^{(t/T)})) \left\{ m_\theta\left(\frac{t}{T}, X_t^{(t/T)}\right) - m_\theta(u, x) \right\} \right].
 \end{aligned}$$

Now, observe that

$$\begin{aligned}
 P_{1,\theta}(u, x) &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \mathbb{E} \left[K_{0,h}(d_\theta(x, X_{t,T})) \right. \\
 &\quad \left. \times \left\{ K_{2,h}(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_t^{(t/T)})) \right\} \left\{ m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right\} \right] \\
 &\leq \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \mathbb{E} \left[\left| K_{2,h}(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_t^{(t/T)})) \right| \right. \\
 &\quad \left. \times K_{0,h}(d_\theta(x, X_{t,T})) \left| m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right| \right].
 \end{aligned}$$

Now, we obtain the bound $h^{1 \wedge \beta}$ using Assumption 3.1. That is,

$$\begin{aligned}
 K_{0,h}(d_\theta(x, X_{t,T})) \left| m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right| &\lesssim K_{0,h}(d_\theta(x, X_{t,T})) \left(d_\theta(x, X_{t,T}) + \left| \frac{t}{T} - u \right| \right)^\beta \\
 &\lesssim h^{1 \wedge \beta}.
 \end{aligned}$$

Furthermore, since $K_2(\cdot)$ is Lipschitz continuous from Assumption 3.3 and that

$$d_\theta(X_{t,T}, X_t^{(t/T)}) \leq \frac{1}{T} U_{t,T}^{(t/T)},$$

as previously shown, we have

$$\begin{aligned}
 P_{1,\theta}(u, x) &\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \mathbb{E} \left[\left| K_{2,h}(d_\theta(x, X_{t,T})) - K_{2,h}(d_\theta(x, X_t^{(t/T)})) \right| \right] \times h^{1 \wedge \beta} \\
 &\lesssim \frac{1}{Th^{1-(1 \wedge \beta)}\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \mathbb{E} \left[\left| \frac{1}{Th} U_{t,T}^{(t/T)} \right| \right]
 \end{aligned}$$

$$\lesssim \frac{1}{Th^{1-(1\wedge\beta)}\phi_\theta(h)},$$

uniformly in u . The cases for $P_{2,\theta}(u, x)$ and $P_{3,\theta}(u, x)$ can be proved in a similar fashion as $P_{1,\theta}(u, x)$. Hence, we have

$$\sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in [C_1h, 1-C_1h]} |P_{2,\theta}(u, x)| \lesssim \frac{1}{Th^{1-(1\wedge\beta)}\phi_\theta(h)},$$

and

$$\sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in [C_1h, 1-C_1h]} |P_{3,\theta}(u, x)| \lesssim \frac{1}{Th^{1-(1\wedge\beta)}\phi_\theta(h)}.$$

Finally, leveraging Assumption 3.1 and Lemma B.1, we derive

$$\begin{aligned} |P_{4,\theta}(u, x)| &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \times \mathbb{E} \left[K_{2,h}\left(d_\theta(x, X_t^{(t/T)})\right) \left\{ m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right\} \right] \\ &\leq \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T \left| K_{1,h}\left(u - \frac{t}{T}\right) \right| \times \mathbb{E} \left[\left| K_{2,h}\left(d_\theta(x, X_t^{(t/T)})\right) \right| \left| m_\theta\left(\frac{t}{T}, X_t^{(t/T)}\right) - m_\theta(u, x) \right| \right] \\ &\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T \left| K_{1,h}\left(u - \frac{t}{T}\right) \right| \times \mathbb{E} \left[\left| K_{2,h}\left(d_\theta(x, X_t^{(t/T)})\right) \right| \left(d_\theta(x, X_t^{(t/T)}) + \left| \frac{t}{T} - u \right| \right)^\beta \right] \\ &\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T \left| K_{1,h}\left(u - \frac{t}{T}\right) - \int_0^1 \frac{1}{h} K_{1,h}(u-v) dv \right| \times \mathbb{E} \left[\left| K_{2,h}\left(d_\theta(x, X_t^{(t/T)})\right) \right| \right] \\ &\quad \times h^\beta + \frac{1}{Th\phi_\theta(h)} \cdot \sum_{t=1}^T \int_0^1 \frac{1}{h} K_{1,h}(u-v) dv \times \mathbb{E} \left[\left| K_{2,h}\left(d_\theta(x, X_t^{(t/T)})\right) \right| \right] \times h^\beta \\ &\lesssim O\left(\frac{1}{Th^2}\right) \cdot h^\beta + h^\beta. \end{aligned}$$

However, we have

$$\frac{1}{Th^2} \cdot h^\beta \leq \frac{1}{T} \cdot h^{\beta-2} \lesssim \phi_\theta(h) \cdot h^2 \ll h^2.$$

We then infer that

$$\sup_{x \in \mathcal{S}} \sup_{\theta \in \Theta} \sup_{u \in [C_1h, 1-C_1h]} |Q_{4,\theta}(u, x)| \ll h^2 + h^\beta.$$

Hence, we obtain

$$\sup_{\theta \in \Theta} \sup_{x \in \mathcal{S}} \sup_{u \in [C_1h, 1-C_1h]} \left| \mathbb{E} \left[\hat{g}_\theta^B(u, x) \right] \right| = O(h^2) + O(h^\beta). \quad (\text{A.4})$$

Finally, given our assumptions, our proof is complete with an approximation error of

$$O\left(\frac{1}{Th^{1-(1\wedge\beta)}\phi_\theta(h)}\right) \ll h^{2\wedge\beta}.$$

Hence the proof is complete. \square

Proof of Theorem 3.2. We begin our proof for this theorem by showing

$$\hat{g}_\theta^B(u, x) - \mathbb{E}[\hat{g}_\theta^B(u, x)] = o_{\mathbb{P}}\left(\sqrt{\frac{1}{Th\phi_\theta(h)}}\right),$$

that is, from

$$\hat{m}_\theta(u, x) - m_\theta(u, x) = \frac{1}{\hat{m}_\theta^{(1)}(u, x)} \left(\hat{g}_\theta^{(1)}(u, x) + \hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x)\hat{m}_\theta^{(1)}(u, x) \right),$$

we define

$$\hat{g}_\theta^B(u, x) = \hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x)\hat{m}_\theta^{(1)}(u, x).$$

Now, we infer

$$\begin{aligned} \hat{g}_\theta^B(u, x) &= \hat{g}_\theta^{(2)}(u, x) - m_\theta(u, x)\hat{m}_\theta^{(1)}(u, x) \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) m_\theta\left(\frac{t}{T}, X_{t,T}\right) \\ &\quad - \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) m_\theta(u, x) \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) \left\{ m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right\} \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \Lambda_{t,T,\theta}(u, x), \end{aligned}$$

where

$$\Lambda_{t,T,\theta}(u, x) = K_{2,h}(d_\theta(x, X_{t,T})) \left\{ m_\theta\left(\frac{t}{T}, X_{t,T}\right) - m_\theta(u, x) \right\}.$$

Then, we readily obtain

$$\begin{aligned} \text{Var}\left(\hat{g}_\theta^B(u, x)\right) &= \text{Var}\left(\frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \Lambda_{t,T,\theta}(u, x)\right) \\ &= \frac{1}{(Th\phi_\theta(h))^2} \text{Var}\left(\sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) \Lambda_{t,T,\theta}(u, x)\right) \\ &= \frac{1}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \cdot \text{Var}\left(\Lambda_{t,T,\theta}(u, x)\right) \\ &\quad + \frac{1}{(Th\phi_\theta(h))^2} \sum_{\substack{t_1, t_2=1 \\ t_1 \neq t_2}}^T K_{1,h}^2\left(u - \frac{t_1}{T}\right) K_{1,h}^2\left(u - \frac{t_2}{T}\right) \text{Cov}\left(\Lambda_{t_1,T,\theta}(u, x), \Lambda_{t_2,T,\theta}(u, x)\right) \\ &=: V_{1,T,\theta}^B + V_{2,T,\theta}^B. \end{aligned}$$

We now discuss $V_{1,T,\theta}^B$ by first noting that

$$\begin{aligned}
 \mathbb{E} \left[K_{2,h}^2 \left(d_\theta(x, X_{t,T}^{(t/T)}) \right) \right] &= \int_0^h K_{2,h}^2(y) \mathbb{P}(dy) \\
 &= -\frac{2}{h} \int_0^h K_{2,h}(y) K'_{2,h}(y) F_{t/T}(y; x, \theta) dy \\
 &\sim -\frac{2}{h} \int_0^h K_{2,h}(y) K'_{2,h}(y) \phi_\theta(y) dy \\
 &= \frac{2}{h} \int_0^h \left(1 - \frac{y}{h} \right) \phi_\theta(y) dy \\
 &= \frac{2}{h^2} \int_0^h \left(\int_0^y \phi_\theta(\varepsilon) d\varepsilon \right) dy \\
 &\sim \frac{2}{h^2} \int_0^h y \phi_\theta(y) dy \sim \phi_\theta(h),
 \end{aligned} \tag{A.5}$$

using integration by parts, change of variables, and the model and kernel assumptions in Assumptions 3.1 and 3.3, respectively, and using the fact that $K_{2,h}(x) = (1-x)\mathbb{1}_{x \in [0,1]}$ is an asymmetrical triangle kernel. On the other hand, since $K_2(\cdot)$ is Lipschitz and by Definition 2.2, we have

$$\begin{aligned}
 \mathbb{E} \left[K_{2,h}^2 \left(d_\theta(x, X_{t,T}) \right) - K_{2,h}^2 \left(d_\theta(x, X_{t,T}^{(t/T)}) \right) \right] &\leq \mathbb{E} \left[K_{2,h} \left(d_\theta(x, X_{t,T}) \right) - K_{2,h} \left(d_\theta(x, X_{t,T}^{(t/T)}) \right) \right] \\
 &\leq \frac{C}{h} \mathbb{E} \left| d_\theta(x, X_{t,T}) - d_\theta(x, X_{t,T}^{(t/T)}) \right| \\
 &\lesssim \frac{1}{Th} \mathbb{E} \left[U_{t,T}^{(t/T)} \right] \lesssim \frac{1}{Th}.
 \end{aligned} \tag{A.6}$$

Moreover, by the smoothness of m_θ in Assumption 3.1, we get

$$\left| m_\theta \left(\frac{t}{T}, X_{t,T} \right) - m_\theta(u, x) \right|^2 \lesssim h^{2\beta}.$$

Therefore, it follows that

$$\begin{aligned}
 |V_{1,T,\theta}^B| &= \frac{1}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2 \left(u - \frac{t}{T} \right) \cdot \text{Var} \left(\Lambda_{t,T,\theta}(u, x) \right) \\
 &= \frac{1}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2 \left(u - \frac{t}{T} \right) \cdot \left\{ \mathbb{E} \left[\left(\Lambda_{t,T,\theta}^2(u, x) \right) \right] - \left(\mathbb{E} \left[\Lambda_{t,T,\theta}(u, x) \right] \right)^2 \right\} \\
 &\leq \frac{1}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2 \left(u - \frac{t}{T} \right) \cdot \mathbb{E} \left[\left(\Lambda_{t,T,\theta}^2(u, x) \right) \right] \\
 &\leq \frac{1}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2 \left(u - \frac{t}{T} \right) \times \mathbb{E} \left[K_{2,h}^2 \left(d_\theta(x, X_{t,T}) \right) \cdot \left(m_\theta \left(\frac{t}{T}, X_{t,T} \right) - m_\theta(u, x) \right)^2 \right] \\
 &\lesssim \frac{h^{2\beta}}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2 \left(u - \frac{t}{T} \right)
 \end{aligned}$$

$$\begin{aligned}
& \times \left\{ \mathbb{E} \left[K_{2,h}^2(d_\theta(x, X_{t,T})) - K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)})) \right] + \mathbb{E} \left[K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)})) \right] \right\} \\
& \lesssim \frac{h^{2\beta}}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \cdot \left(\frac{1}{Th} + \phi_\theta(h)\right) \\
& \lesssim \frac{h^{2\beta}\phi_\theta(h)}{(Th\phi_\theta(h))^2} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \\
& \ll \frac{1}{Th\phi_\theta(h)}. \tag{A.7}
\end{aligned}$$

We now consider $V_{2,T,\theta}^B$ and write

$$V_{2,T,\theta}^B = V_{2,1,T,\theta}^B + V_{2,2,T,\theta}^B, \tag{A.8}$$

where

$$\begin{aligned}
V_{2,1,T,\theta}^B &= \frac{1}{(Th\phi_\theta(h))^2} \sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| \leq \lambda_T}}^T K_{1,h}\left(u - \frac{t_1}{T}\right) K_{1,h}\left(u - \frac{t_2}{T}\right) \times \text{Cov}\left(\Lambda_{t_1, T, \theta}(u, x), \Lambda_{t_2, T, \theta}(u, x)\right), \\
\text{and } V_{2,2,T,\theta}^B &= \frac{1}{(Th\phi_\theta(h))^2} \sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| > \lambda_T}}^T K_{1,h}\left(u - \frac{t_1}{T}\right) K_{1,h}\left(u - \frac{t_2}{T}\right) \times \text{Cov}\left(\Lambda_{t_1, T, \theta}(u, x), \Lambda_{t_2, T, \theta}(u, x)\right).
\end{aligned}$$

Observe that from our previous calculations, we can deduce

$$\begin{aligned}
\mathbb{E}[\Lambda_{t_1, T, \theta}(u, x)] \cdot \mathbb{E}[\Lambda_{t_2, T, \theta}(u, x)] &= \mathbb{E} \left[K_{2,h}(d_\theta(x, X_{t_1, T})) \left\{ m_\theta\left(\frac{t}{T}, X_{t_1, T}\right) - m_\theta(u, x) \right\} \right] \\
&\quad \times \mathbb{E} \left[K_{2,h}(d_\theta(x, X_{t_2, T})) \left\{ m_\theta\left(\frac{t}{T}, X_{t_2, T}\right) - m_\theta(u, x) \right\} \right] \\
&\lesssim h^{2\beta} \cdot \phi_\theta^2(h),
\end{aligned}$$

and that by Assumption 3.1,

$$\mathbb{E}[\Lambda_{t_1, T, \theta}(u, x) \cdot \Lambda_{t_2, T, \theta}(u, x)] \leq \psi_\theta(h) f_2(x).$$

Therefore, $V_{2,1,T,\theta}^B$ becomes

$$\begin{aligned}
|V_{2,1,T,\theta}^B| &\leq \frac{1}{(Th\phi_\theta(h))^2} \sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| \leq \lambda_T}}^T K_{1,h}\left(u - \frac{t_1}{T}\right) K_{1,h}\left(u - \frac{t_2}{T}\right) \\
&\quad \times \left\{ \mathbb{E}[\Lambda_{t_1, T, \theta}(u, x) \cdot \Lambda_{t_2, T, \theta}(u, x)] + \mathbb{E}[\Lambda_{t_1, T, \theta}(u, x)] \cdot \mathbb{E}[\Lambda_{t_2, T, \theta}(u, x)] \right\} \\
&\lesssim \frac{h^{2(1 \wedge \beta)}}{(Th\phi_\theta(h))^2} \sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| \leq \lambda_T}}^T K_{1,h}\left(u - \frac{t_1}{T}\right) K_{1,h}\left(u - \frac{t_2}{T}\right) \left\{ \psi_\theta(h) + \phi_\theta^2(h) \right\} \\
&\lesssim \frac{h^{2(1 \wedge \beta)}}{(Th\phi_\theta(h))^2} \cdot T \lambda_T \cdot \left(\psi_\theta(h) + \phi_\theta^2(h) \right)
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{h^{2(1 \wedge \beta)-1} \lambda_T}{Th} \cdot \frac{\psi_\theta(h)}{\phi_\theta^2(h)} \right) + \left(\frac{h^{2(1 \wedge \beta)-1} \lambda_T}{Th} \right) \\
&\lesssim \frac{1}{Th\phi_\theta(h)} \cdot h^{2(1 \wedge \beta)-1} \phi_\theta(h) \lambda_T,
\end{aligned}$$

where by Assumption 3.1, $\frac{\psi_\theta(h)}{\phi_\theta^2(h)} < \infty$, and

$$\begin{aligned}
h^{2(1 \wedge \beta)-1} \lambda_T \cdot \frac{1}{Th} &= h^{2(1 \wedge \beta)-1} \lambda_T \cdot \frac{1}{Th} \cdot \frac{h^2}{h^2} \\
&= h^{2(1 \wedge \beta)+1} \lambda_T \cdot \frac{1}{Th^3} \rightarrow 0,
\end{aligned}$$

since $Th^3 \rightarrow \infty$ by Assumption 3.5. Here, the goal is to pick λ_T such that $|V_{2,1,T,\theta}^B| \rightarrow 0$ as $T \rightarrow \infty$. Thus, using Lemma B.4, and the fact that $K_2(\cdot)$ is bounded as in Assumption 3.1, we have

$$\begin{aligned}
&\text{Cov}(\Lambda_{t_1,T,\theta}(u, x), \Lambda_{t_2,T,\theta}(u, x)) \\
&\leq \mathbb{E}[\Lambda_{t_1,T,\theta}(u, x) \cdot \Lambda_{t_2,T,\theta}(u, x)] - \mathbb{E}[\Lambda_{t_1,T,\theta}(u, x)] \cdot \mathbb{E}[\Lambda_{t_2,T,\theta}(u, x)] \\
&\lesssim \|\Lambda_{t_1,T,\theta}(u, x)\|_v \cdot \|\Lambda_{t_2,T,\theta}(u, x)\|_v \cdot \alpha(|t_1 - t_2|)^{1-\frac{1}{v}-\frac{1}{v}} \\
&= \mathbb{E}[(\Lambda_{t_1,T,\theta}(u, x))^v]^{\frac{1}{v}} \mathbb{E}[(\Lambda_{t_2,T,\theta}(u, x))^v]^{\frac{1}{v}} \cdot \alpha(|t_1 - t_2|)^{1-\frac{2}{v}} \\
&= \mathbb{E}\left[\left\{K_{2,h}(d_\theta(x, X_{t_1,T})) \cdot \left(m_\theta\left(\frac{t_1}{T}, X_{t_1,T}\right) - m_\theta(u, x)\right)\right\}^v\right]^{\frac{1}{v}} \\
&\quad \times \mathbb{E}\left[\left\{K_{2,h}(d_\theta(x, X_{t_2,T})) \cdot \left(m_\theta\left(\frac{t_2}{T}, X_{t_2,T}\right) - m_\theta(u, x)\right)\right\}^v\right]^{\frac{1}{v}} \cdot \alpha(|t_1 - t_2|)^{1-\frac{2}{v}} \\
&\lesssim h^{2(1 \wedge \beta)} \cdot \mathbb{E}\left[K_{2,h}(d_\theta(x, X_{t_1,T}))^v\right]^{\frac{1}{v}} \mathbb{E}\left[K_{2,h}(d_\theta(x, X_{t_2,T}))^v\right]^{\frac{1}{v}} \cdot \alpha(|t_1 - t_2|)^{1-\frac{2}{v}} \\
&\leq h^{2(1 \wedge \beta)} \cdot \mathbb{E}\left[K_{2,h}(d_\theta(x, X_{t_1,T}))^2\right]^{\frac{1}{v}} \mathbb{E}\left[K_{2,h}(d_\theta(x, X_{t_2,T}))^2\right]^{\frac{1}{v}} \cdot \alpha(|t_1 - t_2|)^{1-\frac{2}{v}} \\
&\lesssim h^{2(1 \wedge \beta)} \cdot (\phi_\theta(h))^{\frac{2}{v}} \cdot (\alpha(k))^{1-\frac{2}{v}}.
\end{aligned}$$

For $V_{2,2,T,\theta}^B$, we use the preceding calculations and obtain

$$\begin{aligned}
V_{2,2,T,\theta}^B &= \frac{1}{(Th\phi_\theta(h))^2} \sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| > \lambda_T}}^T K_{1,h}\left(u - \frac{t_1}{T}\right) K_{1,h}\left(u - \frac{t_2}{T}\right) \times \text{Cov}(\Lambda_{t_1,T,\theta}(u, x), \Lambda_{t_2,T,\theta}(u, x)) \\
&\lesssim \frac{1}{(Th\phi_\theta(h))^2} \cdot h^{2(1 \wedge \beta)} (\phi_\theta(h))^{\frac{2}{v}} \left(\sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| > \lambda_T}}^T \alpha(|t_1 - t_2|)^{1-\frac{2}{v}} \right) \\
&\leq \frac{1}{Th\phi_\theta(h)} \cdot \frac{h^{2(1 \wedge \beta)-1}}{(\phi_\theta(h))^{1-\frac{2}{v}}} \left(\sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| > \lambda_T}}^T \alpha(|t_1 - t_2|)^{1-\frac{2}{v}} \right) \\
&\leq \frac{1}{Th\phi_\theta(h)} \cdot \frac{h^{2(1 \wedge \beta)-1}}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{v}}} \left(\sum_{k=\lambda_T+1}^{\infty} k^\delta \alpha(k)^{1-\frac{2}{v}} \right)
\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{Th\phi_\theta(h)} \cdot \frac{h^{2(1\wedge\beta)-1}}{(\phi_\theta(h))^{1-\frac{2}{v}}} \left(\sum_{\substack{t_1, t_2=1 \\ 1 \leq |t_1 - t_2| > \lambda_T}}^T \alpha (|t_1 - t_2|)^{1-\frac{2}{v}} \right) \\ &\leq \frac{1}{Th\phi_\theta(h)} \cdot \frac{h^{2(1\wedge\beta)-1}}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{v}}} \left(\sum_{k=\lambda_T+1}^\infty k^\delta \alpha(k)^{1-\frac{2}{v}} \right). \end{aligned}$$

As previously mentioned, we pick λ_T that makes our expression approach to zero. That happens if we choose

$$\lambda_T = \left\lfloor (\phi_\theta(h))^{-\frac{1-(2/v)}{\delta}} \right\rfloor,$$

and hence the use of Assumption 3.4. Therefore, the results from (A.7) and (A.8) suggest that

$$\text{Var}(\hat{g}_\theta^B(u, x)) < |V_{1,T,\theta}^B| + |V_{2,T,\theta}^B| = o\left(\frac{1}{Th\phi_\theta(h)}\right).$$

Consequently,

$$\hat{g}_\theta^B(u, x) - \mathbb{E}[\hat{g}_\theta^B(u, x)] = o_{\mathbb{P}}\left(\sqrt{\frac{1}{Th\phi_\theta(h)}}\right),$$

as we desired. Thus, the above results together with (A.3), (A.4), and the fact that $\lim_{T \rightarrow \infty} \mathbb{E}[\hat{m}_\theta^{(1)}(u, x)] > 0$ give us

$$\hat{m}_\theta(u, x) - m_\theta(u, x) = \frac{\hat{g}_\theta^{(1)}(u, x)}{\hat{m}_\theta^{(1)}(u, x)} + B_{T,\theta}(u, x) + o_{\mathbb{P}}\left(\sqrt{\frac{1}{Th\phi_\theta(h)}}\right).$$

We now work with $Th\phi_\theta(h) \text{Var}(\hat{g}_\theta^{(1)}(u, x))$ which concerns the first part of the previous equation. In this part of the proof, we will show that

$$Th\phi_\theta(h) \text{Var}(\hat{g}_\theta^{(1)}(u, x)) \sim \frac{\mathbb{E}[\varepsilon_T^2] \sigma^2(u, \langle \theta, x \rangle)}{Th} \int K_1^2(w) dw. \quad (\text{A.9})$$

Here, we utilize the Assumption 3.2, and the result in (A.5). Also, consider a sequence of independent random variables $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, independent of $\{X_{t,T}\}_{t=1}^T$. Thus, we have

$$\begin{aligned} Th\phi_\theta(h) \text{Var}(\hat{g}_\theta^{(1)}(u, x)) &= Th\phi_\theta(h) \text{Var}\left(\frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T}\right) \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[\left\{K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T}\right\}^2\right] \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T})) \varepsilon_{t,T}^2\right] \\ &= \frac{1}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T})) \cdot \sigma^2\left(\frac{t}{T}, \langle \theta, X_{t,T} \rangle\right) \varepsilon_t^2\right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\mathbb{E}[\varepsilon_t^2]}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T})) \cdot \sigma^2\left(\frac{t}{T}, \langle\theta, X_{t,T}\rangle\right)\right] \\
&= \frac{\mathbb{E}[\varepsilon_t^2]}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T}))\right] \cdot \{\sigma^2(u, \langle\theta, x\rangle) + o(1)\} \\
&= \frac{\mathbb{E}[\varepsilon_t^2]}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T})) - K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)}))\right] \\
&\quad \times \{\sigma^2(u, \langle\theta, x\rangle) + o(1)\} + \frac{\mathbb{E}[\varepsilon_t^2]}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \\
&\quad \times \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)}))\right] \cdot \{\sigma^2(u, \langle\theta, x\rangle) + o(1)\} \\
&= \frac{\mathbb{E}[\varepsilon_t^2]}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \cdot o(\phi_\theta(h)) \cdot \{\sigma^2(u, \langle\theta, x\rangle) + o(1)\} \\
&\quad + \frac{\mathbb{E}[\varepsilon_t^2]}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)}))\right] \cdot \{\sigma^2(u, \langle\theta, x\rangle) + o(1)\} \\
&= \frac{\mathbb{E}[\varepsilon_t^2]o(\phi_\theta(h))(\sigma^2(u, \langle\theta, x\rangle) + o(1))}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \\
&\quad + \frac{\mathbb{E}[\varepsilon_t^2](\sigma^2(u, \langle\theta, x\rangle) + o(1))}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)}))\right] \\
&= \frac{\mathbb{E}[\varepsilon_t^2](\sigma^2(u, \langle\theta, x\rangle) + o(1))}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \mathbb{E}\left[K_{2,h}^2(d_\theta(x, X_{t,T}^{(t/T)}))\right] + o(1) \\
&\sim \frac{\mathbb{E}[\varepsilon_t^2]\phi_\theta(h)(\sigma^2(u, \langle\theta, x\rangle) + o(1))}{Th\phi_\theta(h)} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) + o(1) \\
&\sim \frac{\mathbb{E}[\varepsilon_t^2]\sigma^2(u, \langle\theta, x\rangle)}{Th} \sum_{t=1}^T K_{1,h}^2\left(u - \frac{t}{T}\right) \\
&\sim \frac{\mathbb{E}[\varepsilon_t^2]\sigma^2(u, \langle\theta, x\rangle)}{Th} \int K_1^2(w)dw.
\end{aligned}$$

The final part of the proof covers the asymptotic normality of $\tilde{g}_\theta^{(1)}(u, x)$ which is derived from

$$Th\phi_\theta(h)\hat{g}_\theta^{(1)}(u, x) = \sqrt{Th\phi_\theta(h)}\tilde{g}_\theta^{(1)}(u, x),$$

where

$$\tilde{g}_\theta^{(1)}(u, x) = \sqrt{Th\phi_\theta(h)}\hat{g}_\theta^{(1)}(u, x).$$

That is, we want to show that

$$\tilde{g}_\theta^{(1)}(u, x) \xrightarrow{d} N(0, V_\theta(u, x)) \text{ as } T \rightarrow \infty.$$

To demonstrate the distributional convergence of $\tilde{g}_\theta^{(1)}(u, x)$, we will employ Bernstein's big-block and small-block method. We begin by partitioning the index set $\{1, \dots, T\}$ into $2k_T + 1$ subsets, consisting of large blocks of size a_T and small blocks of size v_T , such that

$$k_T := \left\lfloor \frac{T}{a_T + v_T} \right\rfloor, \quad \frac{v_T}{a_T} \rightarrow 0, \quad \frac{a_T}{T} \rightarrow 0, \quad \text{and} \quad \frac{T}{a_T} \alpha(v_T) \rightarrow 0. \quad (\text{A.10})$$

Using Assumption 3.6, there exists a sequence of positive integers $\{q_T\}$, $q_T \rightarrow \infty$, such that

$$q_T v_T = o\left(\sqrt{Th\phi_\theta(h)}\right), \quad q_T \sqrt{\frac{T}{h\phi_\theta(h)}} \alpha(v_T) \rightarrow 0 \text{ as } T \rightarrow \infty.$$

Consequently, setting $a_T = \left\lfloor \frac{\sqrt{Th\phi_\theta(h)}}{q_T} \right\rfloor$, we have

$$\frac{a_T}{\sqrt{Th\phi_\theta(h)}} \rightarrow 0, \quad \frac{T}{a_T} \alpha(v_T) \rightarrow 0 \text{ as } T \rightarrow \infty.$$

Thus, the decomposition of $\tilde{g}_\theta^{(1)}(u, x)$ would be

$$\begin{aligned} \tilde{g}_\theta^{(1)}(u, x) &= \frac{1}{\sqrt{Th\phi_\theta(h)}} \sum_{j=1}^{k_T} \eta_j(u, x; \theta) + \frac{1}{\sqrt{Th\phi_\theta(h)}} \sum_{j=1}^{k_T} \xi_j(u, x; \theta) + \zeta(u, x; \theta) \\ &=: \tilde{g}_\theta^{(11)}(u, x) + \tilde{g}_\theta^{(12)}(u, x) + \tilde{g}_\theta^{(13)}(u, x), \end{aligned} \quad (\text{A.11})$$

where

$$\begin{aligned} \eta_j(u, x; \theta) &= \sum_{t=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}\left(d_\theta(x, X_{t,T})\right) \varepsilon_{t,T}, \\ \xi_j(u, x; \theta) &= \sum_{t=ja_T+(j-1)v_T+1}^{j(a_T+v_T)} K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}\left(d_\theta(x, X_{t,T})\right) \varepsilon_{t,T}, \\ \text{and } \zeta(u, x; \theta) &= \sum_{t=k_T(a_T+v_T)+1}^T K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}\left(d_\theta(x, X_{t,T})\right) \varepsilon_{t,T}. \end{aligned}$$

First, we show that the summation over the small-blocks ($\tilde{g}_\theta^{(12)}(u, x)$), and over the remainder-blocks ($\tilde{g}_\theta^{(13)}(u, x)$) are asymptotically negligible. That is, as $T \rightarrow \infty$,

$$\mathbb{E} \left[\tilde{g}_\theta^{(12)}(u, x) \right]^2 \rightarrow 0, \quad \text{and} \quad \mathbb{E} \left[\tilde{g}_\theta^{(13)}(u, x) \right]^2 \rightarrow 0.$$

We first prove

$$\mathbb{E} \left[\tilde{g}_\theta^{(12)}(u, x) \right]^2 \rightarrow 0.$$

We have

$$\mathbb{E} \left[\tilde{g}_\theta^{(12)}(u, x) \right]^2 = \text{Var} \left(\frac{1}{\sqrt{Th\phi_\theta(h)}} \sum_{j=1}^{k_T} \xi_j(u, x; \theta) \right)$$

$$\begin{aligned}
&= \frac{1}{Th\phi_\theta(h)} \text{Var} \left(\sum_{j=1}^{k_T} \xi_j(u, x; \theta) \right) \\
&= \frac{1}{Th\phi_\theta(h)} \sum_{j=1}^{k_T} \text{Var}(\xi_j(u, x; \theta)) + \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ i \neq j}}^{k_T} \text{Cov}(\xi_i(u, x; \theta), \xi_j(u, x; \theta)) \\
&=: F_1 + F_2.
\end{aligned}$$

Dealing F_1 , we have

$$\begin{aligned}
\text{Var}(\xi_j(u, x; \theta)) &= \text{Var} \left(\sum_{t=ja_T+(j-1)v_T+1}^{j(a_T+v_T)} K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&= \sum_{t=ja_T+(j-1)v_T+1}^{j(a_T+v_T)} K_{1,h}^2 \left(u - \frac{t}{T} \right) \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&\lesssim v_T \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right).
\end{aligned}$$

We have

$$\begin{aligned}
F_1 &= \frac{1}{Th\phi_\theta(h)} \sum_{j=1}^{k_T} \text{Var}(\xi_j(u, x; \theta)) \\
&\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{j=1}^{k_T} v_T \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&\lesssim \frac{1}{Th\phi_\theta(h)} k_T v_T \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&\sim k_T v_T \cdot \left\{ \frac{\mathbb{E}[\varepsilon_t^2] \sigma^2(u, x)}{Th} \left(\int_{[0,h]} K_1^2(w) dw \right) \right\} \\
&\sim \frac{T}{a_T + v_T} \cdot v_T \sim \frac{T v_T}{a_T} \\
&= o_p(T),
\end{aligned}$$

using the results in (A.9) and (A.10). We next handle F_2 . That is, with $\Delta'_i(u, x; \theta) = K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h}(d_\theta(x, X_{t,T}))$,

$$\begin{aligned}
\text{Cov}(\xi_i(u, x; \theta), \xi_j(u, x; \theta)) &= \sum_{k=ia_T+(i-1)v_T+1}^{i(a_T+v_T)} \sum_{k'=ja_T+(j-1)v_T+1}^{j(a_T+v_T)} \text{Cov}(\Delta'_k(u, x; \theta) \varepsilon_{k,T}, \Delta'_{k'}(u, x; \theta) \varepsilon_{k',T}) \\
&= \sum_{l_1=1}^{v_T} \sum_{l_2=1}^{v_T} \text{Cov}(\Delta'_{\lambda_k+l_1}(u, x; \theta) \varepsilon_{\lambda_k+l_1,T}, \Delta'_{\lambda_{k'}+l_2}(u, x; \theta) \varepsilon_{\lambda_{k'}+l_2,T}),
\end{aligned}$$

where $\lambda_k = ja_T + (j-1)v_T$. Since $i \neq j$, $|\lambda_i - \lambda_j + l_1 - l_2| \geq v_T$, we have

$$|F_2| = \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ i \neq j}}^{k_T} \text{Cov}(\xi_i(u, x; \theta), \xi_j(u, x; \theta))$$

$$\leq \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ |i-j|\geq v_T}}^{k_T} \text{Cov}(\Delta'_i(u, x; \theta)\varepsilon_{i,T}, \Delta'_j(u, x; \theta)\varepsilon_{j,T}).$$

Using Davydov's lemma (see Lemma B.4), Assumption 3.3, and the results in (A.5) and (A.6), we get

$$\begin{aligned} & \text{Cov}(\Delta'_i(u, x; \theta)\varepsilon_{i,T}, \Delta'_j(u, x; \theta)\varepsilon_{j,T}) \\ &= \mathbb{E}[(\Delta'_i(u, x; \theta)\varepsilon_{i,T}) \cdot (\Delta'_j(u, x; \theta)\varepsilon_{j,T})] - \mathbb{E}[\Delta'_i(u, x; \theta)\varepsilon_{i,T}] \mathbb{E}[\Delta'_j(u, x; \theta)\varepsilon_{j,T}] \\ &\leq 8 \|\Delta'_i(u, x; \theta)\varepsilon_{i,T}\|_{\nu} \|\Delta'_j(u, x; \theta)\varepsilon_{j,T}\|_{\nu} \cdot (\alpha(|i-j|))^{1-\frac{1}{\nu}-\frac{1}{\nu}} \\ &= 8 \left(\mathbb{E}|\Delta'_i(u, x; \theta)\varepsilon_{i,T}|^{\nu} \right)^{\frac{1}{\nu}} \left(\mathbb{E}|\Delta'_j(u, x; \theta)\varepsilon_{j,T}|^{\nu} \right)^{\frac{1}{\nu}} \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\ &\lesssim \left(\mathbb{E} \left| K_{1,h} \left(u - \frac{i}{T} \right) K_{2,h}(d_\theta(x, X_{i,T})) \right|^{\nu} \right)^{\frac{1}{\nu}} \mathbb{E}[|\varepsilon_{i,T}|^{\nu}]^{\frac{1}{\nu}} \\ &\quad \times \left(\mathbb{E} \left| K_{1,h} \left(u - \frac{j}{T} \right) K_{2,h}(d_\theta(x, X_{j,T})) \right|^{\nu} \right)^{\frac{1}{\nu}} \mathbb{E}[|\varepsilon_{j,T}|^{\nu}]^{\frac{1}{\nu}} \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\ &\leq \left(\mathbb{E} \left| K_{1,h} \left(u - \frac{i}{T} \right) K_{2,h}(d_\theta(x, X_{i,T})) \right|^2 \right)^{\frac{1}{\nu}} \mathbb{E}[|\varepsilon_{i,T}|^2]^{\frac{1}{\nu}} \\ &\quad \times \left(\mathbb{E} \left| K_{1,h} \left(u - \frac{j}{T} \right) K_{2,h}(d_\theta(x, X_{j,T})) \right|^2 \right)^{\frac{1}{\nu}} \mathbb{E}[|\varepsilon_{j,T}|^2]^{\frac{1}{\nu}} \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\ &\leq \left(\mathbb{E} \left| K_{1,h} \left(u - \frac{i}{T} \right) \right|^2 \right)^{\frac{1}{\nu}} \left(\mathbb{E} \left| K_{2,h}(d_\theta(x, X_{i,T})) \right|^2 \right)^{\frac{1}{\nu}} \mathbb{E}[|\varepsilon_{i,T}|^2]^{\frac{1}{\nu}} \\ &\quad \times \left(\mathbb{E} \left| K_{1,h} \left(u - \frac{j}{T} \right) \right|^2 \right)^{\frac{1}{\nu}} \left(\mathbb{E} \left| K_{2,h}(d_\theta(x, X_{j,T})) \right|^2 \right)^{\frac{1}{\nu}} \mathbb{E}[|\varepsilon_{j,T}|^2]^{\frac{1}{\nu}} \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\ &\lesssim \phi_\theta^2(h) \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}}. \end{aligned}$$

Therefore, with

$$\lambda_T = \left[(\phi_\theta(h))^{-\frac{1-(2/\nu)}{\delta}} \right],$$

and using Assumption 3.4, we obtain

$$\begin{aligned} |F_2| &\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ |i-j|\geq v_T}}^{k_T} \phi_\theta^2(h) \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\ &\lesssim \frac{1}{Th\phi_\theta(h)} \cdot \phi_\theta^2(h) \sum_{\substack{i,j=1 \\ |i-j|\geq v_T}}^{k_T} (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\ &\leq \frac{1}{Th\phi_\theta(h)} \cdot \phi_\theta(h) \cdot \frac{1}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{\nu}}} \left(\sum_{k=\lambda_T+1}^{\infty} k^\delta \alpha(k)^{1-\frac{2}{\nu}} \right) \end{aligned}$$

$$\begin{aligned} &\sim \frac{1}{Th\phi_\theta(h)} \cdot \frac{h^{2(1\wedge\beta)-1}}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{\nu}}} \left(\sum_{k=\lambda_T+1}^{\infty} k^\delta \alpha(k)^{1-\frac{2}{\nu}} \right) \\ &= o\left(\frac{1}{Th\phi_\theta(h)}\right). \end{aligned}$$

Hence,

$$\mathbb{E} \left[\tilde{g}_\theta^{(12)}(u, x) \right]^2 \rightarrow 0.$$

We next show that

$$\mathbb{E} \left[\tilde{g}_\theta^{(13)}(u, x) \right]^2 \rightarrow 0.$$

We again use $\Delta'_t(u, x; \theta)$ for simplicity of notations to denote $K_{1,h}(u - \frac{t}{T})K_{2,h}(d_\theta(x, X_{t,T}))$. Thus,

$$\begin{aligned} \mathbb{E} \left[\tilde{g}_\theta^{(13)}(u, x) \right]^2 &= \text{Var}(\zeta(u, x; \theta)) \\ &= \text{Var} \left(\sum_{t=k_T(a_T+v_T)+1}^T \Delta'_t(u, x; \theta) \varepsilon_{t,T} \right) \\ &= \sum_{i=k_T(a_T+v_T)+1}^T \text{Var}(\Delta'_i(u, x; \theta) \varepsilon_{i,T}) \\ &\quad + \sum_{i=k_T(a_T+v_T)+1}^T \sum_{\substack{j=k_T(a_T+v_T)+1 \\ i \neq j}}^T \text{Cov}(\Delta'_i(u, x; \theta) \varepsilon_{i,T}, \Delta'_j(u, x; \theta) \varepsilon_{j,T}) \\ &=: G_1 + G_2. \end{aligned}$$

Note that

$$\begin{aligned} G_1 &= \sum_{i=k_T(a_T+v_T)+1}^T \text{Var} \left(K_{1,h} \left(u - \frac{i}{T} \right) K_{2,h}(d_\theta(x, X_{i,T})) \varepsilon_{i,T} \right) \\ &= \sum_{i=k_T(a_T+v_T)+1}^T K_{1,h}^2 \left(u - \frac{i}{T} \right) \text{Var} \left(K_{2,h}(d_\theta(x, X_{i,T})) \varepsilon_{i,T} \right) \\ &\lesssim (T - k_T(a_T + v_T)) \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\ &\sim (T - k_T(a_T + v_T)) \cdot \left\{ \frac{\mathbb{E}[\varepsilon_t^2] \sigma^2(u, \langle \theta, x \rangle)}{Th} \left(\int_{[0,h]} K_1^2(w) dw \right) \right\}, \end{aligned}$$

implies that

$$\frac{1}{T} G_1 \lesssim \frac{1}{T} (T - k_T(a_T + v_T)) \cdot \left\{ \frac{\mathbb{E}[\varepsilon_t^2] \sigma^2(u, \langle \theta, x \rangle)}{Th} \left(\int_{[0,h]} K_1^2(w) dw \right) \right\} \rightarrow 0 \text{ as } T \rightarrow \infty.$$

We are left to show the bound of G_2 . Now, setting $\lambda_i = k_T(a_T + v_T)$, we have

$$G_2 = \sum_{i=k_T(a_T+v_T)+1}^T \sum_{\substack{j=k_T(a_T+v_T)+1 \\ i \neq j}}^T \text{Cov}(\Delta'_i(u, x; \theta) \varepsilon_{i,T}, \Delta'_j(u, x; \theta) \varepsilon_{j,T})$$

$$\begin{aligned}
&= \sum_{\substack{l_1, l_2=1 \\ l_1 \neq l_2}}^T \text{Cov} \left(\Delta'_{\lambda_i+l_1}(u, x; \theta) \varepsilon_{\lambda_i+l_1, T}, \Delta'_{\lambda_j+l_2}(u, x; \theta) \varepsilon_{\lambda_j+l_2, T} \right) \\
&= \sum_{\substack{i, j=1 \\ |i-j|>0}}^T \text{Cov} \left(\Delta'_i(u, x; \theta) \varepsilon_{i, T}, \Delta'_j(u, x; \theta) \varepsilon_{j, T} \right) \\
&\lesssim \sum_{\substack{i, j=1 \\ |i-j|>0}}^T \phi_\theta^2(h) \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\
&\leq \frac{\phi_\theta(h)}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{\nu}}} \left(\sum_{k=\lambda_T+1}^{\infty} k^\delta \alpha(k)^{1-\frac{2}{\nu}} \right) \\
&\sim \frac{h^{2(1 \wedge \beta)-1}}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{\nu}}} \left(\sum_{k=\lambda_T+1}^{\infty} k^\delta \alpha(k)^{1-\frac{2}{\nu}} \right) \rightarrow 0 \text{ as } T \rightarrow \infty.
\end{aligned}$$

The third equality follows from the fact that since $i \neq j$,

$$|\lambda_i - \lambda_j + l_1 - l_2| > 0.$$

Moreover, the preceding proof procedure follows similarly as that in $|F_2|$ making use of Davydov's lemma, Assumptions 3.3 and 3.4, the results in (A.5) and (A.6), and picking again

$$\lambda_T = \left\lfloor (\phi_\theta(h))^{-\frac{1-(2/\nu)}{\delta}} \right\rfloor.$$

We then infer that

$$\mathbb{E} \left[\tilde{g}_\theta^{(13)}(u, x) \right]^2 \rightarrow 0,$$

as $T \rightarrow \infty$. Finally, we will prove the last part of the decomposition (A.11). First, we will show that the summation in $\tilde{g}_\theta^{(11)}(u, x)$ are asymptotically independent. This ensures us the use of the conditions of Lindeberg-Feller for finite normality. To establish this, we first note that the processes $(Y_{t,T}, X_{t,T})$ are strongly mixing and apply the Volkonskii and Rosanov inequality in Lemma B.5. Then, for a $\mathcal{F}_{i_a}^{j_a}$ -measurable η_a where

$$i_a = (j-1)(a_T + v_T) + 1, \text{ and } j_a = ja_T + (j-1)v_T,$$

we have

$$\begin{aligned}
\left| \mathbb{E} \left[\exp \left(itT^{-\frac{1}{2}} \tilde{g}_\theta^{(11)}(u, x) \right) \right] - \prod_{j=1}^{k_T} \mathbb{E} \left[\exp \left(itT^{-\frac{1}{2}} \eta_j(u, x; \theta) \right) \right] \right| &\leq 16k_T \alpha(v_T) \\
&\sim 16 \frac{T}{a_T} \alpha(v_T),
\end{aligned}$$

which tends to zero using (A.10) implying that the asymptotic independence is achieved. We now find the variance of $\tilde{g}_\theta^{(11)}(u, x)$. We obtain

$$\text{Var} \left(\tilde{g}_\theta^{(11)}(u, x) \right) = \text{Var} \left(\frac{1}{\sqrt{Th\phi_\theta(h)}} \sum_{j=1}^{k_T} \eta_j(u, x; \theta) \right)$$

$$\begin{aligned}
&= \text{Var} \left(\frac{1}{\sqrt{Th\phi_\theta(h)}} \sum_{j=1}^{k_T} \eta_j(u, x; \theta) \right) \\
&\leq \frac{1}{Th\phi_\theta(h)} \sum_{j=1}^{k_T} \text{Var}(\eta_j(u, x; \theta)) + \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ i \neq j}}^{k_T} \text{Cov}(\eta_i(u, x; \theta), \eta_j(u, x; \theta)).
\end{aligned}$$

Now, we can see that

$$\begin{aligned}
\text{Var}(\eta_j(u, x; \theta)) &= \text{Var} \left(\sum_{t=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&= \sum_{t=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} K_{1,h}^2 \left(u - \frac{t}{T} \right) \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&\lesssim a_T \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right).
\end{aligned}$$

This means that

$$\begin{aligned}
\frac{1}{Th\phi_\theta(h)} \sum_{j=1}^{k_T} \text{Var}(\eta_j(u, x; \theta)) &\lesssim \frac{1}{h\phi_\theta(h)} \cdot \frac{k_T a_T}{T} \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) \\
&\rightarrow \frac{1}{h\phi_\theta(h)} \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right),
\end{aligned}$$

since $\frac{k_T a_T}{T} \rightarrow 1$. On the other hand, similar to our previous calculations, we can deduce with $\Delta'_t(u, x; \theta) = K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h}(d_\theta(x, X_{t,T}))$ that

$$\begin{aligned}
&\frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ i \neq j}}^{k_T} \text{Cov}(\eta_i(u, x; \theta), \eta_j(u, x; \theta)) \\
&= \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ i \neq j}}^{k_T} \text{Cov} \left(\sum_{t=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} \Delta'_t(u, x; \theta) \varepsilon_{t,T}, \sum_{t'=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} \Delta'_{t'}(u, x; \theta) \varepsilon_{t',T} \right) \\
&\lesssim \frac{1}{Th\phi_\theta(h)} \sum_{\substack{i,j=1 \\ i \neq j}}^{k_T} \phi_\theta^2(h) \cdot (\alpha(|i-j|))^{1-\frac{2}{\nu}} \\
&\sim \frac{h^{2(1 \wedge \beta)-1}}{\lambda_T^\delta (\phi_\theta(h))^{1-\frac{2}{\nu}}} \left(\sum_{k=\lambda_T+1}^{\infty} k^\delta \alpha(k)^{1-\frac{2}{\nu}} \right) \rightarrow 0 \text{ as } T \rightarrow \infty,
\end{aligned}$$

where λ_T is as defined in Assumption 3.4. Therefore, we conclude that

$$\text{Var}(\tilde{g}_\theta^{(11)}(u, x)) \lesssim \frac{1}{h\phi_\theta(h)} \left(\int_{[0,h]} K_1^2(w) dw \right) \cdot \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right) := V_\theta(u, x).$$

To complete the proof for the finite-dimensional convergence, we need to show that for sufficiently large T ,

$$\frac{1}{T} \sum_{j=1}^{k_T} \mathbb{E} \left[\eta_j^2(u, x; \theta) \mathbb{1}_{\{|\eta_j(u, x; \theta)| > \varepsilon V_\theta(u, x) \sqrt{T}\}} \right] \rightarrow 0. \quad (\text{A.12})$$

Observe that using A.10,

$$\begin{aligned} \max_{1 \leq j \leq k_T} \frac{|\eta_j(u, x; \theta)|}{\sqrt{T}} &= \frac{1}{\sqrt{T}} \max_{1 \leq j \leq k_T} \left| \sum_{t=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right| \\ &\leq \frac{1}{\sqrt{T}} \max_{1 \leq j \leq k_T} \sum_{t=(j-1)(a_T+v_T)+1}^{ja_T+(j-1)v_T} \left| K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right| \\ &= \frac{a_T}{\sqrt{T}} \max_{1 \leq j \leq k_T} \left| K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})) \varepsilon_{t,T} \right| \rightarrow 0. \end{aligned}$$

Hence, when T is large enough, the set

$$\{|\eta_j(u, x; \theta)| > \varepsilon V_\theta(u, x) \sqrt{T}\}$$

is empty proving (A.12). This further implies that

$$\frac{1}{T} S_T \rightarrow N(0, V_\theta(u, x)),$$

where

$$S_T = \sum_{t=1}^T \left((Y_{t,T} - \mathbb{E}[Y_{t,T}]) K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})) \right).$$

To prove the general case, we employ a truncation argument since the response variable $Y_{t,T}$ is not necessarily bounded. With L being the truncation point, we set

$$\kappa_L(y) = y \mathbb{1}_{\{|y| \leq L\}}, \quad \text{and } m_{\theta,L}(u, x) = \mathbb{E} \left[\kappa_L(Y_{t,T}) | X_{t,T} = x \right].$$

Recall that

$$\Delta'_t(u, x; \theta) = K_{1,h} \left(u - \frac{t}{T} \right) K_{2,h} (d_\theta(x, X_{t,T})),$$

and define

$$Z_{t,T}^L := (\kappa_L(Y_{t,T}) - m_{\theta,L}(u, x)) \Delta'_t(u, x; \theta) - \mu_T^L, \quad (\text{A.13})$$

where μ_T^L is the mean of the first term on the right side, and

$$\tilde{Z}_{t,T}^L := \frac{1}{h\phi_\theta(h)} Z_{t,T}^L \sqrt{h\phi_\theta(h)},$$

so that for each $L > 0$,

$$\text{Var}(\tilde{Z}_{t,T}^L) \rightarrow V_\theta^L(u, x).$$

We have

$$\begin{aligned}
\text{Var}(\tilde{Z}_{t,T}^L) &= \text{Var}\left(\frac{Z_{t,T}^L \sqrt{h\phi_\theta(h)}}{h\phi_\theta(h)}\right) \\
&= \frac{h\phi_\theta(h)}{(h\phi_\theta(h))^2} \text{Var}(Z_{t,T}^L) \\
&= \frac{1}{h\phi_\theta(h)} \text{Var}\left(\left(\kappa_L(Y_{t,T}) - m_{\theta,L}(u, x)\right)\Delta'_t(u, x; \theta) - \mu_T^L\right) \\
&= \frac{1}{h\phi_\theta(h)} \text{Var}\left(\left(Y_{t,T}\mathbb{1}_{\{|Y_{t,T}| \leq L\}} - \mathbb{E}\left[Y_{t,T}\mathbb{1}_{\{|Y_{t,T}| \leq L\}} | X_{t,T} = x\right]\right)\right. \\
&\quad \left. \times K_{1,h}\left(u - \frac{t}{T}\right) K_{2,h}(d_\theta(x, X_{t,T}))\right) \\
&= \frac{1}{h\phi_\theta(h)} K_{1,h}^2\left(u - \frac{t}{T}\right) \text{Var}\left(K_{2,h}(d_\theta(x, X_{t,T}))\right. \\
&\quad \left. \times \left(Y_{t,T}\mathbb{1}_{\{|Y_{t,T}| \leq L\}} - \mathbb{E}\left[Y_{t,T}\mathbb{1}_{\{|Y_{t,T}| \leq L\}} | X_{t,T} = x\right]\right)\right) \\
&= \frac{1}{h\phi_\theta(h)} \left(\int_{[0,h]} K_1^2(w)dw\right) \text{Var}\left(K_{2,h}(d_\theta(x, X_{t,T}))\right. \\
&\quad \left. \times \left(Y_{t,T}\mathbb{1}_{\{|Y_{t,T}| \leq L\}} - \mathbb{E}\left[Y_{t,T}\mathbb{1}_{\{|Y_{t,T}| \leq L\}} | X_{t,T} = x\right]\right)\right) \\
&:= V_\theta^L(u, x) \rightarrow V_\theta(u, x), \text{ as } T \rightarrow \infty. \tag{A.14}
\end{aligned}$$

We also define

$$S_T^L := \sum_{t=1}^T \tilde{Z}_{t,T}^L, \quad \text{and} \quad \bar{S}_T^L := \sum_{t=1}^T (\tilde{Z}_{t,T} - \tilde{Z}_{t,T}^L).$$

Now, we can infer by (A.13), setting $\eta_j^L(u, x; \theta)$ as defined in (A.11), and the fact that $\Delta'_t(u, x; \theta)$ is bounded, we have by replacing $\tilde{Z}_{t,T}$ by $\tilde{Z}_{t,T}^L$,

$$\max_{1 \leq j \leq k_T} \frac{|\eta_j^L(u, x; \theta)|}{\sqrt{T}} \rightarrow 0, \text{ implying } \{|\eta_j^L(u, x; \theta)| > \varepsilon V_\theta^L(u, x) \sqrt{T}\} = \emptyset.$$

This means that

$$\frac{1}{T} S_T^L \xrightarrow{L} N(0, V_\theta^L(u, x)). \tag{A.15}$$

The final part of the proof intends to show that

$$\frac{1}{T} \text{Var}(\bar{S}_T^L) \rightarrow 0 \text{ as first } T \rightarrow \infty, \text{ and then } L \rightarrow \infty. \tag{A.16}$$

We have

$$\left| \mathbb{E}\left[\exp\left(\frac{itS_T}{\sqrt{T}}\right)\right] - \exp\left(-\frac{t^2 V_\theta(u, x)}{2}\right) \right|$$

$$\begin{aligned}
&= \left| \mathbb{E} \left[\exp \left(\frac{itS_T^L}{\sqrt{T}} \right) \right] - \exp \left(-\frac{t^2 V_\theta^L(u, x)}{2} \right) + \exp \left(-\frac{t^2 V_\theta^L(u, x)}{2} \right) - \exp \left(-\frac{t^2 V_\theta(u, x)}{2} \right) \right| \\
&\leq \left| \mathbb{E} \left[\exp \left(\frac{itS_T^L}{\sqrt{T}} \right) \right] - \exp \left(-\frac{t^2 V_\theta^L(u, x)}{2} \right) \right| \\
&\quad + \left| \mathbb{E} \left[\exp \left(\frac{it\bar{S}_T^L}{\sqrt{T}} \right) \right] - 1 \right| + \left| \exp \left(-\frac{t^2 V_\theta^L(u, x)}{2} \right) - \exp \left(-\frac{t^2 V_\theta(u, x)}{2} \right) \right| \\
&:= I + II + III.
\end{aligned}$$

Note that by (A.15), $I \rightarrow 0$ as $T \rightarrow \infty$ for every $L > 0$. Also, by (A.16), $II \rightarrow 0$ as first $T \rightarrow \infty$ and then $L \rightarrow \infty$. Last, since as $L \rightarrow \infty$,

$$V_\theta^L(u, x) \rightarrow V_\theta(u, x),$$

we have $III \rightarrow 0$ using (A.14). Hence, we are left to prove (A.16). Observe that \bar{S}_T^L has the same structure as S_T by replacing Y_t with $Y_t \mathbb{1}_{\{|Y_t|>L\}}$. We can use similar proof procedure as in (A.14), and obtain

$$\begin{aligned}
\lim_{T \rightarrow \infty} \frac{1}{T} \text{Var}(\bar{S}_T^L) &= \lim_{T \rightarrow \infty} \frac{1}{T} \cdot \frac{1}{h\phi_\theta(h)} \left(\int_{[0,h]} K_1^2(w) dw \right) \text{Var} \left(K_{2,h}(d_\theta(x, X_{t,T})) \right. \\
&\quad \left. \times (Y_{t,T} \mathbb{1}_{\{|Y_{t,T}|>L\}} - \mathbb{E}[Y_{t,T} \mathbb{1}_{\{|Y_{t,T}|>L\}} | X_{t,T} = x]) \right),
\end{aligned}$$

which converges to 0. By dominated convergence, the right side converges to 0 as L tends to infinity. This completes the proof for Theorem 3.2. \square

B. Appendix-B

This section provides a detailed overview of the essential lemmas underpinning the primary results of this study. Lemmas B.1 and B.2, both found in [57], play a crucial role in deriving the convergence results outlined in Proposition 3.1. Specifically, [60] established Lemma B.3, which introduces an exponential inequality for strongly mixing sequences, detailed in Theorem 2.1 of his paper. This lemma is instrumental in demonstrating the convergence of the second term in the decomposition of the general kernel estimator. Further, Lemma B.4, attributed to [39] (with the proof presented in Corollary A.2 of [52]), is central to proving the asymptotic negligibility of covariance terms within the decomposition of $\hat{g}_\theta^{(1)}(u, x)$, as specified in Eq (A.11). Last, Lemma B.5 is pivotal in establishing the asymptotic independence of the term $\tilde{g}_\theta^{(11)}(u, x)$, which represents the first term in Eq (A.11). For a comprehensive introduction, refer to Proposition 2.6 in [41].

Lemma B.1. *Suppose that the kernel $K_1(\cdot)$ satisfies Assumption 3.3 (KB1). Then, for $k = 0, 1, 2$,*

$$\sup_{u \in I_h} \left| \frac{1}{Th} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) \left(\frac{u - \frac{t}{T}}{h} \right)^k - \int_0^1 \frac{1}{h} K_{1,h}(u - v) \left(\frac{u - v}{h} \right)^k dv \right| = O \left(\frac{1}{Th^2} \right).$$

Lemma B.2. *Suppose that the kernel $K_1(\cdot)$ satisfies Assumption 3.3 (KB1) and let $g : [0, 1] \times \mathcal{H} \rightarrow \mathbb{R}$ such that $(u, x) \mapsto g(u, x)$ be continuously differentiable with respect to u . Then,*

$$\sup_{u \in I_h} \left| \frac{1}{Th} \sum_{t=1}^T K_{1,h} \left(u - \frac{t}{T} \right) g \left(\frac{t}{T}, x \right) - g(u, x) \right| = O \left(\frac{1}{Th^2} \right) + o(h).$$

Lemma B.3. Let $Z_{t,T}$ be a zero-mean triangular array such that $|Z_{t,T}| \leq b_T$ with α -mixing coefficient $\alpha(k)$. Then, for any $\varepsilon > 0$ and $S_T \leq T$ with $\varepsilon > 4S_T b_T$,

$$\mathbb{P}\left(\left|\sum_{t=1}^T Z_{t,T}\right| \geq \varepsilon\right) \leq 4 \exp\left(-\frac{\varepsilon^2}{64\sigma_{S_T,T}^2 \frac{T}{S_T} + \frac{8}{3}\varepsilon b_T S_T}\right) + 4\frac{T}{S_T}\alpha(S_T),$$

where

$$\sigma_{S_T,T}^2 = \sup_{0 \leq j \leq T-1} \mathbb{E}\left[\left(\sum_{t=j+1}^{(j+S_T) \wedge T} Z_{t,T}\right)^2\right].$$

Lemma B.4. (Corollary 2 in [52]) Suppose that X and Y are random variables which are \mathcal{G} - and \mathcal{H} -measurable, respectively, and that $\mathbb{E}|X|^p < \infty$, $\mathbb{E}|Y|^q < \infty$, where $p, q > 1$, $\frac{1}{p} + \frac{1}{q} < 1$. Then

$$\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \leq 8\|X\|_p\|Y\|_q[\alpha(\mathcal{G}, \mathcal{H})]^{1-\frac{1}{p}-\frac{1}{q}}.$$

Lemma B.5. (Proposition 2.6 in [41]) Suppose that $\{X_t : t = 0, \pm 1, \pm 2, \dots\}$ Let \mathcal{F}_i^j denote the σ -algebra generated by $\{X_t : i \leq t \leq j\}$ and

$$\alpha(n) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty} |\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)|.$$

Let ξ_1, \dots, ξ_k be complex-valued random variables measurable with respect to the σ -algebras $\mathcal{F}_{i_1}^{j_1}, \dots, \mathcal{F}_{i_k}^{j_k}$, respectively. Suppose that $i_{l+1} - j_l \geq n$ for $l = 1, \dots, k-1$ and $j_l \geq i_l$ and $\mathbb{P}(|\xi_l| \leq 1) = 1$ for $l = 1, \dots, k$. Then

$$|\mathbb{E}(\xi_1 \cdots \xi_k) - \mathbb{E}(\xi_1) \cdots \mathbb{E}(\xi_k)| \leq 16(k-1)\alpha(n).$$



AIMS Press

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