



Research article

Weak convergence of the conditional single index U -statistics for locally stationary functional time series

Salim Bouzebda*

Université de technologie de Compiègne, Laboratory of Applied Mathematics of Compiègne (LMAC), France

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

Abstract: In recent years, there has been a notable shift in focus towards the analysis of non-stationary time series, driven largely by the complexities associated with delineating significant asymptotic behaviors inherent to such processes. The genesis of the theory of locally stationary processes arises from the quest for asymptotic inference grounded in nonparametric statistics. This paper endeavors to formulate a comprehensive framework for conducting inference within the realm of locally stationary functional time series by harnessing the conditional U -statistics methodology as propounded by W. Stute in 1991. The proposed methodology extends the Nadaraya-Watson regression function estimations. Within this context, a novel estimator was introduced for the single index conditional U -statistics operator, adept at accommodating the non-stationary attributes inherent to the data-generating process. The primary objective of this paper was to establish the weak convergence of conditional U -processes within the domain of locally stationary functional mixing data. Specifically, the investigation delved into scenarios of weak convergence involving functional explanatory variables, considering both bounded and unbounded sets of functions while adhering to specific moment requirements. The derived findings emanate from broad structural specifications applicable to the class of functions and models under scrutiny. The theoretical insights expounded in this study constitute pivotal tools for advancing the domain of functional data analysis.

Keywords: almost complete convergence; convergence rates; exponential inequality; kernel regression; empirical processes; random measures; U -processes; conditional U -statistics

Mathematics Subject Classification: 60F17, 62E17, 62G05, 62G08, 62G20, 62G35, 62G07, 62G32

1. Introduction and motivations

In the evolutionary trajectory of asymptotic outcomes related to U -statistics, with a particular emphasis on independent and identically distributed random variables, pivotal contributions can be

ascribed to esteemed figures like [93, 99, 181], among others. When extending these advancements to accommodate scenarios involving weak dependency assumptions, notable references include [30, 39, 40, 66, 119, 120]. For a comprehensive grasp of U -statistics and U -processes, scholars are directed to seminal works such as [11, 12, 14, 31, 113, 117]. A substantial leap forward in the theoretical landscape of U -processes is accredited to [64], who made significant contributions by assimilating insights from empirical process theory. Their introduction of innovative techniques, including decoupling inequality and randomization, played a pivotal role in propelling the theoretical framework forward. The applications of U -processes traverse diverse statistical domains, encompassing testing for qualitative features of functions in nonparametric statistics [1, 88, 118], cross-validation for density estimation [144], and establishing limiting distributions of M-estimators [11, 64, 157, 158]. In the realm of machine learning, U -statistics find multifaceted applications in clustering, image recognition, ranking, and learning on graphs. The natural estimates of risk prevalent in various machine learning contexts often manifest in the form of U -statistics, as elucidated in [54]. Instances of U -statistics are also discerned in various contexts, such as empirical performance measures in metric learning, exemplified by [50]. When confronted with U -statistics characterized by random kernels exhibiting diverging orders, pertinent literature includes contributions from [85, 98, 152, 162]. Infinite-order U -statistics manifest as invaluable tools for constructing simultaneous prediction intervals, providing insights into the uncertainty inherent in ensemble methods like subbagging and random forests, as explicated in [146]. The MeanNN approach, introduced by [75] for estimating differential entropy, intricately involves the utilization of the U -statistic. Additionally, [129] proposes a novel test statistic for goodness-of-fit tests, employing U -statistics. A model-free approach to clustering and classifying genetic data based on U -statistics is explored by [55], presenting alternative perspectives driven by the adaptability of U -statistics to a diverse array of genetic issues and their capability to accommodate various data types. Furthermore, [125] advocates for the natural application of U -statistics in examining random compressed sensing matrices in the non-asymptotic regime. For the latest references in this context, please consult [43, 163, 164]. In the realm of nonparametric density and regression function estimation, [166] introduces a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, referred to as conditional U -statistics. These estimators can be perceived as an extension of the Nadaraya-Watson estimates for regression functions, initially proposed by [139, 184]. The nonparametric domain of density and regression function estimation has been a focal point for statisticians and probabilists over numerous years, resulting in the evolution of various methodologies. Kernel nonparametric function estimation methods, in particular, have garnered substantial attention. For a thorough exploration of the research literature and statistical applications in this field, one is encouraged to consult [67, 74, 95, 140, 160, 182], and the pertinent references therein.

This investigation delves into the intricacies of nonparametric conditional U -statistics. To facilitate our exploration, we commence by introducing the estimators proposed by [166]. Consider a regular sequence of random elements $\{(\mathbf{X}_i, Y_i) : i \in \mathbb{N}^*\}$, where $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathcal{Y}$, a Polish space, with $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. Let $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ be a measurable function. In this paper, our central focus revolves around the estimation of the conditional expectation or regression function:

$$r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(Y_1, \dots, Y_m) \mid (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t}),$$

for $\mathbf{t} \in \mathbb{R}^{dm}$, provided it exists, namely, when $\mathbb{E}(|\varphi(Y_1, \dots, Y_m)|) < \infty$. We introduce a kernel function

$K : \mathbb{R}^d \rightarrow \mathbb{R}$ with support contained in $[-B, B]^d$, where $B > 0$, adhering to the following conditions:

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty \quad \text{and} \quad \int K(\mathbf{x}) d\mathbf{x} = 1.$$

Reference [166] introduced a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, known as conditional U -statistics, defined for each $\mathbf{t} \in \mathbb{R}^{dm}$ as:

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1, \dots, i_m) \in I_n^m} \varphi(Y_{i_1}, \dots, Y_{i_m}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \cdots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}{\sum_{(i_1, \dots, i_m) \in I_n^m} K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \cdots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}, \quad (1.1)$$

where I_n^m is the set of all m -tuples of different integers between 1 and n :

$$I_n^m = \left\{ \mathbf{i} = (i_1, \dots, i_m) : 1 \leq i_j \leq n \quad \text{and} \quad i_j \neq i_r \quad \text{if} \quad j \neq r \right\},$$

and $\{h_n\}_{n \geq 1}$ is a sequence of positive constants converging to zero at the rate $nh_n^{dm} \rightarrow \infty$. In the specific scenario of $m = 1$, where $r^{(m)}(\varphi, t)$ simplifies to $r^{(1)}(\varphi, t) = \mathbb{E}(\varphi(\mathbf{Y}) \mid \mathbf{X} = t)$, the estimator by Stute transforms into the Nadaraya-Watson estimator of $r^{(1)}(\varphi, t)$. The study conducted by [154] focused on estimating the rate of uniform convergence in \mathbf{t} of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ to $r^{(m)}(\varphi, \mathbf{t})$. In [148], the paper discusses and compares the limit distributions of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ with those obtained by Stute. [97] extended the results of [166] to weakly dependent data under appropriate mixing conditions (also see [17]). They applied these findings to verify the Bayes risk consistency of corresponding discrimination rules similar to [167] and Section 5.1. In [169], symmetrized nearest neighbor conditional U -statistics are proposed as alternatives to the usual kernel-type estimators, and reference can also be made to [48]. [86] explored the functional conditional U -statistic and established its finite-dimensional asymptotic normality. Despite the subject's importance, nonparametric estimation of conditional U -statistics in a functional data framework has received relatively limited attention. Recent advancements are presented in [40, 48], addressing problems related to uniform bandwidth consistency in a general setting. In [104], the test of independence in the functional framework based on the Kendall statistics was investigated, which can be considered as particular cases of U -statistics. Extending this exploration to conditional empirical U -processes in the functional setting is practically useful and technically more challenging. Two perspectives on conditional U -processes are presented 1) they are infinite-dimensional versions of conditional U -statistics (with one kernel) and 2) they are stochastic processes that are nonlinear generalizations of conditional empirical processes. Both views are valuable because: 1) from a statistical standpoint, considering a rich class of statistics is more interesting than a single statistic; 2) mathematically, insights from empirical process theory can be applied to derive limit or approximation theorems for U -processes. Importantly, extending U -statistics to U -processes demands substantial effort and different techniques, and generalization from conditional empirical processes to conditional U -processes is highly nontrivial.

The prevalent practice of assuming stationarity in time series modeling has prompted the development of various models, techniques, research, and methodologies. However, this assumption may not always be suitable for spatio-temporal data, even with detrending and deseasonalization.

Many pivotal time series models exhibit nonstationarity, observed in diverse physical phenomena and economic data, rendering classical methods ineffective. To address this challenge, the concept of the locally stationary random process was introduced by [161]. This type of process approximates a non-stationary process by a stationary one locally over short periods. The intuitive concept of local stationarity is also explored in the works of [57, 58, 142, 149, 153], among others. The groundbreaking work of [57] notably serves as a robust foundation for the inference of locally stationary processes. In addition to generalizing stationary processes, this innovative approach eliminates time-varying parameters. Over the past decade, the theory of empirical processes for locally stationary time series has garnered significant attention. Empirical processes theory plays a crucial role in addressing statistical problems and has expanded into time series analysis and regression estimation. Relevant references in this context include [59, 179] and more recent contributions such as [134, 147]. The extension of the previously discussed exploration to conditional empirical U -processes bears significant interest from both practical and theoretical standpoints. We specifically delve into the domain of conditional U -processes indexed by a class of functions within the framework of functional data. Building upon insights from [8], functional data analysis (FDA) emerges as a statistical field dedicated to analyzing infinite-dimensional variables such as curves, sets, and images. Experiencing remarkable growth over the past two decades, FDA has become a crucial area of investigation in data science, fueled by advancements in data collection technology during the “Big Data” revolution. For an introduction to FDA, readers can refer to the books by [78, 151], providing fundamental analysis methods and case studies across various domains like criminology, economics, archaeology, and neurophysiology. Notably, the extension of probability theory to random variables taking values in normed spaces predates recent literature on functional data, with foundational knowledge available in [9, 87]. In the context of regression estimation and nonparametric models for data in normed vector spaces, valuable references include [78, 136], along with additional contributions from [32, 102, 126]. Modern empirical process theory has been applied to functional data, as demonstrated by [82], who established uniform consistency rates for functionals of the conditional distribution, including the regression function, conditional cumulative distribution, and conditional density. [109] extended this by providing consistency rates for various functional nonparametric models, uniformly in bandwidth (UIB consistency). Recent advancements in this field can be explored through references such as [3, 40, 42, 49, 68, 69, 137]. This strongly motivates the consideration of regression models that offer dimension reduction. Single index models are widely used to achieve this by assuming that the predictors’ influence on the response can be simplified to a single index. This index represents a projection in a specified direction and is combined with a nonparametric link function, simplifying the predictors to a one-dimensional index while still incorporating important characteristics. Additionally, because the nonparametric link function only operates on a one-dimensional index, these models are not affected by the problem of having a high number of dimensions, known 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; for further details, interested readers can refer to [23, 91, 123, 138, 171, 183].

Recent progress in functional data analysis underscores the need for developing models to address the challenges of dimensionality reduction (refer to [90, 126] for recent surveys, and also [4, 5, 40, 41, 165]). In response to this, semiparametric approaches emerge as promising solutions. The functional single-index model (FSIM) has gained attention in this context, with exploration

by [2, 16, 79]. Furthermore, [106] proposed functional single-index composite quantile regression, estimating the unknown slope function and link function through B -spline basis functions. A functional single index model with coefficient functions restricted to a subregion was introduced by [143]. The estimation of a general functional single index model, where the conditional distribution depends on the functional predictor via a single index structure, was investigated by [188]. Innovatively, [173] developed a new estimation method that combines functional principal component analysis, B -spline modeling, and profile estimation for parameters and functions. Addressing the estimation of the functional single index regression model with missing responses for strong mixing time series data, [127, 128] made valuable contributions. [76] introduced a functional single-index varying coefficient model with the functional predictor as the single-index part. Utilizing functional principal component analysis and basis function approximation, they obtained estimators for slope and coefficient functions, proposing an iterative estimating procedure. An automatic and location-adaptive procedure for estimating regression in an FSIM based on k -Nearest Neighbors (k NN) principles was presented by [145]. Motivated by imaging data analysis, [121] proposed a novel functional varying-coefficient single-index model for regression analysis of functional response data on a set of covariates. Investigating a functional Hilbertian regressor for nonparametric estimation of the conditional cumulative distribution with a scalar response variable in a single index structure, [15] made notable contributions. An alternative approach was introduced by [52], extending to the multi-index case without anchoring the true parameter on a prespecified sieve. Their detailed theoretical analysis of a direct kernel-based estimation scheme establishes a polynomial convergence rate.

The primary objective of this paper is to scrutinize a comprehensive framework for the single-index conditional U -process of any fixed order, indexed by a class of functions within a nonparametric context. Specifically, we explore the conditional U -process in the realm of functional covariates, considering the potential non-stationary nature of functional time series. The main aim of this study is to offer an initial and comprehensive theoretical examination within this specific context. To achieve this, we skillfully apply large sample theory techniques developed for empirical processes and U -empirical processes. This paper meticulously tackles various technical hurdles. Initially, it delves into the nonlinear expansion of the single index concept and conditional U -statistics. Subsequently, it addresses the extension of the Hoeffding decomposition to non-stationary time series. Finally, it confronts the complexity stemming from the unbounded function class, leading to extensive and intricate proofs.

The manuscript's organization is structured as follows. Section 2 provides a detailed exposition of our theoretical framework, elucidating essential definitions and contextual explanations while introducing technical assumptions. Our principal findings are presented in Sections 3 and 4. Specifically, Section 3 unveils convergence rate results, reintroducing the pivotal Hoeffding decomposition technique. Accommodating our outcomes on weak convergence, Section 4 delves into the details of these results. Section 5 accentuates selected applications. In Section 6, we explore bandwidth selection methodologies utilizing cross-validation procedures. Concluding reflections are encapsulated in Section 7. The comprehensive proofs are furnished in Section 8. Lastly, Appendix 8 provides technical properties and lemmas for easy reference.

2. Background and preliminaries

2.1. Notation

In this document, the notation $a_n \lesssim b_n$ is employed to signify the existence of a constant C , which is independent of n and may vary between lines, unless explicitly specified otherwise. This constant is such that $a_n \leq Cb_n$ holds for all n . Additionally, the notation $a_n \ll b_n$ indicates that $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$. When both $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold, their equivalence is denoted as $a_n \sim b_n$. Moreover, (i_1, \dots, i_m) is denoted as \mathbf{i} , and $(i_1/n, \dots, i_m/n)$ as \mathbf{i}/n . For any $c, d \in \mathbb{R}$, the expressions $c \vee d = \max\{c, d\}$ and $c \wedge d = \min\{c, d\}$ are employed. The notation $\lfloor a \rfloor$ signifies the integer part of a number. Additionally, for $m < n$, where m and n are positive integers, $C_m^n = \frac{n!}{(n-m)!m!}$ is defined. The set I_n^m is introduced as

$$I_n^m := \left\{ \mathbf{i} = (i_1, \dots, i_m) : 1 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r \right\},$$

comprising all m -tuples of distinct integers between 1 and n .

2.2. Model

Consider the stochastic processes $\{Y_{i,n}, X_{i,n}\}_{i=1}^n$, where $Y_{i,n}$ takes values in a space \mathcal{Y} , and $X_{i,n}$ is in an abstract space \mathcal{H} . We assume \mathcal{H} is a semi-metric vector space with a semi-metric $d(\cdot, \cdot)^*$ which, in most applications, would be a Hilbert or Banach space. We consider the semi-metric $d_\theta(\cdot, \cdot)$ associated with the single-index $\theta \in \mathcal{H}$, defined as

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

Consider any function $\varphi(\cdot)$ of k variables (the U -kernel) such that $\varphi(Y_1, \dots, Y_m)$ is integrable. For $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{H}^m$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \Theta^m \subset \mathcal{H}^m$, define the regression functional parameter as

$$\begin{aligned} r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right) &:= \mathbb{E}(\varphi(Y_1, \dots, Y_m) \mid \langle X_1, \theta_1 \rangle = \langle x_1, \theta_1 \rangle, \dots, \langle X_m, \theta_m \rangle = \langle x_m, \theta_m \rangle) \\ &=: \mathbb{E}(\varphi(Y_{\mathbf{i}}) \mid \langle X_{\mathbf{i}}, \boldsymbol{\theta} \rangle = \langle \mathbf{x}, \boldsymbol{\theta} \rangle), \quad \mathbf{i} = (1, \dots, m). \end{aligned} \quad (2.1)$$

In this study, we consider the following model:

$$\varphi(Y_{\mathbf{i},n}) = r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right) + \sigma\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i},n}\right) \varepsilon_{\mathbf{i}}, \quad \mathbf{i} = (i_1, \dots, i_m), \quad 1 \leq i_j \leq n, \quad (2.2)$$

where $\{\varepsilon_{\mathbf{i}}\}_{\mathbf{i} \in I_n^m}$ is a sequence of univariate independent and identically distributed random variables, independent of $\{X_{i,n}\}_{i=1}^n$. We denote $\sigma\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i},n}\right) \varepsilon_{\mathbf{i}}$ as $\varepsilon_{\mathbf{i},n}$. Furthermore, we assume that the process is a locally stationary functional time series. In a heuristic sense, a process $\{X_{i,n}\}$ is considered locally stationary if it displays approximately stationary behavior locally in time. The regression function $r^{(m)}(\varphi, \cdot, \mathbf{x}, \boldsymbol{\theta})$ is allowed to change smoothly over time, depending on a rescaled quantity \mathbf{i}/n rather than on the specific point \mathbf{i} (where \mathbf{i} typically represents time in a time series framework).

*A semi-metric (or pseudo-metric) $d(\cdot, \cdot)$ is a metric allowing $d_{\theta_1}(x_1, x_2) = 0$ for some $x_1 \neq x_2$.

2.3. Local stationarity

We delve into the exploration of non-stationary processes characterized by dynamics evolving gradually over time, manifesting behaviors akin to stationarity at a local level. This conceptual realm has undergone comprehensive scrutiny, exemplified by references such as [57, 105, 115, 131, 141, 185]. For illustration, consider a continuous function $a : [0, 1] \rightarrow \mathbb{R}$ and a sequence of i.i.d. random variables $(\varepsilon_i)_{i \in \mathbb{N}}$. The stochastic process $X_{i,n} = a(i/n) + \varepsilon_i$, where $i \in \{1, \dots, n\}$ and $n \in \mathbb{N}$, can exhibit “almost” stationary behavior for i close to a specific point i^* (e.g., i^* in $\{1, \dots, n\}$), given that $a(i^*/n) \approx a(i/n)$. However, this process is not strictly weakly stationary. To capture this type of gradual change, the concept of local stationarity was introduced by [57], wherein the spectral representation of the underlying stochastic process is locally approximated. In our framework, the process $\{X_{i,n}\}$ can be stochastically approximated by a stationary process $\{X_{i,n}^{(u)}\}$ around each rescaled time point u , specifically for those values of i where $i/n - u$ is small. Since our focus is on functional data, we define a functional time series as locally stationary if it can be locally approximated by a stationary functional time series. We will provide a standard definition of local stationarity.

Definition 2.1 (local stationarity). *For a sequence of stochastic processes, indexed by $n \in \mathbb{N}$ and taking values in \mathcal{H} , denoted as $\{X_{i,n}\}$, it is deemed locally stationary if, for all rescaled times $u \in [0, 1]$, there exists an associated \mathcal{H} -valued process $\{X_i^{(u)}\}$ that is strictly stationary. This association is characterized by the inequality:*

$$d_\theta(X_{i,n}, X_i^{(u)}) \leq \left(\left| \frac{i}{n} - u \right| + \frac{1}{n} \right) U_{i,n}^{(u)} \text{ a.s.}, \quad (2.3)$$

This holds for all $1 \leq i \leq n$, where $\{U_{i,n}^{(u)}\}$ is a positive valued process satisfying $\mathbb{E}[(U_{i,n}^{(u)})^\rho] < C$ for some $\rho > 0$, $C < \infty$. These conditions are independent of u , i , and n .

Definition 2.1 represents a natural extension of the concept of local stationarity for real-valued time series introduced in [57]. In a more specific context, [176] and [175] elaborate on this definition, considering \mathcal{H} as a Hilbert space $L_{\mathbb{R}}^2[0, 1]$. Here, all real-valued functions are square-integrable with respect to the Lebesgue measure on the interval $[0, 1]$, equipped with the inner product L_2 -norm:

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

where $f, g \in L_{\mathbb{R}}^2([0, 1])$. These authors also provide sufficient conditions for an $L_{\mathbb{R}}^2([0, 1])$ -valued stochastic process $\{X_{i,n}\}$ to satisfy (2.3) with $d(f, g) = \|f - g\|_2$ and $\rho = 2$. Additionally, they define $L_E^p(T, \mu)$ as the Banach space of all strongly measurable functions $f : T \rightarrow E$ with finite norm:

$$\|f\|_p = \|f\|_{L_E^p(T, \mu)} = \left(\int \|f(\tau)\|_E^p d\mu(\tau) \right)^{\frac{1}{p}},$$

for $1 \leq p < \infty$ and with finite norm

$$\|f\|_\infty = \|f\|_{L_E^\infty(T, \mu)} = \inf_{\mu(N)=0} \sup_{\tau \in T \setminus N} \|f(\tau)\|_E,$$

for $p = \infty$. In this context, $\mathcal{H}_{\mathbb{C}} = L_{\mathbb{C}}^2([0, 1])$.

Remark 2.2. [176] generalizes the definition of local stationary processes, initially proposed by [56], 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 $\mathcal{H}_{\mathbb{C}}$;

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

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

They have proved in [176, Proposition 2.2] that:

Proposition 2.3. Suppose that assumptions (A1) and (A2) hold. Then, $\{X_{i,n}\}$ is a locally stationary process in \mathcal{H} .

In our investigation of non-stationary processes characterized by gradually evolving dynamics over time, exhibiting behaviors reminiscent of stationarity at a localized scale, we draw upon an extensive body of research documented in notable references such as [57, 105, 115, 131, 141, 185]. As an illustrative example, let $a : [0, 1] \rightarrow \mathbb{R}$ be a continuous function, and consider a sequence of independent and identically distributed random variables $(\varepsilon_i)_{i \in \mathbb{N}}$. The stochastic process $X_{i,n} = a(i/n) + \varepsilon_i$, where $i \in \{1, \dots, n\}$ and $n \in \mathbb{N}$, may demonstrate "almost" stationary behavior for i in proximity to a specific point i^* (e.g., i^* in $\{1, \dots, n\}$), under the condition that $a(i^*/n) \approx a(i/n)$. However, it is important to note that this process does not strictly adhere to weak stationarity. To capture this gradual transition, the concept of local stationarity was introduced by [57], wherein the spectral representation of the underlying stochastic process is locally approximated. Within our framework, the process $\{X_{i,n}\}$ can be stochastically approximated by a stationary process $\{X_{i,n}^{(u)}\}$ around each rescaled time point u , particularly for those values of i where $i/n - u$ is small. Given our emphasis on functional data, we define a functional time series as locally stationary if it can be approximated locally by a stationary functional time series. The subsequent section will present a standard definition of local stationarity, building upon the foundation laid out in [176].

Theorem 2.4. Consider a white noise process $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ in $L_H^2(\Omega, \mathbb{P})$ where $H = L^2([0, 1])$, and let $\{X_{i,n}\}$ be a sequence of functional autoregressive processes defined as

$$\sum_{j=0}^m B_{\frac{i}{n}, j}^{i,n} (X_{i-j,n}) = C_{\frac{i}{n}}(\varepsilon_i),$$

with $B_{u,j} = B_{0,j}, C_u = C_0$ for $u < 0$, and $B_{u,j} = B_{1,j}, C_u = C_1$ for $u > 1$. If the process satisfies, for all $u \in [0, 1]$ and $p = 2$ or $p = \infty$, the conditions

- (i) C_u is an invertible element of $S_\infty(H)$;
- (ii) $B_{u,j} \in S_p(\mathcal{H})$ for $j = 1, \dots, m$ with $\sum_{j=1}^m \|B_{u,j}\|_l < 1$ and $B_{u,0} = I_H$;
- (iii) the mappings $u \mapsto B_{u,j}$ for $j = 1, \dots, m$ and $u \mapsto C_u$ are continuous in $u \in [0, 1]$ and differentiable on $u \in (0, 1)$ with bounded derivatives,

then the process $\{X_{i,n}\}$ satisfies (A2) with

$$\mathcal{A}_{\frac{i}{n}, \omega}^{(n)} = \frac{1}{\sqrt{2\pi}} \left(\sum_{j=0}^m e^{-i\omega j} B_{\frac{i}{n}, j} \right)^{-1} C_{\frac{i}{n}},$$

and, consequently, is locally stationary.

2.4. Small-ball probability

Handling infinite-dimensional spaces presents a notable technical hurdle due to the absence of a universal reference measure, such as the Lebesgue measure. Consequently, defining a density function for the functional variable becomes elusive. To surmount this challenge, we leverage the concept of “small-ball probability”. In particular, we address the concentration of the probability measure for the functional variable within a small-ball using the function $\phi_x(\cdot)$. For a fixed $x \in \mathcal{H}$ and every $r > 0$, the function $\phi_{x,\theta}(r)$ is defined as:

$$\mathbb{P}(X \in B_\theta(x, r)) =: \phi_{x,\theta}(r) > 0. \quad (2.4)$$

Here, \mathcal{H} is equipped with the semi-metric $d(\cdot, \cdot)$, and $B_\theta(x, r)$ represents a ball in \mathcal{H} with center $x \in \mathcal{H}$ and radius r . For $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{H}^m$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \Theta^m$, we define

$$\phi_{\mathbf{x}, \boldsymbol{\theta}}(r) = \prod_{i=1}^m \phi_{x_i, \theta_i}(r).$$

Further elucidation and examples concerning small-ball probability can be explored in Remark 3.6.

2.5. Mixing conditions

Statistical observations commonly exhibit a certain degree of dependence rather than complete independence. The concept of *mixing* serves as a quantitative measure of the proximity of a sequence of random variables to independence, facilitating the extension of traditional results applicable to independent sequences to sequences that are weakly dependent or mixing. The development of *the theory of mixing conditions* has emerged from the recognition that time series manifest “asymptotic independence” properties, thereby facilitating their analysis and statistical inference. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $Z_{1,n}, Z_{2,n}, \dots$ be a sequence of random variables on a probability space $(\Omega, \mathcal{D}, \mathbb{P})$. For an array $\{Z_{i,n} : 1 \leq i \leq n\}$, the coefficients are defined as

$$\beta(k) = \sup_{i,n: 1 \leq i \leq n-k} \beta \left(\sigma(Z_{s,n}, 1 \leq s \leq i), \sigma(Z_{s,n}, i+k \leq s \leq n) \right),$$

where $\sigma(Z)$ represents the σ -field generated by Z . The array $\{Z_{i,n}\}$ is considered β -mixing if $\beta(k) \rightarrow 0$. It is crucial to note that β -mixing implies α -mixing. Throughout the ensuing discussion, we assume that the sequence of random elements $\{(X_{i,n}, Y_{i,n}), i = 1, \dots, n; n \geq 1\}$ is absolutely regular. Remarkably, Markov chains exhibit β -mixing under the milder Harris recurrence condition, provided the underlying space is finite [62]. Additional rationale for favoring regular processes over strongly mixing processes is provided in the concluding remarks (Section 7).

2.6. Kernel estimation

We endeavor to estimate the regression function as denoted in (2.1). The kernel estimator is formally defined as

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \varphi(Y_{\mathbf{i},n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\}}, \quad (2.5)$$

where $K_1(\cdot)$ and $K_2(\cdot)$ denote one-dimensional kernel functions. Here, let $h = h_n$ be a bandwidth with the property that $h \rightarrow 0$ as $n \rightarrow \infty$. The function $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ is symmetric and measurable, belonging to a class of functions denoted as \mathcal{F}_m . Importantly, this estimator is a conditional U -statistic utilizing the sequence of random variables $\{Y_{i,n}, X_{i,n}\}_{i=1}^n$ and the kernel $\varphi \times K_1 \times K_2$. The introduction of such statistics was pioneered by [166]. To investigate the weak convergence of the conditional empirical process and the conditional U -process within the functional data framework, we introduce some necessary notations. Consider the class

$$\mathcal{F}_m = \{\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}\},$$

which consists of real-valued symmetric measurable functions on \mathcal{Y}^m with a measurable envelope function :

$$F(\mathbf{y}) \geq \sup_{\varphi \in \mathcal{F}_m} |\varphi(\mathbf{y})|, \text{ for } \mathbf{y} \in \mathcal{Y}^m. \quad (2.6)$$

For kernel functions $K_1(\cdot)$ and $K_2(\cdot)$, as well as a subset $S_{\mathcal{H}} \subset \mathcal{H}$, we define the pointwise measurable class of functions for $1 \leq m \leq n$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$:

$$\mathcal{H}_{\boldsymbol{\theta}}^m := \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m K_1 \left(\frac{u_k - \cdot}{h_n} \right) K_2 \left(\frac{d_{\theta_i}(x_i, \cdot)}{h_i} \right), (\mathbf{x}, \mathbf{u}) \in \mathcal{H}^m \times [0, 1]^m \right\}$$

and

$$\mathfrak{S}_{\boldsymbol{\theta}}^m := \bigcup_{\boldsymbol{\theta} \in \Theta^m} \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m K_1 \left(\frac{u_k - \cdot}{h_n} \right) K_2 \left(\frac{d_{\theta_i}(x_i, \cdot)}{h_i} \right), (\mathbf{x}, \mathbf{u}) \in \mathcal{H}^m \times [0, 1]^m \right\}.$$

The conditional U -process indexed by $\mathcal{F}_m \mathfrak{S}_{\boldsymbol{\theta}}^m$ is defined by

$$\left\{ \mathbb{G}_n(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) := \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} \left(\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right) \right\}_{\mathcal{F}_m \mathfrak{S}_{\boldsymbol{\theta}}^m}. \quad (2.7)$$

Observing the importance of point-wise measurability in our context, it enables us to state our results conventionally, adhering to the classical definition of probability, without invoking the abstract notions of outer probability or outer expectation [178].

Remark 2.5. *The bandwidth h remains consistent across all directions, simplifying the analysis in cases involving product kernels. Nevertheless, the results can be readily adapted to scenarios with non-product kernels and varying bandwidths.*

Remark 2.6. *Our estimator differs from the conventional conditional U -statistics not only in the nature of the sequence $\{X_i\}_i$ but also in the inclusion of a kernel in the time direction. As a result, we attain smoothness from both the covariate direction $(X_{i,n})$ and the temporal dimension, allowing us to capture the characteristics of a regression model evolving over time.*

2.7. VC-type classes of functions

Examining functional data through asymptotic methods involves delving into concentration properties elucidated by the concept of small-ball probability. When scrutinizing a process characterized by a set of functions, it becomes imperative to consider additional topological concepts, including metric entropy and VC-subgraph classes (referred to as “VC”, inspired by Vapnik and Cervonenkis).

Definition 2.7. *Let $\mathcal{S}_{\mathcal{E}}$ be a subset of a semi-metric space \mathcal{E} . A finite set of points $\{e_1, \dots, e_N\} \subset \mathcal{E}$ is considered a ε -net of $\mathcal{S}_{\mathcal{E}}$ for a given $\varepsilon > 0$ if:*

$$\mathcal{S}_{\mathcal{E}} \subseteq \bigcup_{j=1}^N B(e_j, \varepsilon).$$

If $N_{\varepsilon}(\mathcal{S}_{\mathcal{E}})$ is the cardinality of the smallest ε -net (i.e., the minimal number of open balls of radius ε) in \mathcal{E} needed to cover $\mathcal{S}_{\mathcal{H}}$, then the Kolmogorov’s entropy (metric entropy) of the set $\mathcal{S}_{\mathcal{E}}$ is defined as the quantity:

$$\psi_{\mathcal{S}_{\mathcal{E}}}(\varepsilon) := \log N_{\varepsilon}(\mathcal{S}_{\mathcal{E}}).$$

Kolmogorov introduced the concept of metric entropy, extensively explored in various metric spaces, as indicated by its name (cf. [112]). Dudley ([70]) utilized this concept to establish sufficient conditions for the continuity of Gaussian processes, forming the foundation for significant generalizations of Donsker’s theorem regarding the weak convergence of the empirical process. Consider two subsets, $\mathcal{B}_{\mathcal{H}}$ and $\mathcal{S}_{\mathcal{H}}$, in the semi-metric space \mathcal{H} with Kolmogorov’s entropy (for radius ε) denoted as $\psi_{\mathcal{B}_{\mathcal{H}}}(\varepsilon)$ and $\psi_{\mathcal{S}_{\mathcal{H}}}(\varepsilon)$, respectively. The Kolmogorov entropy for the subset $\mathcal{B}_{\mathcal{H}} \times \mathcal{S}_{\mathcal{H}}$ of the semi-metric space \mathcal{H}^2 is given by:

$$\psi_{\mathcal{B}_{\mathcal{H}} \times \mathcal{S}_{\mathcal{H}}}(\varepsilon) = \psi_{\mathcal{B}_{\mathcal{H}}}(\varepsilon) + \psi_{\mathcal{S}_{\mathcal{H}}}(\varepsilon).$$

Thus, $m\psi_{\mathcal{S}_{\mathcal{H}}}(\varepsilon)$ represents the Kolmogorov entropy of the subset $\mathcal{S}_{\mathcal{H}}^m$ in the semi-metric space \mathcal{H}^m . If d denotes the semi-metric on \mathcal{H} , a semi-metric on \mathcal{H}^m can be defined as:

$$d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z}) := \frac{1}{m} d_{\theta_1}(x_1, z_1) + \dots + \frac{1}{m} d_{\theta_m}(x_m, z_m), \quad (2.8)$$

for $\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{z} = (z_1, \dots, z_m) \in \mathcal{H}^m$. The choice of the semi-metric is crucial in this type of analysis, and readers can find insightful discussions on this topic in [78, Chapters 3 and 11]. Furthermore, in this context, we must also address another topological concept: VC-subgraph classes.

Definition 2.8. A class of subsets C on a set C is called a VC-class if there exists a polynomial $P(\cdot)$ such that, for every set of N points in C , the class C picks out at most $P(N)$ distinct subsets.

Definition 2.9. A class of functions \mathcal{F} is called a VC-subgraph class if the graphs of the functions in \mathcal{F} form a VC-class of sets. In other words, if we define the subgraph of a real-valued function f on S as the following subset \mathcal{G}_f on $S \times \mathbb{R}$:

$$\mathcal{G}_f = \{(s, t) : 0 \leq t \leq f(s) \text{ or } f(s) \leq t \leq 0\}$$

the class $\{\mathcal{G}_f : f \in \mathcal{F}\}$ is a VC-class of sets on $S \times \mathbb{R}$. Informally, a VC-class of functions is characterized by having a polynomial covering number (the minimal number of required functions to make a covering on the entire class of functions).

A VC-class of functions \mathcal{F} with an envelope function F has the following entropy property. For a given $1 \leq q < \infty$, there exist constants a and b such that:

$$N(\epsilon, \mathcal{F}, \|\cdot\|_{L_q(Q)}) \leq a \left(\frac{(QF^q)^{1/q}}{\epsilon} \right)^b, \quad (2.9)$$

for any $\epsilon > 0$ and each probability measure such that $QF^q < \infty$. Several references provide sufficient conditions under which (2.9) holds, such as [144, Lemma 22], [72, §4.7.], [178, Theorem 2.6.7], [114, §9.1], [65, §3.2], [33, 34, 45], offering further discussions.

2.8. Assumptions

For the reader's convenience, we have compiled the essential assumptions as follows:

Assumption 1. [Model and distribution assumptions]

i) The process $\{X_{i,n}\}$ is locally stationary and satisfies that for each time point $u \in [0, 1]$, there exists a stationary process $\{X_i^{(u)}\}$ such that

$$d_{\theta_i}(X_{i,n}, X_i^{(u)}) \leq \left(\left| \frac{i}{n} - u \right| + \frac{1}{n} \right) U_{i,n}^{(u)} \text{ a.s.},$$

with $\mathbb{E}[(U_{i,n}^{(u)})^\rho] < C$ for some $\rho > 0$, $C < \infty$.

ii) Let $B(x, h)$ be a ball centered at $x \in \mathcal{H}$ with radius h , defined in Section 2.4, and let $c_d < C_d$ be positive constants. For all $\mathbf{u} \in [0, 1]^m$,

$$0 < c_d \phi^m(h_n) f_1(\mathbf{x}) \leq \mathbb{P} \left((X_{i_1}^{(u_1)}, \dots, X_{i_m}^{(u_m)}) \in \mathbf{B}_\theta(\mathbf{x}, h) \right) =: F_{\mathbf{u}, \theta}(h; \mathbf{x}) \leq C_d \phi^m(h_n) f_1(\mathbf{x}), \quad (2.10)$$

where $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin, $f_1(x)$ is a non-negative functional in $x \in \mathcal{H}$, and

$$\mathbf{B}_\theta(\mathbf{x}, h) = \prod_{i=1}^m B_{\theta_i}(x_i, h).$$

iii) There exist constants $C_\phi > 0$ and $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$,

$$\int_0^\varepsilon \phi(u) du > C_\phi \varepsilon \phi(\varepsilon). \quad (2.11)$$

iv) Let $\psi(h_n) \rightarrow 0$ as $h \rightarrow 0$, and $f_2(\mathbf{x})$ is a non-negative functional in $\mathbf{x} := (x_1, \dots, x_m) \in \mathcal{H}^m$

$$\sup_{\mathbf{i} \in I_n^m} \mathbb{P} \left(\left((X_{i_1, n}, \dots, X_{i_m, n}), (X'_{i_1, n}, \dots, X'_{i_m, n}) \right) \in \mathbf{B}_\theta(\mathbf{x}, h) \times \mathbf{B}_\theta(\mathbf{x}, h) \right) \leq \psi^m(h_n) f_2(\mathbf{x}).$$

We will also assume that the ratio $\psi(h_n)/\phi^2(h_n)$ is bounded.

Assumption 2. [Kernel assumptions]

i) $K_1(\cdot)$ is a symmetric kernel around zero, bounded, and possesses compact support, i.e., $K_1(v) = 0$ for all $|v| > C_1$ for some $C_1 < \infty$. Additionally,

$$\int K_1(z) dz = 1$$

and $K_1(\cdot)$ is Lipschitz continuous, i.e.,

$$|K_1(v_1) - K_1(v_2)| \leq C_2 |v_1 - v_2|$$

for some $C_2 < \infty$ and all $v_1, v_2 \in \mathbb{R}$.

ii) The kernel $K_2(\cdot)$ is non-negative, bounded, and has a compact support in $[0, 1]$, such that $0 < K_2(0)$ and $K_2(1) = 0$. Alternatively, $K_2(\cdot)$ can be viewed as an asymmetrical triangular kernel, i.e., $K_2(x) = (1 - x)\mathbb{1}_{(x \in [0, 1])}$, and $K_2(\cdot)$ is Lipschitz continuous, i.e.,

$$|K_2(v_1) - K_2(v_2)| \leq C_2 |v_1 - v_2|.$$

Moreover, $K'_2(v) = dK_2(v)/dv$ exists on $[0, 1]$, and for two real constants $-\infty < C'_1 < C'_2 < 0$, we have:

$$C'_2 \leq K'_2(v) \leq C'_1.$$

Assumption 3. [Smoothness]

i) $r^{(m)}(\mathbf{u}, \mathbf{x})$ is twice continuously partially differentiable with respect to \mathbf{u} . We also assume that

$$\sup_{\mathbf{u}_1, \mathbf{u}_2 \in [0, 1]^m} \left| r^{(m)}(\mathbf{u}_1, \mathbf{x}, \boldsymbol{\theta}) - r^{(m)}(\mathbf{u}_2, \mathbf{z}, \boldsymbol{\theta}) \right| \leq c_m (d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z})^\alpha + \|\mathbf{u}_1 - \mathbf{u}_2\|^\alpha) \quad (2.12)$$

for some $c_m > 0$, $\alpha > 0$ and $\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{z} = (z_1, \dots, z_m) \in \mathcal{H}^m$.

ii) $\sigma : [0, 1] \times \mathcal{H}^m \rightarrow \mathbb{R}$ is bounded by some constant $C_\sigma < \infty$ from above and by some constant $c_\sigma > 0$ from below, that is, for all \mathbf{u} and \mathbf{x} ,

$$0 < c_\sigma \leq \sigma(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \leq C_\sigma < \infty.$$

iii) $\sigma(\cdot, \cdot, \cdot)$ is Lipschitz continuous with respect to \mathbf{u} .

iv) $\sup_{\mathbf{u} \in [0, 1]^m} \sup_{z: d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z}) \leq \varepsilon} |\sigma(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \sigma(\boldsymbol{\theta}, \mathbf{u}, \mathbf{z})| = o(1)$ as $\varepsilon \rightarrow 0$.

Let $\mathfrak{B}_{\mathbf{i}, \varphi, n}$ be an array of one-dimensional random variables. In this study, this array will be equal to $\mathfrak{B}_{\mathbf{i}, \varphi, n} = 1$ or $\mathfrak{B}_{\mathbf{i}, \varphi, n} = \varepsilon_{\mathbf{i}, n}$.

Assumption 4. [Mixing]

i) For $\zeta > 2$ and $C < \infty$, we have

$$\sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E} |\mathfrak{B}_{i,n}|^\zeta \leq C,$$

and

$$\sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E} \left[|\mathfrak{B}_{i,n}|^\zeta \mid X_{i,n} = \mathbf{x} \right] \leq C.$$

ii) The β -mixing coefficients of the array $\{X_{i,n}, \mathfrak{B}_{i,n}\}$ satisfy $\beta(k) \leq Ak^{-\gamma}$ for some $A > 0$ and $\gamma > 2$. Additionally, we assume that $\delta + 1 < \gamma(1 - \frac{2}{\nu})$ for some $\nu > 2$ and $\delta > 1 - \frac{2}{\nu}$, along with the condition

$$h^{2(1 \wedge \alpha) - 1} \left(\phi(h_n) a_n + \sum_{k=a_n}^{\infty} k^\delta (\beta(k))^{1 - \frac{2}{\nu}} \right) \rightarrow 0, \quad (2.13)$$

as $n \rightarrow \infty$, where $a_n = \left\lceil (\phi(h_n))^{-(1 - \frac{2}{\nu})/\delta} \right\rceil$ and for all $\alpha > 0$.

iii) For some $\zeta_0 > 0$, as $n \rightarrow \infty$, we have

$$\frac{(\log n)^{-\frac{m+\gamma+1}{2} + \zeta_0(\gamma+1)}}{n^{\frac{-m+\gamma+1}{2} - 1 - \frac{\gamma+1}{\zeta}} h_n^{\frac{m+\gamma+1}{2}} \phi(h_n)^{\frac{-m+\gamma+1}{2}}} \rightarrow 0.$$

iv) Both nh_n^{2m+1} and $nh^m \phi(h_n)^m$ tend to infinity as n goes to infinity.

Assumption 5. [Blocking assumptions] There exists a sequence of positive integers $\{v_n\}$ satisfying $v_n \rightarrow \infty$, $v_n = o(\sqrt{nh\phi(h_n)})$ and $\sqrt{\frac{n}{h\phi(h_n)}} \beta(v_n) \rightarrow \infty$ as $n \rightarrow \infty$.

Assumption 6. [Class of functions assumptions]

The classes of functions \mathfrak{S}_Θ^m and \mathcal{F}_m are such that :

i) The class of functions \mathcal{F}_m is bounded and its envelope function satisfies for some $0 < M < \infty$:

$$F(\mathbf{y}) \leq M, \quad \mathbf{y} \in \mathcal{Y}^m.$$

ii) The class of functions $\mathcal{F}_m \mathfrak{S}_\Theta^m$ is supposed to be of VC-type with envelope function previously defined. Hence, there are two finite constants b and ν such that:

$$N(\epsilon, \mathcal{F}_m \mathfrak{S}_\Theta^m, \|\cdot\|_{L_2(Q)}) \leq \left(\frac{b \|F\|_{L_2(Q)}}{\epsilon} \right)^\nu$$

for any $\epsilon, \nu > 0$ and each probability measure such that $Q(F)^2 < \infty$.

iii) The class of functions \mathcal{F}_m is unbounded and its envelope function satisfies for some $\zeta > 2$:

$$\theta_\zeta := \sup_{\mathbf{x} \in \mathcal{S}_\mathcal{H}^m} \mathbb{E} \left(F^\zeta(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x} \right) < \infty, \quad \mathcal{S}^m \subset \mathcal{H}^m.$$

iv) The metric entropy of the class $\mathcal{F} \mathcal{K}$ satisfies, for some $1 \leq \zeta < \infty$:

$$\int_0^\infty (\log N(u, \mathcal{F} \mathcal{K}, \|\cdot\|_\zeta))^{\frac{1}{2}} du < \infty.$$

2.9. Comments on the assumptions

To establish the groundwork for our analysis, we draw inspiration from seminal works such as [78, 87, 116, 133, 179]. Our assumptions play a pivotal role in shaping the properties of the random processes under consideration. Starting with Assumption 1, we formalize the local stationarity property of X_i and introduce conditions related to the distribution behavior of the variables. Equation (2.10) governs the small-ball probability around zero, representing the standard condition for small-ball probability. This equation implies that the small-ball probability can be approximated as the product of two independent functions $\phi^m(\cdot)$ and $f_1(\cdot)$. For instance, when $m = 1$, references such as [135] (diffusion process), [25] (Gaussian measure), [122] (general Gaussian process), and [133] (strongly mixing processes) provide context. The function $\phi(\cdot)$ can take various forms, such as $\phi(\epsilon) = \epsilon^\delta \exp(-C/\epsilon^a)$ for Ornstein-Uhlenbeck and general diffusion processes. Further examples and discussions can be found in [81] and Remark 3.6. Assumption 1 iv) details the behavior of the joint distribution near the origin, aligning with assumptions made by [87] in the context of density estimation for functional data. Assumption 2 encompasses the kernel assumptions commonly used in nonparametric functional estimation. Notably, the Parzen symmetric kernel is inadequate due to the positivity of the random process $D_i = d(x, X_i)$; thus, $K_2(\cdot)$ with support $[0, 1]$ is considered. The kernel $K_2(\cdot)$ is a symmetric type II kernel belonging to the family of continuous kernels (triangle, quadratic, etc.). Compact support on $[0, 1]$ is assumed for the kernels to derive an expression for the asymptotic variance. The Lipschitz-type assumptions on $K_2(\cdot)$ and $\sigma(\cdot, \cdot)$ (Assumption 2ii) and Assumption 3iii)) are crucial for obtaining the convergence rate. Assumption 3 restricts the growth of $r^{(m)}(\cdot)$ and $\sigma(\cdot)$ and places bounds on these functions to prevent rapid growth outside a large bound. It is tailored to ensure the convergence rate and forms an integral part of the overall analysis. Assumption 4 ii) is a standard mixing condition necessary for establishing asymptotic normality and the asymptotic negligibility of the bias, consistent with [133]. The variables $\mathfrak{B}_{i,n}$ are not necessarily bounded, and there is a tradeoff between the decay of the mixing condition and the order ζ of the moment $\sup_{x \in \mathcal{H}^m} \mathbb{E}|\mathfrak{B}_{i,n}|^\zeta \leq C$. Assumption 4 iii) and iv) are technical conditions crucial for obtaining the desired results, addressing the uniform convergence rate and the bias and convergence rate of the general estimator. Assumption 6 asserts that the class of functions satisfies certain entropy conditions. Part ii) and iii) are interconnected, with the former stating that the class is bounded. However, in the context of proving the functional central limit theorem for conditional U -processes indexed by an unbounded class of functions, part iii) supersedes the first one. Assumption 6 ii) ensures that \mathcal{F} is VC type with characteristics b and n for the envelope $F\kappa^m$. Since $F \in L^2(\mathbb{P})$ by Assumption 6, Dudley's criterion on the sample continuity of Gaussian processes implies that the function class \mathcal{F} is \mathbb{P} -pre-Gaussian. These general assumptions, inclusive of the mentioned conditions, provide adequate flexibility given the diverse components in our main results. They encapsulate and leverage the topological structure of functional variables, the probability measure within the functional space, the concept of measurability applied to the function class, and the uniformity regulated by entropy properties.

Remark 2.10. *It is worth noting that Assumption 6 iii) can be replaced by more general hypotheses regarding the moments of Y , as discussed in [65]. The alternative assumption takes the following form:*

iii)' We introduce $\{\mathcal{M}(x) : x \geq 0\}$ as a non-negative continuous function, increasing on $[0, \infty)$, and

such that, for some $s > 2$, eventually as $x \rightarrow \infty$:

$$x^{-s} \mathcal{M}(x) \downarrow; \quad x^{-1} \mathcal{M}(x) \uparrow. \quad (2.14)$$

For each $t \geq \mathcal{M}(0)$, we define $\mathcal{M}^{\text{inv}}(t) \geq 0$ such that $\mathcal{M}(\mathcal{M}^{\text{inv}}(t)) = t$. Additionally, we assume that:

$$\mathbb{E}(\mathcal{M}(|F(Y)|)) < \infty.$$

The following choices for $\mathcal{M}(\cdot)$ are particularly interesting:

- (i) $\mathcal{M}(x) = x^\xi$ for some $\xi > 2$;
- (ii) $\mathcal{M}(x) = \exp(sx)$ for some $s > 0$.

These alternative formulations provide broader flexibility in defining the moments of Y , accommodating various scenarios and enhancing the applicability of the analysis.

3. Uniform convergence rates for kernel estimators

Before expressing the asymptotic behavior of our estimator represented in (2.5), we will generalize the study to a U -statistic estimator defined by:

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) = \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\} \mathfrak{B}_{\mathbf{i}, \varphi, n}, \quad (3.1)$$

where $\mathfrak{B}_{\mathbf{i}, \varphi, n}$ is an array of one-dimensional random variables. In this study, we use the results with $\mathfrak{B}_{\mathbf{i}, \varphi, n} = 1$ and $\mathfrak{B}_{\mathbf{i}, \varphi, n} = \varepsilon_{\mathbf{i}, n}$.

3.1. Hoeffding's decomposition

Note that $\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi)$ is a classical U -statistic with a kernel depending on n . We define

$$\xi_k := \frac{1}{h} K_1 \left(\frac{u_k - k/n}{h_n} \right),$$

$$H(Z_1, \dots, Z_m) := \prod_{k=1}^m \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_k}(x_k, X_{k, n})}{h_n} \right) \mathfrak{B}_{\mathbf{i}, \varphi, n},$$

thus, the U -statistic in (3.1) can be viewed as a weighted U -statistic of degree m :

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) = \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \dots \xi_{i_m} H(Z_{i_1}, \dots, Z_{i_m}). \quad (3.2)$$

We can write Hoeffding's decomposition in this case as in [94]. If we do not assume symmetry for $\mathfrak{B}_{\mathbf{i}, \varphi, n}$ or H , we must define:

- The expectation of $H(Z_{i_1}, \dots, Z_{i_m})$:

$$\vartheta(\mathbf{i}) := \mathbb{E}H(Z_{i_1}, \dots, Z_{i_m}) = \int \mathfrak{B}_{\mathbf{i}, \varphi, n} \prod_{k=1}^m \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_k}(x_k, v_{k, n})}{h_n} \right) d\mathbb{P}_{\mathbf{i}}(z_{\mathbf{i}}). \quad (3.3)$$

- For all $\ell \in \{1, \dots, m\}$ the position of the argument, construct the function π_ℓ such that:

$$\pi_\ell(z; z_1, \dots, z_{m-1}) := (z_1, \dots, z_{\ell-1}, z, z_\ell, \dots, z_{m-1}). \quad (3.4)$$

- Define:

$$H^{(\ell)}(z; z_1, \dots, z_{m-1}) := H\{\pi_\ell(z; z_1, \dots, z_{m-1})\} \quad (3.5)$$

$$\vartheta^{(\ell)}(i; i_1, \dots, i_{m-1}) := \vartheta\{\pi_\ell(i; i_1, \dots, i_{m-1})\}. \quad (3.6)$$

Hence, the first-order expansion of $H(\cdot)$ will be seen as:

$$\begin{aligned} \widetilde{H}^{(\ell)}(z) &:= \mathbb{E}\{H^{(\ell)}(z, Z_1, \dots, Z_{m-1})\} \\ &= \int \mathfrak{B}_{(1, \dots, \ell-1, i, \ell, \dots, m-1)} \prod_{\substack{k=1 \\ k \neq i}}^{m-1} \frac{1}{\phi(h_n)} K_2\left(\frac{d_{\theta_k}(x_k, v_k)}{h_n}\right) \times \frac{1}{\phi(h_n)} K_2\left(\frac{d_{\theta_i}(x_i, v_i)}{h_n}\right) \\ &\quad \mathbb{P}(dv_1, \dots, dv_{\ell-1}, dv_\ell, \dots, dv_{m-1}) \\ &= \frac{1}{\phi(h_n)} K_2\left(\frac{d_{\theta_i}(x_i, x)}{h_n}\right) \times w \times \int \mathfrak{B}_{(1, \dots, \ell-1, \ell, \dots, m-1)} \prod_{\substack{k=1 \\ k \neq i}}^{m-1} \frac{1}{\phi(h_n)} K_2\left(\frac{d_{\theta_k}(x_k, v_k)}{h_n}\right) \\ &\quad \mathbb{P}(dv_1, \dots, dv_{\ell-1}, dv_\ell, \dots, dv_{m-1}), \end{aligned} \quad (3.7)$$

with \mathbb{P} as the underlying probability measure, and define

$$f_{i, i_1, \dots, i_{m-1}}^{(\ell)} := \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} (\widetilde{H}^{(\ell)}(z) - \vartheta^{(\ell)}(i; i_1, \dots, i_{m-1})). \quad (3.8)$$

Then, the first-order projection can be defined as:

$$\widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) := \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} f_{i, i_1, \dots, i_{m-1}}^{(\ell)}, \quad (3.9)$$

where

$$I_{n-1}^{m-1}(-i) := \{1 \leq i_1 < \dots < i_{m-1} \leq n \text{ and } i_j \neq i \text{ for all } j \in \{1, \dots, m-1\}\}.$$

For the remainder terms, we denote by $\mathbf{i} \setminus i_\ell := (i_1, \dots, i_{\ell-1}, i_{\ell+1}, \dots, i_m)$ and for $\ell \in \{1, \dots, m\}$, let

$$H_{2,i}(\mathbf{z}) := H(\mathbf{z}) - \sum_{\ell=1}^m \widetilde{H}_{\mathbf{i} \setminus i_\ell}^{(\ell)}(z_\ell) + (m-1)\vartheta(\mathbf{i}), \quad (3.10)$$

where

$$\widetilde{H}_{\mathbf{i} \setminus i_\ell}^{(\ell)}(z_\ell) = \mathbb{E}\{H(Z_1, \dots, Z_{\ell-1}, z, Z_{\ell+1}, \dots, Z_{m-1})\},$$

defined in (3.7), this projection derives us to the following remainder term:

$$\widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) := \frac{(n-m)!}{(n)!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H_{2,i}(\mathbf{z}). \quad (3.11)$$

Finally, using Eqs (3.9) and (3.11), and under conditions that :

$$\mathbb{E}\{\widehat{H}_{1,i}(\mathbf{u}, X, \boldsymbol{\theta}, \varphi)\} = 0, \quad (3.12)$$

$$\mathbb{E}\{H_{2,i}(\mathbf{Z} | Z_k)\} = 0 \text{ a.s.}, \quad (3.13)$$

we get the [99] decomposition:

$$\begin{aligned} & \widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) - \mathbb{E}\{\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi)\} \\ &= \frac{1}{n} \sum_{i=1}^n \widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) + \widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) \\ &=: \widehat{\psi}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) + \widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi). \end{aligned} \quad (3.14)$$

For more details, the interested reader can refer to [94, Lemma 2.2].

3.2. Uniform convergence rate

We commence by presenting the following general proposition.

Proposition 3.1. *Let $\mathcal{F}_m \mathfrak{S}_{\Theta}^m$ denote a measurable VC-subgraph class of functions, adhering to Assumption 6. Suppose that Assumptions 1–4 are satisfied. In such case, the ensuing result holds:*

$$\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [0,1]^m} |\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) - \mathbb{E}[\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi)]| = O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh_n^m \phi^m(h_n)}} \right).$$

The proof of Proposition 3.1 is deferred to Section 4.1.

Remark 3.2. *Elaborating on Proposition 3.1, we can delve into the uniform convergence rate of the kernel estimator $\widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n)$. It is crucial to emphasize that when $m = 1$ and the function φ remains constant, the outcomes align with the pointwise convergence rate of the regression function for a strictly stationary functional time series, as discussed in [78].*

The subsequent theorem presents the uniform convergence rate of the kernel estimator (2.5).

Theorem 3.3. *Let $\mathcal{F}_m \mathfrak{S}_{\Theta}^m$ be a measurable VC-subgraph class of functions complying with Assumption 6. Suppose Assumptions 1–4 are fulfilled. Then, we have:*

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |\widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})| \\ &= O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh_n^m \phi^m(h_n)}} + h_n^{2m \wedge \alpha} \right). \end{aligned} \quad (3.15)$$

The proof of Theorem 3.15 is postponed until Section 4.1.

Remark 3.4. *It is possible to consider the setting of $\Theta = \Theta_n$ and assume that*

$$\text{card}(\Theta_n) = n^{\alpha} \quad \text{with} \quad \alpha > 0$$

and

$$\forall \boldsymbol{\theta} \in \Theta_n, \quad \langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, \boldsymbol{\theta} - \boldsymbol{\theta}_0 \rangle^{1/2} \leq C_7 b_n,$$

where b_n tends to zero, as in [145].

Remark 3.5. In contrast to Theorem 4.2 in [179] and akin to Theorem 3.1 in [116], our formulation excludes the bias term arising from the approximation error of $X_{i,n}$ by $X_i^{(u)}$. Under our assumptions, the approximation error is negligibly small compared to $h_n^{2m\wedge\beta}$.

Remark 3.6. In nonparametric problems, the infinite dimensionality of the target function is typically determined by the smoothness condition, specifically in Assumptions 2 i) and 3 i). This primarily affects the bias component of the convergence rates, represented by terms like $O(h_n^{2m\wedge\alpha})$ in Theorem 3.3. Other terms in the convergence rates stem directly from dispersion effects and are inherently linked to the concentration properties of the probability measure of the variable X . These terms can be expressed as

$$O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh_n^m\phi^m(h_n)}}\right),$$

where small-ball probabilities are quantified and controlled by the mean of the function $\phi(\cdot)$ defined in (2.4). The rate of convergence is influenced by the concentration of the measure of the process X ; less concentration leads to a slower rate of convergence. Unfortunately, solutions to $\mathbb{P}(X \in B(x, r))$ are known for very few random variables (or processes) X , even when $x = 0$. In certain functional spaces, considering $x \neq 0$ introduces considerable difficulties that might not be surmountable. Authors often focus on Gaussian random elements, and for a comprehensive overview of main results on small-ball probability, refer to [122]. In many scenarios, it is convenient to assume that

$$\mathbb{P}(X \in B(x, r)) \sim \psi(x)\phi(r) \quad \text{as } r \rightarrow 0, \quad (3.16)$$

where, to ensure the identifiability of the decomposition, a normalizing restriction is necessary, such as

$$\mathbb{E}[\psi(X)] = 1.$$

The factorization (3.16) is not a stringent assumption; it holds under appropriate hypotheses (see, for instance, [27, 122]). The advantage of assuming (3.16) is two-fold. Firstly, the function $\psi(x)$ can be viewed as a surrogate density of the functional random element X and can be leveraged in various contexts. The interested reader can explore its potential in works like [26, 83, 87], where the surrogate density is estimated and used to define a notion of mode or for classification purposes. Second, the function $\phi(h_n)$ acts as the volumetric term and can be employed to assess the complexity of the probability law of the process X (see [28]). In the special multi-(but finite)-dimensional scenario where $\mathbf{X} \in \mathbb{R}^d$, the relation (2.4) is satisfied under standard assumptions with $\phi_x(h_n) \sim C_x h^d$, commonly known as the curse of dimensionality (see [77, 80]). Here, we would instead refer to it as the curse of infinite dimension, specifically highlighting the effects of small-ball probabilities. The inherent nature of these probability effects involving small-balls becomes apparent within our infinite-dimensional framework. The remainder of this remark will focus on applying our approach to various continuous-time processes, where the probabilities associated with small-balls have already been identified. For more details on the following examples, refer to [80].

- (i) Consider the space $C([0, 1], \mathbb{R})$ equipped with the supremum norm, and its associated Cameron-Martin space

$$\mathcal{F} = C([0, 1], \mathbb{R})^{\text{CM}}.$$

Let's examine the fractional Brownian motion ζ^{FBM} with parameter $\delta, 0 < \delta < 2$. small-ball probabilities in this context have been extensively studied. According to [122, Theorems 3.1 and 4.6], we have

$$\forall x_0 \in \mathcal{F}, \quad C'_{x_0} e^{h^{-2/\delta}} \leq \mathbb{P} \left(\|\zeta^{\text{FBM}} - x_0\|_{\infty} \leq h \right) \leq C_{x_0} e^{h^{-2/\delta}}.$$

Notably, our crucial relation (2.4) is trivially satisfied for the fractional Brownian motion by choosing the function $\phi(\cdot)$ in the form

$$\phi_x^{\text{FBM}}(h_n) \sim C_x e^{h^{-2/\delta}}.$$

(ii) Consider a centered Gaussian process $\zeta^{\text{GP}} = \{\zeta_t^{\text{GP}}, 0 \leq t \leq 1\}$. This process can be expressed using the Karhunen-Loève decomposition as follows:

$$\zeta_t^{\text{GP}} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \mathbb{B}_i f_i(t),$$

where λ_i are the eigenvalues of the covariance operator of ζ^{GP} , f_i are the associated orthonormal eigenfunctions, and \mathbb{B}_i are independent standard normal real random variables. For any fixed $k \in \mathbb{N}^*$, let Π_k be the orthogonal projection onto the subspace spanned by the eigenfunctions $\{f_1, \dots, f_k\}$. Define a semi-metric by

$$d^2(x, y) = \int_0^1 [\Pi_k(x - y)(t)]^2 dt.$$

Using the Karhunen-Loève expansion, we obtain

$$d^2(\zeta^{\text{GP}}, x) = \sum_{i=1}^k (\sqrt{\lambda_i} \mathbb{B}_i - x_i)^2 = \sum_{i=1}^k Z_i^2,$$

where $x_i = \int_0^1 x(t) f_i(t) dt$, and Z_i are the components of the vector $Z = (Z_1, \dots, Z_k)$, exhibiting the Euclidean norm structure on \mathbb{R}^k . Due to the independence of Z_i with densities with respect to the Lebesgue measure, we have

$$\mathbb{P}(d^2(\zeta^{\text{GP}}, x) < h) \sim C_x h^k.$$

(iii) Consider the space $C([0, 1], \mathbb{R})$ of continuous real-valued functions on $[0, 1]$, equipped with the supremum norm denoted by $\|\cdot\|_{\infty}$. Let \mathbb{P}^{W} be the Wiener measure on $C([0, 1], \mathbb{R})$, and define the Cameron-Martin space of $C([0, 1], \mathbb{R})$ as

$$\mathcal{F} = C([0, 1], \mathbb{R})^{\text{CM}}.$$

Consider the Ornstein-Uhlenbeck process ζ^{OU} with $\zeta_0^{\text{OU}} = 0$ and

$$d\zeta_t^{\text{OU}} = d\mathbb{B}_t - \frac{1}{2}\zeta_t^{\text{OU}}, \quad \forall t, 0 < t \leq 1.$$

Wiener measures for small centered balls are known to be of the form [[25], p. 187]:

$$\mathbb{P}^{\text{W}}(\|x\|_{\infty} \leq h) \sim \frac{4}{\pi} e^{-\pi^2/8h^2}.$$

By extending this result to any small-ball probability measure through the Cameron-Martin space characterization, we have

$$\forall x_0 \in \mathcal{F}, \quad \mathbb{P}^W \left(\|x - x_0\|_{\text{sup}} \leq h \right) \sim C_{x_0} e^{-\pi^2/8h^2}.$$

Since the Ornstein-Uhlenbeck process has a probability measure absolutely continuous with respect to P^W , we can directly state

$$\forall x_0 \in \mathcal{F}, \quad \mathbb{P} \left(\zeta^{\text{OU}} \in \mathcal{B}(x_0, h) \right) \sim C_{x_0} e^{-\pi^2/8h^2}.$$

Our crucial relation (2.4) is trivially satisfied for this Ornstein-Uhlenbeck process by choosing the function $\phi_x(\cdot)$ in the form

$$\phi_x^{\text{OU}}(h_n) \sim C_x e^{-\pi^2/8h^2}.$$

4. Weak convergence for kernel estimators

In this section, we are interested in studying the weak convergence of the conditional U -processes under absolute regular observations. Observe that

$$\begin{aligned} & \widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \\ &= \frac{1}{\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})} \left(\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + \widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right) \\ &= \frac{1}{\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})} \left(\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + \widehat{g}^B(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right), \end{aligned} \quad (4.1)$$

where

$$\begin{aligned} \widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\}, \\ \widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\} \mathfrak{B}_{\mathbf{i}, \varphi, n}, \\ \widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\} r^{(m)} \left(\frac{\mathbf{i}}{n}, X_{i, n}, \boldsymbol{\theta} \right). \end{aligned}$$

Under the same assumption in Theorem 3.3, we will show in the next theorem that

$$\text{Var}(\widehat{g}^B(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) = o \left(\frac{1}{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}(h_n)} \right)$$

and

$$1/\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = O_{\mathbb{P}}(1).$$

Then, we have

$$\widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \frac{\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})}{\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})} + B_n(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + o_{\mathbb{P}} \left(\sqrt{\frac{1}{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}(h_n)}} \right),$$

where

$$B_n(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \mathbb{E}[\widehat{g}^B(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})] / \mathbb{E}[\widehat{r}_1(\boldsymbol{\varphi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})],$$

is the “bias” term and $\frac{\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})}{\widehat{r}_1(\boldsymbol{\varphi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})}$ is the “variance” term. Let us define, for $\varphi_1, \varphi_2 \in \mathcal{F}_m$

$$\begin{aligned} \sigma(\varphi_1, \varphi_2) &= \lim_{n \rightarrow \infty} nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n) \mathbb{E}((\widehat{r}_n^{(m)}(\varphi_1, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi_1, \mathbf{u}, \mathbf{x})) \\ &\quad \times (\widehat{r}_n^{(m)}(\varphi_2, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi_2, \mathbf{u}, \mathbf{x}))). \end{aligned} \quad (4.2)$$

In the following, we would set $K_2(\cdot)$ as the asymmetrical triangle kernel, that is, $K_2(x) = (1-x)\mathbb{1}_{(x \in [0,1])}$ to simplify the proof. The main results of this section are given in the following theorems.

Theorem 4.1. *Let $\mathcal{F}_m \mathfrak{R}_{\Theta}^m$ be a measurable VC-subgraph class of functions, and assume that all the assumptions of Section 2.8 are satisfied for both cases $\mathfrak{B}_{i, \varphi, n} = 1$ and $\mathfrak{B}_{i, \varphi, n} = \varepsilon_{i, n}$. Then, as $n \rightarrow \infty$, the U -process*

$$\sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} (\widehat{r}_n^{(m)}(\boldsymbol{\varphi}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\boldsymbol{\varphi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - B_n(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}))$$

converges to a Gaussian process \mathfrak{G}_n over $\mathcal{F}_m \mathfrak{R}_{\Theta}^m$, whose sample paths are bounded and informally continuous with respect to the $\|\cdot\|_2$ -norm, with the covariance function given in (4.2).

The proof of Theorem 4.2 is postponed until Section 4.1.

To examine the weak convergence of our estimator using the standard procedure, involving Hoeffding decomposition, finite-dimensional convergence, and equicontinuity, we can turn to the following theorem. In the proof of this theorem, we express the conditional U -process in terms of a U -process based on a stationary sequence, illustrating its convergence to a Gaussian process. This convergence is established in the distribution sense within $l^\infty(\mathcal{F}_m \mathfrak{R}_{\Theta}^m)$, the space of bounded real functions on $\mathcal{F}_m \mathfrak{R}_{\Theta}^m$, as defined in [100]. For further details, refer to [6, 71], or [178].

Theorem 4.2. *Assume $\mathcal{F}_m \mathfrak{R}_{\Theta}^m$ is a measurable VC-subgraph class of functions, and all assumptions in Section 2.8 are satisfied. If in addition $n\phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)h^{m+2(2m \wedge \alpha)} \rightarrow 0$ as $n \rightarrow \infty$, then we have*

$$\sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} (\widehat{r}_n^{(m)}(\boldsymbol{\varphi}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\boldsymbol{\varphi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}))$$

converges in law to a Gaussian process $\{\mathfrak{G}(\psi) : \psi \in \mathcal{F}_m \mathfrak{R}_{\Theta}^m\}$ in $l^\infty(\mathcal{F}_m \mathfrak{R}_{\Theta}^m)$ that admits a version with uniformly bounded and uniformly continuous paths with respect to the $\|\cdot\|_2$ -norm, and its covariance function is given in (4.2).

The proof of Theorem 4.2 is deferred to Section 8.

Remark 4.3. *To eliminate the bias term, it is necessary to have $n\phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)h^{m+2(2m \wedge \alpha)} \rightarrow 0$ as $n \rightarrow \infty$. As a consequence, the last condition, along with $nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n) \rightarrow \infty$, holds as long as $h_n = n^{-\xi}$ and $\phi_{\mathbf{x}, \boldsymbol{\theta}}(h_n) = h_n^{mc}$, where $0 < c < 1 - \frac{1}{\xi m}$ and $\frac{1}{m(1+c)+2(2m \wedge \alpha)} < \xi < \frac{1}{m(1-c)}$.*

Remark 4.4. *The validity of the results remains intact even when replacing the entropy condition with the bracketing condition. In particular, the existence of constants $C_0 > 0$ and $v_0 > 0$ ensures the specified inequality, as described in the remark. In our framework, the choice of the kernel function is flexible, with minimal restrictions, as long as some mild conditions are satisfied. However, the selection of the bandwidth introduces challenges, and it is crucial for achieving a favorable rate of consistency.*

The bandwidth choice significantly impacts the bias and variance trade-off in the estimator. Therefore, adopting a bandwidth that adjusts based on specific criteria, available data, and location is more suitable. The discussion on this topic can be found in [44, 46, 132]. Establishing uniform-in-bandwidth central limit theorems in our context would be particularly interesting.

Remark 4.5. We can consider the scenario where $\Theta = \Theta_n$, where Θ_n satisfies the conditions $\text{card}(\Theta_n) = n^\alpha$ with $\alpha > 0$, and for every $\theta \in \Theta_n$, we have

$$\langle \theta - \theta_0, \theta - \theta_0 \rangle^{1/2} \leq C_7 b_n,$$

where b_n converges to zero, as discussed in [145].

The functional directions set, Θ_n , is constructed following a similar approach to [2, 145], as outlined below:

(i) Each direction $\theta \in \Theta_n$ is derived from a d_n -dimensional space formed by B-spline basis functions, denoted by $\{e_1(\cdot), \dots, e_{d_n}(\cdot)\}$. Thus, we express directions as:

$$\theta(\cdot) = \sum_{j=1}^{d_n} \alpha_j e_j(\cdot) \text{ where } (\alpha_1, \dots, \alpha_{d_n}) \in \mathcal{V}. \quad (4.3)$$

(ii) The set of coefficient vectors in (4.3), denoted by \mathcal{V} , is generated through the following steps:

Step 1. For each $(\beta_1, \dots, \beta_{d_n}) \in C^{d_n}$, where $C = \{c_1, \dots, c_J\} \subset \mathbb{R}^J$ represents a set of J 'seed-coefficients', construct the initial functional direction as

$$\theta_{\text{init}}(\cdot) = \sum_{j=1}^{d_n} \beta_j e_j(\cdot).$$

Step 2. For each θ_{init} from Step 1 satisfying $\theta_{\text{init}}(t_0) > 0$, where t_0 denotes a fixed value in the domain of $\theta_{\text{init}}(\cdot)$, compute $\langle \theta_{\text{init}}, \theta_{\text{init}} \rangle$ and form $(\alpha_1, \dots, \alpha_{d_n}) = (\beta_1, \dots, \beta_{d_n}) / \langle \theta_{\text{init}}, \theta_{\text{init}} \rangle^{1/2}$.

Step 3. Define \mathcal{V} as the collection of vectors $(\alpha_1, \dots, \alpha_{d_n})$ obtained in Step 2. Consequently, the final set of permissible functional directions is represented as

$$\Theta_n = \left\{ \theta(\cdot) = \sum_{j=1}^{d_n} \alpha_j e_j(\cdot); (\alpha_1, \dots, \alpha_{d_n}) \in \mathcal{V} \right\}.$$

5. Applications

While only the subsequent examples will be provided in this section, they serve as prototypes for a range of problems that can be explored in a comparable manner.

5.1. Discrimination

Now, we apply the results to the discrimination problem described in Section 3 of [168], also referring to [167]. We will employ similar notation and settings. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say $1, \dots, M$. The sets

$$A_j = \{(y_1, \dots, y_m) : \varphi(y_1, \dots, y_m) = j\}, \quad 1 \leq j \leq M,$$

then yield a partition of the feature space. Predicting the value of $\varphi(Y_1, \dots, Y_m)$ is tantamount to predicting the set in the partition to which (Y_1, \dots, Y_m) belongs. For any discrimination rule g , we have

$$\mathbb{P}(g(\mathbf{X}, \boldsymbol{\theta}) = \varphi(\mathbf{Y})) \leq \sum_{j=1}^M \int_{\{\mathbf{x}: g(\mathbf{x})=j\}} \max_{\mathbf{i}} \mathfrak{W}^j\left(\frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right) d\mathbb{P}(\mathbf{x}),$$

where

$$\mathfrak{W}^j\left(\frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right) = \mathbb{P}(\varphi(\mathbf{Y}_i) = j \mid \langle \mathbf{X}_i, \boldsymbol{\theta} \rangle = \langle \mathbf{x}, \boldsymbol{\theta} \rangle), \quad \mathbf{x} \in \mathcal{H}^m.$$

The above inequality becomes equality if

$$\mathfrak{G}_0(\mathbf{x}, \boldsymbol{\theta}) = \arg \max_{1 \leq j \leq M} \mathfrak{W}^j\left(\frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right).$$

$\mathfrak{G}_0(\cdot)$ is called the Bayes rule, and the pertaining probability of error

$$\mathbf{L}^* = 1 - \mathbb{P}(\mathfrak{G}_0(\mathbf{X}, \boldsymbol{\theta}) = \varphi(\mathbf{Y})) = 1 - \mathbb{E} \left\{ \max_{1 \leq j \leq M} \mathfrak{W}^j\left(\frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right) \right\}$$

is called the Bayes risk. Each of the above unknown functions \mathfrak{W}^j 's can be consistently estimated by one of the methods discussed in the preceding sections. Let, for $1 \leq j \leq M$,

$$\mathfrak{W}_n^j(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) = \frac{\sum_{\mathbf{i} \in I_n^m} \mathbb{1}\{\varphi(Y_{i_1}, \dots, Y_{i_m}) = j\} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h_n}\right) K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h_n}\right) K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right\}}.$$

Set

$$\mathfrak{G}_{0, n}(\mathbf{x}, \boldsymbol{\theta}) = \arg \max_{1 \leq j \leq M} \mathfrak{W}_n^j\left(\frac{\mathbf{i}}{n}, \mathbf{x}, \boldsymbol{\theta}\right).$$

Let us introduce

$$\mathbf{L}_n^* = \mathbb{P}(\mathfrak{G}_{0, n}(\mathbf{X}, \boldsymbol{\theta}) \neq \varphi(\mathbf{Y})).$$

Then, one can show that the discrimination rule $\mathfrak{G}_{0, n}(\cdot)$ is asymptotically Bayes' risk consistent

$$\mathbf{L}_n^* \rightarrow \mathbf{L}^*.$$

This follows from the apparent relation:

$$|\mathbf{L}^* - \mathbf{L}_n^*| \leq 2\mathbb{E} \left[\max_{1 \leq j \leq M} \left| \mathfrak{W}_n^j\left(\frac{\mathbf{i}}{n}, \mathbf{X}, \boldsymbol{\theta}\right) - \mathfrak{W}^j\left(\frac{\mathbf{i}}{n}, \mathbf{X}, \boldsymbol{\theta}\right) \right| \right].$$

5.2. Metric learning

Metric learning, a field that has garnered significant attention in recent years, revolves around adapting the metric to the underlying data. Extensive discussions on metric learning and its applications are available in [19, 54]. This concept has proven valuable in diverse domains, including computer vision, information retrieval, and bioinformatics. To demonstrate the practicality of metric learning, we delve into the metric learning problem for supervised classification outlined in [54]. Consider independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of a $\mathcal{H} \times \mathcal{Y}$ -valued random couple (X, Y) , where \mathcal{H} is a feature space, and $\mathcal{Y} = \{1, \dots, C\}$ (with $C \geq 2$), representing a finite set of labels. Let \mathcal{D} be a set of distance measures $D : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$. In this context, the objective of metric learning is to find a metric under which pairs of points with the same label are close to each other, while those with different labels are far apart. The natural way to define the risk of a metric D is given by:

$$R(D) = \mathbb{E}[\phi((1 - D(X, X')) \cdot (2 \mathbb{1}_{\{Y=Y'\}} - 1))], \quad (5.1)$$

where $\phi(u)$ is a convex loss function upper-bounding the indicator function $\mathbb{1}\{u \geq 0\}$, for instance, the hinge loss $\phi(u) = \max(0, 1 - u)$. To estimate $R(D)$, we consider the natural empirical estimator:

$$R_n(D) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \phi((D(X_i, X_j) - 1) \cdot (2 \mathbb{1}_{\{Y_i=Y_j\}} - 1)), \quad (5.2)$$

which is a one-sample U -statistic of degree two with a kernel given by:

$$\varphi_D((x, y), (x', y')) = \phi((D(x, x') - 1) \cdot (2 \mathbb{1}_{\{y=y'\}} - 1)).$$

The convergence to (5.1) of a minimizer of (5.2) has been investigated within the frameworks of algorithmic stability [107], algorithmic robustness [18], and the theory of U -processes under suitable regularization [50].

5.3. Kendall rank correlation coefficient

To test the independence of one-dimensional random variables Y_1 and Y_2 , [110] proposed a method based on the U -statistic K_n with the kernel function :

$$\varphi((s_1, t_1), (s_2, t_2)) = \mathbb{1}_{\{(s_2-s_1)(t_2-t_1) > 0\}} - \mathbb{1}_{\{(s_2-s_1)(t_2-t_1) \leq 0\}}. \quad (5.3)$$

Its rejection on the region is of the form $\{\sqrt{n}K_n > \gamma\}$. In this example, we consider a multivariate case. To test the conditional independence of $\xi, \eta : Y = (\xi, \eta)$ given X , we propose a method based on the conditional U -statistic :

$$\widehat{T}_n^{(2)}(\varphi, \mathbf{u}, \mathbf{x}) = \frac{\sum_{i \in I_n^2} \varphi(Y_{i_1}, Y_{i_2}) \prod_{k=1}^2 \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\}}{\sum_{i \in I_n^2} \prod_{k=1}^2 \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\}},$$

where $\mathbf{x} = (x_1, x_2) \in \mathbb{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (5.3). Suppose that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are d_1 and d_2 -dimensional random vectors respectively and $d_1 + d_2 = d$. Furthermore, suppose that Y_1, \dots, Y_n are observations of $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we are interested in testing :

$$H_0 : \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X. \text{ vs } H_a : H_0 \text{ is not true.} \quad (5.4)$$

Let $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2) \in \mathbb{R}^d$ such as $\|\mathbf{a}\| = 1$ and $\mathbf{a}_1 \in \mathbb{R}^{d_1}, \mathbf{a}_2 \in \mathbb{R}^{d_2}$, and $F(\cdot), G(\cdot)$ be the distribution functions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, respectively. Suppose $F^{a_1}(\cdot)$ and $G^{a_2}(\cdot)$ to be continuous for any unit vector $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$ where $F^{a_1}(t) = \mathbb{P}(\mathbf{a}_1^\top \boldsymbol{\xi} < t)$ and $G^{a_2}(t) = \mathbb{P}(\mathbf{a}_2^\top \boldsymbol{\eta} < t)$ and \mathbf{a}_i^\top means the transpose of the vector $\mathbf{a}_i, 1 \leq i \leq 2$. For $n = 2$, let $Y^{(1)} = (\boldsymbol{\xi}^{(1)}, \boldsymbol{\eta}^{(1)})$ and $Y^{(2)} = (\boldsymbol{\xi}^{(2)}, \boldsymbol{\eta}^{(2)})$ such as $\boldsymbol{\xi}^{(i)} \in \mathbb{R}^{d_1}$ and $\boldsymbol{\eta}^{(i)} \in \mathbb{R}^{d_2}$ for $i = 1, 2$, and :

$$\varphi^a(Y^{(1)}, Y^{(2)}) = \varphi\left(\left(\mathbf{a}_1^\top \boldsymbol{\xi}^{(1)}, \mathbf{a}_2^\top \boldsymbol{\eta}^{(1)}\right), \left(\mathbf{a}_1^\top \boldsymbol{\xi}^{(2)}, \mathbf{a}_2^\top \boldsymbol{\eta}^{(2)}\right)\right).$$

An application of Theorem 3.3 gives

$$\left| \widehat{r}_n^{(2)}(\varphi^a, \mathbf{u}, \mathbf{x}) - r^{(2)}(\varphi^a, \mathbf{u}, \mathbf{x}) \right| \longrightarrow 0, \quad a.s. \quad (5.5)$$

5.4. Conditional U-statistics for censored data

Consider a triple (Y, C, X) of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathcal{H}$. Here, Y is the variable of interest, C is a censoring variable, and X is a concomitant variable. Throughout, we work with a sample $\{(Y_i, C_i, X_i)_{1 \leq i \leq n}\}$ of independent and identically distributed replication of (Y, C, X) , $n \geq 1$. Actually, in the right censorship model, the pairs (Y_i, C_i) , $1 \leq i \leq n$, are not directly observed, and the corresponding information is given by $Z_i := \min\{Y_i, C_i\}$ and $\Delta_i := \mathbb{1}\{Y_i \leq C_i\}$, $1 \leq i \leq n$. Accordingly, the observed sample is

$$\mathcal{D}_n = \{(Z_i, \Delta_i, X_i), i = 1, \dots, n\}.$$

For example, survival data in clinical trials or failure time data in reliability studies are often subject to such censoring. To be more specific, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of diseases are usually censored by other competing risks to life, which result in death. In the sequel, we impose the following assumptions upon the distribution of (X, Y) . For $-\infty < t < \infty$, set

$$F_Y(t) = \mathbb{P}(Y \leq t), \quad G(t) = \mathbb{P}(C \leq t), \quad \text{and} \quad H(t) = \mathbb{P}(Z \leq t),$$

the right-continuous distribution functions of Y , C , and Z , respectively. For any right-continuous distribution function \mathcal{Q} defined on \mathbb{R} , denote by

$$T_{\mathcal{Q}} = \sup\{t \in \mathbb{R} : \mathcal{Q}(t) < 1\}$$

the upper point of the corresponding distribution. Now consider a pointwise measurable class \mathcal{F} of real measurable functions defined on \mathbb{R} , and assume that \mathcal{F} is of VC-type. We recall the regression function of $\psi(Y)$ evaluated at $\langle X, \theta \rangle = \langle t, \theta \rangle$, for $\psi \in \mathcal{F}$ and $t \in \mathcal{H}$, given by

$$r^{(1)}\left(\psi, \frac{i}{n}, t, \theta\right) = \mathbb{E}(\psi(Y_i) \mid \langle X_i, \theta \rangle = \langle t, \theta \rangle),$$

when Y is right-censored. To estimate $r^{(1)}(\psi, \cdot)$, we make use of the inverse probability of censoring weighted (I.P.C.W.) estimators that have recently gained popularity in the censored data literature (see [51, 111]). The key ideas of I.P.C.W. estimators are as follows. Introduce the real-valued function $\Phi_\psi(\cdot, \cdot)$ defined on \mathbb{R}^2 by

$$\Phi_\psi(y, c) = \frac{\mathbb{1}\{y \leq c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \quad (5.6)$$

Assuming the function $G(\cdot)$ to be known, first note that $\Phi_\psi(Y_i, C_i) = \Delta_i\psi(Z_i)/(1 - G(Z_i))$ is observed for every $1 \leq i \leq n$. Moreover, under the Assumption **(I)** below:

(I) C and (Y, \mathbf{X}) are independent.

We have

$$\begin{aligned} r^{(1)}(\Phi_\psi, \frac{i}{n}, t, \theta) &:= \mathbb{E}(\Phi_\psi(Y_i, C_i) \mid \langle X_i, \theta \rangle = \langle t, \theta \rangle) \\ &= \mathbb{E}\left\{ \frac{\mathbb{1}\{Y_i \leq C_i\}\psi(Z_i)}{1 - G(Z_i)} \mid \langle X_i, \theta \rangle = \langle t, \theta \rangle \right\} \\ &= \mathbb{E}\left\{ \frac{\psi(Y_i)}{1 - G(Y_i)} \mathbb{E}(\mathbb{1}\{Y_i \leq C_i\} \mid \mathbf{X}_i, Y_i) \mid \langle X_i, \theta \rangle = \langle t, \theta \rangle \right\} \\ &= r^{(1)}(\psi, \frac{i}{n}, t, \theta). \end{aligned} \quad (5.7)$$

Therefore, any estimate of $r^{(1)}(\Phi_\psi, \cdot)$, which can be built on fully observed data, turns out to be an estimate for $r^{(1)}(\psi, \cdot)$ too. Thanks to this property, most statistical procedures that provide estimates of the regression function in the uncensored case can be naturally extended to the censored case. For instance, kernel-type estimates are straightforward to construct. Set, for $\mathbf{x} \in \mathcal{I}$, $h \geq 0$, $1 \leq i \leq n$,

$$\bar{\omega}_{n,h,j}^{(1)}(u, x) := K_1\left(\frac{u - j/n}{h_n}\right) K_2\left(\frac{d_{\theta_k}(x, X_{j,n})}{h_n}\right) / \sum_{j=1}^n K_1\left(\frac{u - j/n}{h_n}\right) K_2\left(\frac{d_{\theta_k}(x, X_{j,n})}{h_n}\right). \quad (5.8)$$

In view of (5.6)–(5.8), whenever $G(\cdot)$ is known, a kernel estimator of $r^{(1)}(\psi, \cdot)$ is given by

$$\check{r}_n^{(1)}(\psi, u, x; h_n) = \sum_{i=1}^n \bar{\omega}_{n,h,i}^{(1)}(u, x) \frac{\Delta_i\psi(Z_i)}{1 - G(Z_i)}. \quad (5.9)$$

The function $G(\cdot)$ is generally unknown and has to be estimated. We will denote by $G_n^*(\cdot)$ the Kaplan-Meier estimator of the function $G(\cdot)$ [108]. Namely, adopting the conventions

$$\prod_{\emptyset} = 1$$

and $0^0 = 1$ and setting

$$N_n(u) = \sum_{i=1}^n \mathbb{1}\{Z_i \geq u\},$$

we have

$$G_n^*(u) = 1 - \prod_{i: Z_i \leq u} \left\{ \frac{N_n(Z_i) - 1}{N_n(Z_i)} \right\}^{(1-\Delta_i)}, \quad \text{for } u \in \mathbb{R}.$$

Given this notation, we will investigate the following estimator of $r^{(1)}(\psi, \cdot)$

$$\check{r}_n^{(1)*}(\psi, u, x; h_n) = \sum_{i=1}^n \bar{\omega}_{n,h,i}^{(1)}(u, x) \frac{\Delta_i \psi(Z_i)}{1 - G_n^*(Z_i)}, \quad (5.10)$$

refer to [111, 130]. Adopting the convention $0/0 = 0$, this quantity is well defined, since $G_n^*(Z_i) = 1$ if and only if $Z_i = Z_{(n)}$ and $\Delta_{(n)} = 0$, where $Z_{(k)}$ is the k th ordered statistic associated with the sample (Z_1, \dots, Z_n) for $k = 1, \dots, n$ and $\Delta_{(k)}$ is the Δ_j corresponding to $Z_k = Z_j$. A right-censored version of an unconditional U -statistic with a kernel of degree $m \geq 1$ is introduced by the principle of a mean preserving reweighting scheme in [60]. Reference [170] has proved almost sure convergence of multi-sample U -statistics under random censorship and provided application by considering the consistency of a new class of tests designed for testing equality in distribution. To overcome potential biases arising from right-censoring of the outcomes and the presence of confounding covariates, [53] proposed adjustments to the classical U -statistics. [186] proposed a different way in the estimation procedure of the U -statistic by using a substitution estimator of the conditional kernel given the observed data. We also refer to [45]. To our best knowledge, the problem of the estimation of the conditional U -statistics in the censored setting with variable bandwidth was opened up to the present, and it gives the primary motivation for the study of this section. A natural extension of the function defined in (5.6) is given by

$$\Phi_\psi(y_1, \dots, y_k, c_1, \dots, c_k) = \frac{\prod_{i=1}^k \{\mathbb{1}\{y_i \leq c_i\} \psi(y_1 \wedge c_1, \dots, y_k \wedge c_k)\}}{\prod_{i=1}^k \{1 - G(y_i \wedge c_i)\}}. \quad (5.11)$$

From this, we have an analogous relation to (5.7) given by

$$\begin{aligned} & \mathbb{E}(\Phi_\psi(Y_1, \dots, Y_k, C_1, \dots, C_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{t}) \\ &= \mathbb{E} \left(\frac{\prod_{i=1}^k \mathbb{1}\{Y_i \leq C_i\} \psi(Y_1 \wedge C_1, \dots, Y_k \wedge C_k)}{\prod_{i=1}^k \{1 - G(Y_i \wedge C_i)\}} \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{t} \right) \\ &= \mathbb{E} \left(\frac{\psi(Y_1, \dots, Y_k)}{\prod_{i=1}^k \{1 - G(Y_i)\}} \mathbb{E} \left(\prod_{i=1}^k \mathbb{1}\{Y_i \leq C_i\} \mid (Y_1, X_1), \dots, (Y_k, X_k) \right) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{t} \right) \\ &= \mathbb{E}(\psi(Y_1, \dots, Y_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{t}). \end{aligned} \quad (5.12)$$

An analogue estimator to (1.1) in the censored case is given by

$$\check{r}_n^{(k)}(\psi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \sum_{(i_1, \dots, i_k) \in I_n^k} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G(Z_{i_1})) \cdots (1 - G(Z_{i_k}))} \bar{\omega}_{n,\mathbf{i}}^{(k)}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \quad (5.13)$$

where, for $\mathbf{i} = (i_1, \dots, i_k) \in I(k, n)$,

$$\bar{\omega}_{n,\mathbf{i}}^{(k)}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \frac{\prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\}}. \quad (5.14)$$

The estimator that we will investigate is given by

$$\check{r}_n^{(k)*}(\boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \sum_{(i_1, \dots, i_k) \in I_n^k} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G_n^*(Z_{i_1})) \cdots (1 - G_n^*(Z_{i_k}))} \bar{\omega}_{n,\mathbf{i}}^{(k)}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}). \quad (5.15)$$

The main result of this section is given in the following corollary.

Corollary 5.1. *Let $\mathcal{F}_m \mathfrak{S}_{\Theta}^m$ be a measurable VC-subgraph class of functions complying with Assumption 6. Suppose Assumptions 1–4 are fulfilled. Then we have:*

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} \left| \check{r}_n^{(k)*}(\boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\boldsymbol{\varphi}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right| \\ &= O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{n h_n^m \phi^m(h_n)}} + h_n^{2m \wedge \alpha} \right). \end{aligned} \quad (5.16)$$

This last result is a direct consequence of Theorem 3.3 and the law of iterated logarithm for $G_n^*(\cdot)$ established in [84] ensures that

$$\sup_{t \leq \tau} |G_n^* - G(t)| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \quad \text{almost surely as } n \rightarrow \infty.$$

For more details refer to [34].

5.5. Conditional U -statistics for left truncated and right censored data

Keeping in mind the notation of the preceding section, we now introduce a truncation variable denoted as L and assume that (L, C) is independent of Y . Let us consider a situation where we have random vectors $(Z_i \epsilon_i, \Delta_i)$, with $\epsilon_i = \mathbb{1}(L_i \leq Z_i)$. In this section, we aim to define U -statistics conditional for data that are left truncated and right censored (LTRC), by following ideas from [172] in the unconditional setting. To achieve this, we propose an extension of the function (5.6) for LTRC data as follows:

$$\tilde{\Phi}_{\psi}(y_1, \dots, y_k, l_1, \dots, l_k, c_1, \dots, c_k) = \frac{\psi(y_1 \wedge c_1, \dots, y_k \wedge c_k) \prod_{i=1}^k \mathbb{1}\{y_i \leq c_i\} \mathbb{1}\{l_i \leq z_i\}}{\prod_{i=1}^k \mathbb{P}(l_i < z_i < c_i)}.$$

According to (5.12), we get that

$$\mathbb{E}(\Phi_{\psi}(Y_1, \dots, Y_k, L_1, \dots, L_k, C_1, \dots, C_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{t})$$

$$= \mathbb{E}(\psi(Y_1, \dots, Y_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{t}).$$

An analog estimator to (1.1) for LTRC data can be expressed as follows:

$$\check{r}_n^{(k)}(\psi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \sum_{(i_1, \dots, i_k) \in I_n^k} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \epsilon_{i_1} \cdots \epsilon_{i_k} \psi(Z_{i_1}, \dots, Z_{i_k})}{\mathbb{P}(L_{i_1} < Z_{i_1} < C_{i_1}) \cdots \mathbb{P}(L_{i_k} < Z_{i_k} < C_{i_k})} \bar{\omega}_{n,i}^{(k)}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \quad (5.17)$$

where $\bar{\omega}_{n,i}^{(k)}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})$ is defined as in (5.14). As $\mathbb{P}(L_i < Z_i < C_i)$ is not known, we need to estimate it. We introduce $N_i(t) = \mathbb{1}(L_i < Z_i \leq t, \Delta_i = 1)$ and $N_i^c(t) = \mathbb{1}(L_i < Z_i \leq t, \Delta_i = 0)$ as the counting process corresponding to the variable of interest and the censoring variable. Furthermore, let

$$N(t) = \sum_{i=1}^n N_i(t)$$

and

$$N^c(t) = \sum_{i=1}^n N_i^c(t).$$

We introduce the risk indicators as $R_i(t) = \mathbb{1}(Z_i \geq t \geq L_i)$ and

$$R(t) = \sum_{i=1}^n R_i(t).$$

It is important to note that the risk set $R(t)$ at t contains the subjects who entered the study before t and are still under study at t . Indeed, $N_i^c(t)$ is a local sub-martingale with the appropriate filtration \mathbb{F}_t . The martingale associated with the censoring counting process with filtration \mathbb{F}_t is given by

$$M_i^c(t) = N_i^c(t) - \int_0^t R_i(u) \lambda_c(u) du, \quad i = 1, 2, \dots, n.$$

Here, $\lambda_c(\cdot)$ represents the hazard function associated with the censoring variable C under left truncation. The cumulative hazard function for the censoring variable C is defined as

$$\Lambda_c(t) = \int_0^t \lambda_c(u) du.$$

Denote

$$M^c(t) = \sum_{i=1}^n M_i^c(t).$$

Now, we define the sub-distribution function of T_1 corresponding to $\Delta_1 = 1$ and $\epsilon_1 = 1$ as

$$S(x) = \mathbb{P}(T_1 \leq x, \Delta_1 \epsilon_1 = 1).$$

Let

$$w(t) = \int_0^\infty \frac{h_1(x)}{\mathbb{P}(L_1 \leq x \leq C_1)} \mathbb{1}(x > t) dS(x),$$

where $h_1(x) = \mathbb{E}(\psi((T_1, \Delta_1), \dots, (T_k, \Delta_k)) \mid (T_1, \Delta_1) = (x, \Delta_1))$. Also, denote $\tilde{z}(t) = \mathbb{P}(T_1 \geq t \geq L_1)$. Then, an estimate for the survival function of the censoring variable C under left truncation, denoted as $\widehat{K}_c(\cdot)$, see [174], can be formulated as follows:

$$\widehat{K}_c(\tau) = \prod_{t \leq \tau} \left(1 - \frac{dN^c(t)}{\tilde{Z}(t)} \right). \quad (5.18)$$

Similar to the Nelson-Aalen estimator (for instance, see [7]), the estimator for the cumulative hazard function of the censoring variable C under left truncation is represented as:

$$\widehat{\Lambda}_c(\tau) = \int_0^\tau \frac{dN^c(t)}{\tilde{Z}(t)}. \quad (5.19)$$

In both the definitions presented in (5.18) and (5.19), we make the assumption that $\tilde{Z}(t)$ is non-zero with probability one. The interrelation between $\widehat{K}_c(\tau)$ and $\widehat{\Lambda}_c(\tau)$ can be expressed as:

$$\widehat{K}_c(\tau) = \exp[-\widehat{\Lambda}_c(\tau)].$$

Let $a_K = \inf\{t : K(t) > 0\}$ and $b_K = \sup\{t : K(t) < 1\}$ denote the left and right endpoints of the support. For LTRC data, as in [187], $F(\cdot)$ is identifiable if $a_G \leq a_W$ and $b_G \leq b_W$. By Corollary 2.2. [187], for $b < b_W$ we readily infer that

$$\sup_{a_W < \tau < b} |\widehat{K}_c(\tau) - K_c(\tau)| = O(\sqrt{n^{-1} \log \log n}). \quad (5.20)$$

From the above, the estimator (5.17) can be rewritten directly as follows:

$$\check{r}_n^{(k)*}(\psi, \theta, \mathbf{u}, \mathbf{x}) = \sum_{(i_1, \dots, i_k) \in I_n^k} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \epsilon_{i_1} \cdots \epsilon_{i_k} \psi(Z_{i_1}, \dots, Z_{i_k}) \omega_{n,1}^{(k)}(\theta, \mathbf{u}, \mathbf{x})}{\widehat{K}_c(Z_{i_1}) \cdots \widehat{K}_c(Z_{i_k})}. \quad (5.21)$$

The last estimator is the conditional version of that studied in [172]. Following the same reasoning of the Corollary 5.1, one can infer that as $n \rightarrow \infty$,

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{S}_0^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} \left| \check{r}_n^{(k)*}(\psi, \theta, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right| \\ &= O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh^m \phi^m(h_n)}} + h^{2m \wedge \alpha} \right). \end{aligned} \quad (5.22)$$

6. The bandwidth selection criterion

Various methodologies have been devised to formulate asymptotically optimal bandwidth selection rules for nonparametric kernel estimators, particularly for the Nadaraya-Watson regression estimator. Key contributions have been made by [42, 92, 96, 150]. The proper selection of bandwidth is pivotal, whether in the standard finite-dimensional case or in the infinite-dimensional framework, to ensure robust practical performance. Currently, to the best of our knowledge, such investigations do not exist for addressing a general functional conditional U -statistic. However, an extension of the leave-one-out cross-validation procedure allows us to define, for any fixed $\mathbf{j} = (j_1, \dots, j_m) \in I_n^m$:

$$\tilde{r}_{n,\mathbf{j}}^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m(\mathbf{j})} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \varphi(Y_{i,n})}{\sum_{\mathbf{i} \in I_n^m(\mathbf{j})} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\}}, \quad (6.1)$$

where:

$$I_n^m(\mathbf{j}) := \{\mathbf{i} \in I_n^m \text{ and } \mathbf{i} \neq \mathbf{j}\} = I_n^m \setminus \{\mathbf{j}\}.$$

Equation (6.1) denotes the leave-out- $(\mathbf{X}_j, \mathbf{Y}_j)$ estimator of the functional regression and can also serve as a predictor of $\varphi(\mathbf{Y}_{j_1}, \dots, \mathbf{Y}_{j_m}) := \varphi(\mathbf{Y}_j)$. To minimize the quadratic loss function, we introduce the following criterion. Let $\mathcal{W}(\cdot)$ be a known non-negative weight function:

$$CV(\varphi, h_n) := \frac{(n-m)!}{n!} \sum_{\mathbf{j} \in I_n^m} \left(\varphi(\mathbf{Y}_j) - \tilde{r}_{n,\mathbf{j}}^{(m)}(\varphi, \mathbf{u}, \mathbf{X}_j, \boldsymbol{\theta}; h_n) \right)^2 \widetilde{\mathcal{W}}(\mathbf{X}_j). \quad (6.2)$$

Building upon the concepts advanced by [150], a judicious approach for selecting the bandwidth is to minimize the aforementioned criterion. Therefore, we choose $\widehat{h}_n \in [a_n, b_n]$, minimizing among $h \in [a_n, b_n]$:

$$CV(\varphi, h_n).$$

Following the approach proposed by [20], where bandwidths are locally determined through a data-driven method involving the minimization of a functional version of a cross-validated criterion, we can substitute (6.2) with:

$$CV(\varphi, h_n) := \frac{(n-m)!}{n!} \sum_{\mathbf{j} \in I_n^m} \left(\varphi(\mathbf{Y}_j) - \tilde{r}_{n,\mathbf{j}}^{(m)}(\varphi, \mathbf{u}, \mathbf{X}_j; h_n) \right)^2 \widehat{\mathcal{W}}(\mathbf{X}_j, \mathbf{x}), \quad (6.3)$$

where

$$\widehat{\mathcal{W}}(\mathbf{s}, \mathbf{t}) := \prod_{i=1}^m \widehat{W}(\mathbf{s}_i, \mathbf{t}_i).$$

In practice, one takes, for $\mathbf{i} \in I_n^m$, the uniform global weights $\widetilde{\mathcal{W}}(\mathbf{X}_i) = 1$, and the local weights

$$\widehat{W}(\mathbf{X}_i, \mathbf{t}) = \begin{cases} 1 & \text{if } d_{\theta}(\mathbf{X}_i, \mathbf{t}) \leq h_n, \\ 0 & \text{otherwise.} \end{cases}$$

For conciseness, we have exclusively discussed the widely used cross-validated selected bandwidth method. However, this approach can be generalized to other bandwidth selectors, including those based on Bayesian principles [156].

7. Concluding remarks

This manuscript introduces the theory of single-index U -processes tailored specifically for locally stationary variables within the functional data paradigm. The primary objective is to leverage

functional local stationary approximations to facilitate asymptotic analyses in the statistical inference of non-stationary time series. We underscore the significance of adopting absolutely regular conditions or β -mixing conditions, which are independent of the entropy dimension of the class, a departure from other mixing conditions. In contrast to strong mixing, β -mixing offers greater flexibility, allowing decoupling and accommodation of diverse examples. Reference [103] provided a comprehensive characterization of stationary Gaussian processes satisfying the β -mixing condition. Additionally, β -mixing aligns with the $L_2(\mathbb{P})$ -norm, playing a crucial role. Unlike strong mixing, which demands a polynomial rate of decay for strong mixing coefficients contingent on the entropy dimension of the function class, β -mixing involves the L_1 -norm and the metric entropy function $H(\cdot, T, d)$. This function is defined concerning the pseudo-metric $d(s, t) = \sqrt{\text{Var}(\mathbb{G}(s) - \mathbb{G}(t))}$ for a Gaussian process $\mathbb{G}(\cdot)$. The definition satisfies the integrability condition:

$$\int_0^1 \sqrt{H(u, T, d)} du < +\infty.$$

Consequently, we establish the rate of convergence, demonstrating that, under suitable conditions, the kernel estimator $\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n)$ constructed with bandwidth h converges to the regression operator $r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})$ with a rate:

$$O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh^m \phi^m(h_n)}} + h^{2m \wedge \alpha} \right).$$

The presented rate underscores the importance of the small-ball probability function, impacting the concentration of functional variables X_i . The second term is linked to the bias of the estimate, dependent on the smoothness of the operator $r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})$ and its parameter α , as indicated by the Lipschitz condition. The interconnected nature of the concentration of functional variables X , the small-ball probability, and the convergence rate are crucial for achieving a more efficient estimator with less dispersed variables and a higher small-ball probability. In the context of empirical process settings, the rate of convergence is established over a subset $[Ch, Ch - 1]^m \times \{\mathbf{x}\}^m$, and for forecasting purposes, it can be extended to a subset $[Ch - 1, 1]^m \times \{\mathbf{x}\}^m$ using one-sided kernels or boundary-corrected kernels. This extension necessitates ensuring that the kernels have compact support and are Lipschitz. The weak convergence is established through classical procedures, involving finite-dimensional convergence and the equicontinuity of conditional U -processes. Finite-dimensional convergence is achieved through Hoeffding decomposition, followed by approaching independence through a block decomposition strategy, ultimately leading to proving a central limit theorem for independent variables. The equicontinuity aspect requires meticulous attention due to the comprehensive framework considered. These results provide a solid theoretical foundation for our methodologies, extending non-parametric functional principles to a generic dependent structure—a relatively new and significant research area. It is crucial to note that, when dealing with highly dependent data, mixing, often adopted for simplicity, may not be suitable. The ergodic framework eliminates the need for frequently used strong mixing conditions, along with their variations for measuring dependence, and the more intricate probabilistic computations they entail.

(see [36, 37, 47]). An intriguing avenue for exploration is the k NN estimator:

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{H_{n,k}(x_k)} \right) \right\} \varphi(Y_{\mathbf{i}, n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{H_{n,k}(x_k)} \right) \right\}},$$

where

$$H_{n,k}(x_j) = \min \left\{ h \in \mathbb{R}^+ : \sum_{i=1}^n \mathbb{1}_{B(x_j, h)}(X_i) = k \right\},$$

with $B(t, h) = \{z \in \mathcal{H} : d(z, t) \leq h\}$ representing a ball in \mathcal{H} with center $t \in \mathcal{H}$ and radius h , and $\mathbb{1}_A$ being the indicator function of the set A , as detailed in [3]. These findings open avenues for various applications, such as data-driven automatic bandwidth selection and confidence band construction. We propose the intriguing notion that bootstrap methods, as outlined in [35, 38], provide valuable insights when applied in the functional context, especially the functional variant of the wild bootstrap, employed for smoothing parameter selection. It is crucial to acknowledge that the theoretical underpinnings of this bandwidth selection method using the functional wild bootstrap are still an area with unresolved challenges. Finally, change-point detection has emerged as a widely utilized technique for recognizing specific points within a data series when a stochastic system experiences abrupt external perturbations. The occurrence of these alterations can be attributed to several circumstances, hence their identification can greatly enhance their comprehension. The application of change-point analysis has been observed in numerous stochastic processes across a wide range of scientific domains. However, the investigation of change-point analysis for conditional U -statistics remains an unexplored and demanding research issue.

8. Mathematical developments

In this section, we focus on proving our results, using the notation introduced earlier. We start by presenting the following lemma before delving into the proof of the main results. The proof techniques follow and extend those of [165] to the single index setting. Additionally, we incorporate certain intricate steps from [13], as observed in [39, 40].

Proof of Proposition 3.1. As mentioned earlier, our statistic is a weighted U -statistic, expressed as a sum of U -statistics through the Hoeffding decomposition. We will delve into the details of this decomposition in Sub-section 3.1 to achieve our desired outcomes. In that specific section, we observed that:

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) - \mathbb{E}(\widehat{\psi}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi)) = \widehat{\psi}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) + \widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi),$$

where the linear term $\widehat{\psi}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi)$ and the residual term $\widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi)$ are precisely defined in (3.9) and (3.11), respectively. Our goal is to establish that the linear term governs the convergence rate of this statistic, while the remaining term converges to zero almost surely as $n \rightarrow \infty$. We will initiate the analysis by addressing the first term in the decomposition. Consider $B = [0, 1]$, with

$\alpha_n = \sqrt{\log n/nh^m\phi^m(h_n)}$ and $\tau_n = \rho_n n^{1/\zeta}$, where ζ is a positive constant specified in Assumption 4, part i), and $\rho_n = (\log n)^{\zeta_0}$ for some $\zeta_0 > 0$. Define

$$\tilde{\mathbb{H}}_1^{(\ell)}(z) := \tilde{\mathbb{H}}^{(\ell)}(z) \mathbb{1}_{\{|\mathfrak{B}_{i,n}| \leq \tau_n\}}, \quad (8.1)$$

$$\tilde{\mathbb{H}}_2^{(\ell)}(z) := \tilde{\mathbb{H}}^{(\ell)}(z) \mathbb{1}_{\{|\mathfrak{B}_{i,n}| > \tau_n\}}, \quad (8.2)$$

and

$$\begin{aligned} \widehat{\psi}_1^{(1)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \vartheta(\mathbf{i}) &= \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{\ell} \xi_{i_{\ell}} \cdots \xi_{i_{m-1}} \tilde{\mathbb{H}}_1^{(\ell)}(z), \\ \widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \vartheta(\mathbf{i}) &= \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{\ell} \xi_{i_{\ell}} \cdots \xi_{i_{m-1}} \tilde{\mathbb{H}}_2^{(\ell)}(z). \end{aligned}$$

It is evident that we have

$$\begin{aligned} &\widehat{\psi}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) - \mathbb{E}\widehat{\psi}_{1,i}(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}, \varphi) \\ &= \left[\widehat{\psi}_1^{(1)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \mathbb{E}\widehat{\psi}_1^{(1)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right] + \left[\widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \mathbb{E}\widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right]. \end{aligned} \quad (8.3)$$

First, we can see that

$$\begin{aligned} &\mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \vartheta(\mathbf{i})| > \alpha_n \right) \\ &= \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \vartheta(\mathbf{i})| > \alpha_n \right) \\ &\quad \cap \left\{ \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \bigcup_{i=1}^n |\mathfrak{B}_{i,n}| > \tau_n \right\} \cup \left\{ \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left\{ \bigcup_{i=1}^n |\mathfrak{B}_{i,n}| > \tau_n \right\}^c \right\} \\ &\leq \mathbb{P} \left\{ \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \vartheta(\mathbf{i})| > \alpha_n \right. \\ &\quad \left. \cap \left\{ \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \bigcup_{i=1}^n |\mathfrak{B}_{i,n}| > \tau_n \right\} \right\} \\ &\quad + \mathbb{P} \left\{ \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})| > \alpha_n \right. \\ &\quad \left. \cap \left\{ \left\{ \sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \bigcup_{i=1}^n |\mathfrak{B}_{i,n}| > \tau_n \right\}^c \right\} \right\} \\ &\leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\mathfrak{B}_{i,n}| > \tau_n \text{ for some } i = 1, \dots, n \right) + \mathbb{P}(\emptyset) \\ &\leq \tau_n^{-\zeta} \sum_{i=1}^n \mathbb{E} \left[\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\mathfrak{B}_{i,n}|^{\zeta} \right] \leq n \tau_n^{-\zeta} = \rho_n^{-\zeta} \rightarrow 0. \end{aligned}$$

We deduce that

$$\mathbb{E} \left[\left| \widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right| \right] \leq \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \mathbb{E} \left(\left| \tilde{H}_2^{(\ell)}(z) \right| \right), \quad (8.4)$$

where

$$\begin{aligned} \mathbb{E} \left(\left| \tilde{H}_2^{(\ell)}(z) \right| \right) &= \mathbb{E} \left[\frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) \mathfrak{B}_{i,n} \times \int \mathfrak{B}_{(1, \dots, \ell-1, \ell, \dots, m)} \right. \\ &\quad \left. \prod_{\substack{k=1 \\ k \neq i}}^m \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_k}(x_k, v_k)}{h_n} \right) \mathbb{P}(dv_1, \dots, dv_{\ell-1}, dv_\ell, \dots, dv_{m-1}) \mathbb{1}_{\{|\mathfrak{B}_{i,n}| > \tau_n\}} \right] \\ &\lesssim \tau_n^{-(\zeta-1)} \mathbb{E} \left[\frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) \left| \mathfrak{B}_{i,n} \right|^\zeta \right] \\ &\lesssim \frac{\tau_n^{-(\zeta-1)}}{\phi(h_n)} \mathbb{E} \left[K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) \right] \\ &\lesssim \frac{\tau_n^{-(\zeta-1)}}{\phi(h_n)} \times \left[\frac{1}{nh} + \phi(h_n) \right] \\ &\lesssim \frac{\tau_n^{-(\zeta-1)}}{nh\phi(h_n)} + \tau_n^{-(\zeta-1)}, \end{aligned} \quad (8.5)$$

where

$$\begin{aligned} &\mathbb{E} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) \right) \\ &= \mathbb{E} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) + K_2 \left(\frac{d_{\theta_i}(x_i, X_i^{i/n})}{h_n} \right) - K_2 \left(\frac{d_{\theta_i}(x_i, X_i^{i/n})}{h_n} \right) \right) \\ &\leq \mathbb{E} \left| K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) - K_2 \left(\frac{d_{\theta_i}(x_i, X_i^{i/n})}{h_n} \right) \right| + \mathbb{E} \left| K_2 \left(\frac{d_{\theta_i}(x_i, X_i^{i/n})}{h_n} \right) \right| \\ &\lesssim Ch^{-1} \mathbb{E} \left| d_{\theta_i}(x_i, X_{i,n}) - d_{\theta_i}(x_i, X_i^{i/n}) \right| + \mathbb{E} \left[\mathbb{1}_{(d(x, X_{i,n}^{i/n}) \leq h)} \right] (K_2 \text{ is Lipschitz}) \\ &\lesssim \frac{1}{nh} \mathbb{E} \left| U_i^{(i/n)} \right| + F_{i/n}(h; x_i) (\text{using Assumption 1 i}) \\ &\lesssim \frac{1}{nh} + \phi(h_n). \end{aligned}$$

Thus, we acquire

$$\begin{aligned} \mathbb{E} \left[\left| \widehat{\psi}_1^{(2)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right| \right] &\leq \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \mathbb{E} \left(\left| H_2^{(\ell)}(z) \right| \right) \\ &\lesssim \underbrace{\frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}}}_{\leq C \text{ uniformly in } \mathbf{u}} \times \left[\frac{\tau_n^{-(\zeta-1)}}{nh\phi(h_n)} + \tau_n^{-(\zeta-1)} \right] \end{aligned}$$

$$\lesssim \left[\frac{\tau_n^{-(\zeta-1)}}{nh\phi(h_n)} + \tau_n^{-(\zeta-1)} \right] \lesssim \tau_n^{-(\zeta-1)} = (\rho_n n^{1/\zeta})^{-(\zeta-1)} \lesssim \alpha_n.$$

Consequently, we deduce that

$$\sup_{\mathcal{F}_m \mathcal{R}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_1^{(2)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E} \widehat{\psi}_1^{(2)}(\varphi, \theta, \mathbf{u}, \mathbf{x})| = O_{\mathbb{P}}(\alpha_n). \quad (8.6)$$

Next, let's consider

$$\sup_{\mathcal{F}_m \mathcal{R}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E} \widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x})|.$$

To achieve the aimed result, we will cover the region $B^m = [0, 1]^m$ by

$$\bigcup_{k_1, \dots, k_m=1}^{N(\mathbf{u})} \prod_{j=1}^m \mathbf{B}(\mathbf{u}_{k_j}, r),$$

for some radius r . Hence, for each $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_m) \in [0, 1]^m$, there exists $\mathbf{l}(\mathbf{u}) = (l(\mathbf{u}_1), \dots, l(\mathbf{u}_m))$, where $\forall 1 \leq i \leq m, 1 \leq l(\mathbf{u}_i) \leq N(\mathbf{u})$ and such that

$$\mathbf{u} \in \prod_{i=1}^m \mathbf{B}(\mathbf{u}_{l(\mathbf{u}_i)}, r) \text{ and } |\mathbf{u}_i - \mathbf{u}_{l(\mathbf{u}_i)}| \leq r, \text{ for } 1 \leq i \leq m,$$

then, for each $\mathbf{u} \in [0, 1]^m$, the closest center will be $\mathbf{u}_1(\mathbf{u})$, and the ball with the closest centre will be defined by

$$\mathcal{B}(\mathbf{u}, \mathbf{l}(\mathbf{u}), r) := \prod_{j=1}^m \mathbf{B}(\mathbf{u}_{k_j}, r).$$

In the same way, $\Theta^m \times \mathcal{H}^m$ should be covered by

$$\bigcup_{\tilde{k}_1, \dots, \tilde{k}_m=1}^{N(\theta)} \bigcup_{k_1, \dots, k_m=1}^{N(\mathbf{x})} \prod_{j=1}^m \mathbf{B}(x_{k_j}, r),$$

for some radius r . Hence, for each $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{H}^m$, there exists $\mathbf{l}(\mathbf{x}) = (l(x_1), \dots, l(x_m))$, where $\forall 1 \leq i \leq m, 1 \leq l(x_i) \leq N(\mathbf{x})$ and such that

$$\mathbf{x} \in \prod_{i=1}^m B(u_{l(x_i)}, r) \text{ and } d_{\theta_i}(x_i, x_{l(x_i)}) \leq r, \text{ for } 1 \leq i \leq m,$$

then, for each $\mathbf{x} \in \mathcal{H}^m$, the closest center will be $\mathbf{x}_1(\mathbf{x})$, and the ball with the closest centre will be defined by

$$\mathbf{B}_\theta(\mathbf{x}, \mathbf{l}(\mathbf{x}), r) := \prod_{i=1}^m B_{\theta_i}(x_{l(x_i)}, r).$$

We define

$$K^*(\omega, \nu) = C \prod_{k=1}^m \mathbb{1}_{(|\omega_k| \leq 2C_1)} \prod_{k=1}^m K_2(\nu_k) \text{ for } (\omega, \nu) \in \mathbb{R}^2.$$

We can show that, for $(u, x) \in B_{j,n}$ and n large enough,

$$\left| \prod_{k=1}^m K_1 \left(\frac{u_k - \frac{ik}{n}}{h_n} \right) - \prod_{k=1}^m K_1 \left(\frac{u_{j,k} - \frac{ik}{n}}{h_n} \right) \right| K_2 \left(\frac{d_{\theta_i}(x_i, X_{i,n})}{h_n} \right) \leq \alpha_n K^* \left(\frac{u_n - \frac{i}{n}, d_{\theta_i}(x_i, X_{i,n})}{h_n} \right).$$

Let

$$\begin{aligned} \bar{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) &= \frac{1}{nh^m \phi(h_n)} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{\substack{i=1 \\ i \neq k}}^m \prod_{k=1}^m K^* \left(\frac{u_k - \frac{ik}{n}, d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \mathfrak{W}_{i,n} \\ &\times \int \mathfrak{W}_{(1, \dots, \ell-1, \ell, \dots, m)} \prod_{\substack{k=1 \\ k \neq i}}^m \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_k}(x_k, \nu_k)}{h_n} \right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_{\ell}, \dots, d\nu_{m-1}) \mathbb{1}_{\{|\mathfrak{W}_{i,n}| \leq \tau_n\}}. \end{aligned}$$

Note that $\mathbb{E} \left[\left| \bar{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right| \right] \leq M < \infty$ for some sufficiently large M . Then, we obtain

$$\begin{aligned} &\sup_{\mathcal{F}_m \mathfrak{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right] \right| \tag{8.7} \\ &\leq \sup_{\mathcal{F}_m \mathfrak{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) \right] \right| \\ &\quad + \sup_{\mathcal{F}_m \mathfrak{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \alpha_n \left(\left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) \right| + \mathbb{E} \left[\left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) \right| \right] \right) \\ &\leq \sup_{\mathcal{F}_m \mathfrak{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) \right] \right| \\ &\quad + \sup_{\mathcal{F}_m \mathfrak{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) - \mathbb{E} \left[\bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}, \theta) \right] \right| + 2MF(\mathbf{y})\alpha_n. \end{aligned}$$

Therefore

$$\begin{aligned} &\mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right] \right| > 4M\alpha_n \right) \\ &\leq N_{\mathcal{F}_m \mathfrak{S}_\Theta^m} N_{(\theta)}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1}, r)} \\ &\quad \mathbb{P} \left(\left| \widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right] \right| > 4M\alpha_n \right) \\ &\leq Q_{1,n} + Q_{2,n}, \tag{8.8} \end{aligned}$$

where, for $N_{\mathcal{F}_m \mathfrak{S}_\Theta^m} N_{(\theta)}^m N_{(x)}^m N_{(u)}^m$ denotes the covering number related respectively to the class of functions $\mathcal{F}_m \mathfrak{S}_\Theta^m$, the balls that cover $[0, 1]^m$ and the balls that cover $\mathcal{H}^m \times \Theta^m$.

$$\begin{aligned} Q_{1,n} &= N_{\mathcal{F}_m \mathfrak{S}_\Theta^m} N_{(\theta)}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1}, r)} \\ &\quad \mathbb{P} \left(\left| \widehat{\psi}_1(\mathbf{u}_j, \mathbf{x}) - \mathbb{E} \left[\widehat{\psi}_1(\mathbf{u}_j, \mathbf{x}) \right] \right| > M\alpha_n \right), \end{aligned}$$

$$Q_{2,n} = N_{\mathcal{F}_m \mathcal{S}_\Theta^m} N_{(\theta)}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1}, r)} \mathbb{P} \left(\left| \tilde{\psi}_1(\mathbf{u}_j, \mathbf{x}) - \mathbb{E} \left[\tilde{\psi}_1(\mathbf{u}_j, \mathbf{x}) \right] \right| > M \alpha_n \right).$$

Notice that $Q_{1,n}$ and $Q_{2,n}$ might be treated in the same way, so, we restrict our attention to $Q_{1,n}$. Write:

$$\begin{aligned} & \mathbb{P} \left(\left| \widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E} \widehat{\psi}_1^{(1)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right|_{\mathcal{F}_m \mathcal{S}_\Theta^m} > M \alpha_n \right) \\ &= \mathbb{P} \left[\left| h^m \phi^m(h_n) \sum_{i=1}^n \sum_{\substack{m-1 \\ n-1}}^{m-1} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{i_\ell} \cdots \xi_{i_{m-1}} H_1^{(\ell)}(z) \right. \right. \\ & \quad \left. \left. - \mathbb{E} \left[h^m \phi^m(h_n) \sum_{i=1}^n \sum_{\substack{m-1 \\ n-1}}^{m-1} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{i_\ell} \cdots \xi_{i_{m-1}} H_1^{(\ell)}(z) \right] \right|_{\mathcal{F}_m \mathcal{S}_\Theta^m} \right. \\ & \quad \left. > M n \frac{(n-1)!}{(n-m)!} \alpha_n h^m \phi^m(h_n) \right] \\ &= \mathbb{P} \left(\left| \sum_{i=1}^n \Phi_{i,n}(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right|_{\mathcal{F}_m \mathcal{S}_\Theta^m} > M n \frac{(n-1)!}{(n-m)!} \alpha_n h^m \phi^m(h_n) \right). \end{aligned}$$

Note that the array $\{\Phi_{i,n}(u, x)\}$ is α -mixing for each fixed (u, x) with mixing coefficients $\beta_{\Phi,n}$ such that $\beta_{\Phi,n}(k) \leq \beta(k)$. We apply Lemma A.4 with

$$\varepsilon := M n \frac{(n-1)!}{(n-m)!} h^m \phi^m(h_n) \alpha_n,$$

and $b_n = C \tau_n$ for sufficiently large $C > 0$ and $S_n = \alpha_n^{-1} \tau_n^{-1}$. As same as [133, Theorem 2], we can see that $\sigma_{S_n, n}^2 \leq C' S_n h^m \phi^m(h_n)$, and we obtain:

$$\begin{aligned} \mathbb{P} \left(\left| \sum_{i=1}^n Z_{i,n}(u, x) \right| \geq \varepsilon \right) &\leq 4 \exp \left(- \frac{\varepsilon^2}{64 \sigma_{S_n, n}^2 \frac{n}{S_n} + \frac{8}{3} \varepsilon b_n S_n} \right) + 4 \frac{n}{S_n} \beta(S_n) \\ &\leq 4 \exp \left(- \frac{M^2 \alpha_n^2 n^2 \left(\frac{(n-1)!}{(n-m)!} \right)^2 h^{2m} \phi^{2m}(h_n)}{64 C' S_n h \phi(h_n) \frac{n}{S_n} + \frac{8}{3} M n \frac{(n-1)!}{(n-m)!} h^m \phi^m(h_n) \alpha_n b_n S_n} \right) + 4 \frac{n}{S_n} \beta(S_n) \\ &\leq 4 \exp \left(- \frac{M \left(\sqrt{\log n / n h^m \phi^m(h_n)} \right)^2 n \left(\frac{(n-1)!}{(n-m)!} \right)}{64 C' h^m \phi^m(h_n) \frac{(n-m)!}{M(n-1)!} + \frac{8}{3} C h^m \phi^m(h_n)} \right) + 4 \frac{n}{S_n} \beta(S_n) \\ &\lesssim \exp \left(- \frac{M \frac{(n-1)!}{(n-m)!} \log n}{64 \frac{(n-m)!}{(n-1)!} \frac{C'}{M} + \frac{8}{3} C} \right) + n S_n^{-\gamma-1}. \end{aligned}$$

To get the last inequality, we must choose $M > C'$. Since $N \leq C h^{-m} \phi(h_n) \alpha_n^{-m}$, it follows that

$$\widehat{Q}_n \leq O(R_{1n}) + O(R_{2n}),$$

with

$$R_{1T} = h_n^{-m} \alpha_n^{-m} n^{-\frac{M \frac{(n-1)!}{(n-m)!}}{64 \frac{(n-m)!}{(n-1)!} + 3C}},$$

$$R_{2T} = h_n^{-m} \alpha_n^{-m} n S_n^{-\gamma-1}.$$

For M sufficiently large, we can see that $R_{1n} \leq n^{-\varsigma}$ for some small $\varsigma > 0$. As $\frac{\phi_n \log T}{T^\theta n^{d+1}} = o(1)$ by assumption, we further get that

$$\begin{aligned} R_{2n} &= h_n^{-m} \alpha_n^{-m} n S_n^{-\gamma-1} \\ &= h_n^{-m} n \left(\sqrt{\frac{\log n}{nh^m \phi^m(h_n)}} \right)^{-m} (\alpha_n^{-1} \tau_n^{-1})^{-\gamma-1} \\ &= h_n^{-m} \left(\sqrt{\frac{\log n}{nh^m \phi^m(h_n)}} \right)^{-m+\gamma+1} ((\log n)^{\zeta_0} n^{1/\zeta})^{\gamma+1} \\ &= \frac{(\log n)^{-\frac{m+\gamma+1}{2} + \zeta_0(\gamma+1)}}{n^{-\frac{m+\gamma+1}{2} - 1 - \frac{\gamma+1}{\zeta}} h^{\frac{m+\gamma+1}{2}} \phi(h_n)^{-\frac{m+\gamma+1}{2}}}. \end{aligned}$$

Under Assumption 4 ii), we establish that R_{2n} tends to zero, confirming the obtained result. Now, let's move on to the nonlinear segment of the Hoeffding decomposition. Here, the goal is to illustrate that

$$\mathbb{P} \left[\sup_{\mathcal{F}_m, \mathcal{S}_\theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{\psi}_{2,i}(\varphi, \theta, \mathbf{u}, \mathbf{x})| > \lambda \right] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

The conclusive phase in establishing Proposition 3.1 involves leveraging Lemma 8.2 to demonstrate the convergence of the nonlinear term to zero. \square

Proof of Theorem 3.3. Equation (4.1) in Section 4 shows that

$$\widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \theta; h_n) - r^{(m)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) = \frac{1}{\widetilde{r}_1(\varphi, \theta, \mathbf{u}, \mathbf{x})} \left(\widehat{g}_1(\theta, \mathbf{u}, \mathbf{x}) + \widehat{g}_2(\theta, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \widetilde{r}_1(\varphi, \theta, \mathbf{u}, \mathbf{x}) \right),$$

where

$$\begin{aligned} \widetilde{r}_1(\varphi, \theta, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\}, \\ \widehat{g}_1(\theta, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \mathfrak{B}_{\mathbf{i}, \varphi, n}, \\ \widehat{g}_2(\theta, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} r^{(m)} \left(\frac{\mathbf{i}}{n}, X_{i,n} \right). \end{aligned}$$

The proof of this theorem is intricate and divided into the following four steps; wherein each step, our objective is to demonstrate that:

Step 1.

$$\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})| = O_{\mathbb{P}} \left(\sqrt{\log n / nh^m \phi(h_n)} \right).$$

Step 2.

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n) \\ & - \mathbb{E}(\widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n))| = O_{\mathbb{P}} \left(\sqrt{\log n / nh^m \phi(h_n)} \right). \end{aligned}$$

Step 3.

$$\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\mathbb{E}(\widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n))| = O(h^2) + O(h^\alpha).$$

Step 4.

$$\frac{1}{\inf_{\mathcal{F}_m \mathcal{S}_\Theta^m} \inf_{\mathbf{x} \in \mathcal{H}^m} \inf_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})|} = O_{\mathbb{P}}(1).$$

Step 1 is an immediate consequence of Proposition 3.1. The validity of the second step is ensured by substituting $\varphi(Y_{i_1}, \dots, Y_{i_m})$ with

$$\widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n),$$

and applying Proposition 3.1. We will now proceed with the demonstration of Step 4. Consider

$$\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \widehat{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + \bar{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \quad (8.9)$$

where

$$\begin{aligned} \widehat{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right\} \\ \bar{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \\ & \quad \prod_{k=1}^m \left[K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right]. \end{aligned}$$

For $W \equiv 1$, the preceding proposition established that

$$\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \mathbb{E}(\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}))| = o_{\mathbb{P}}(1).$$

Therefore, it is evident that

$$\begin{aligned} \widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + \mathbb{E}(\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) - \mathbb{E}(\widetilde{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) \\ &= o_{\mathbb{P}}(1) + \mathbb{E}[\widehat{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})] + \mathbb{E}[\bar{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})]. \end{aligned} \quad (8.10)$$

Furthermore, we have

$$\begin{aligned}\mathbb{E}(\bar{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) &= \mathbb{E}\left(\frac{(n-m)!}{n!h^m\phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \right. \\ &\quad \left. \prod_{k=1}^m \left[K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n}\right) - K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n}\right) \right] \right) \\ &\approx \frac{(n-m)!}{n!h^m\phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \left(\frac{m\phi^{m-1}(h_n)}{nh_n}\right) = o(1).\end{aligned}\quad (8.11)$$

The final outcome arises from the Lipschitz continuity of $K_2(\cdot)$ (as stipulated in Assumption 2, i)), along with the utilization of Assumption 1 i) and Lemma A.2). This holds uniformly in \mathbf{u} . Furthermore,

$$\begin{aligned}\mathbb{E}[\widehat{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})] &= \frac{(n-m)!}{n!h^m\phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \mathbb{E}\left[\prod_{k=1}^m K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n}\right)\right] \\ &= \frac{(n-m)!}{n!h^m\phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \int_0^h \prod_{k=1}^m K_2\left(\frac{y_k}{h_n}\right) dF_{i_k/n}(y_k, x_k) \\ &\gtrsim \frac{(n-m)!}{n!h^m\phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \phi^m(h_n) f_1(\mathbf{x}) \sim f_1(\mathbf{x}) > 0,\end{aligned}$$

uniformly in \mathbf{u} . Then, we obtain

$$\begin{aligned}&\frac{1}{\inf_{\mathcal{F}_m} \inf_{\boldsymbol{\theta}} \inf_{\mathbf{x} \in \mathcal{H}^m} \inf_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |\bar{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})|} \\ &= \frac{1}{\inf_{\mathcal{F}_m} \inf_{\boldsymbol{\theta}} \inf_{\mathbf{x} \in \mathcal{H}^m} \inf_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} o(1) + o_{\mathbb{P}}(1) + \mathbb{E}[\widehat{r}_1(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x})]} = O_{\mathbb{P}}(1).\end{aligned}\quad (8.12)$$

Take $K_0 : [0, 1] \rightarrow \mathbb{R}$ to be a Lipschitz continuous function with its support in $[0, q]$ for some $q > 1$, and ensure that $K_0(x) = 1$ for all $x \in [0, 1]$. Importantly,

$$\mathbb{E}[\widehat{g}_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \bar{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n)] = \sum_{i=1}^4 Q_i(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \quad (8.13)$$

where Q_i can be defined as follows

$$Q_i(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \frac{(n-m)!}{n!h^m\phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \right\} q_i(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \quad (8.14)$$

such that

$$q_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) = \mathbb{E}\left[\prod_{k=1}^m K_0\left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n}\right) \left\{ \prod_{k=1}^m K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n}\right) \right\}\right]$$

$$\begin{aligned}
& - \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \times \left\{ r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, X_{i, n}\right) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right\} \Bigg], \\
q_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\prod_{k=1}^m \left\{ K_0 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right\} \right. \\
& \quad \left. \left\{ r^{(m)}\left(\varphi, \frac{i_k}{n}, X_{i_k, n}\right) - r^{(m)}\left(\varphi, \frac{i_k}{n}, X_{i_k, n}^{(i_k/n)}\right) \right\} \right], \\
q_3(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\left\{ \prod_{k=1}^m K_0 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^m K_0 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right\} \right. \\
& \quad \left. \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \times \left\{ r^{(m)}\left(\varphi, \frac{i_k}{n}, X_{i_k, n}^{(i_k/n)}\right) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right\} \right], \\
q_4(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \left\{ r^{(m)}\left(\varphi, \frac{i_k}{n}, X_{i_k, n}^{(i_k/n)}\right) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right\} \right].
\end{aligned}$$

Observe that

$$\begin{aligned}
Q_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &\lesssim \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \mathbb{E} \left[\prod_{k=1}^m K_0 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right. \right. \\
& \quad \left. \left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| \right] \\
& \quad \left. \times \left| r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, X_{i, n}\right) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right| \right\}.
\end{aligned}$$

Utilizing Assumption 3 (i), the properties of $r(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})$ allow us to establish that

$$\begin{aligned}
& \prod_{k=1}^m K_0 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \left| r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, X_{i, n}\right) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right| \\
& \lesssim \prod_{k=1}^m K_0 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \left(d_{\mathcal{H}^m}(X_{i, n}, \mathbf{x}) + \|\mathbf{u} - \frac{\mathbf{i}}{n}\| \right)^\alpha \lesssim h^{m \wedge \alpha}.
\end{aligned}$$

With Assumption 2, part ii) in mind, we will employ Lemma 8.1 and Eq (8.58) to verify that:

$$\begin{aligned}
Q_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &\lesssim \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \right\} \times h^{m \wedge \alpha} \times \frac{m \phi^{m-1}(h_n)}{n h_n}. \\
&\lesssim \frac{1}{n \phi(h_n) h^{m-(m \wedge \alpha)}} \quad \text{uniformly in } \mathbf{u}. \tag{8.15}
\end{aligned}$$

Likewise, it can be observed that

$$\sup_{\mathcal{F}_m \mathcal{S}_\theta^m} \sup_{\boldsymbol{\theta} \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} Q_2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \lesssim \frac{1}{n \phi(h_n) h^{m-(m \wedge \alpha)}}, \tag{8.16}$$

and

$$\sup_{\mathcal{F}_m \mathcal{R}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} Q_3(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \lesssim \frac{1}{n \phi(h_n) h^{m-(1 \wedge \alpha)}}. \quad (8.17)$$

Concerning the final term, we derive

$$\begin{aligned} Q_4(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \right\} \\ &\quad \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \left\{ r^{(m)}(\varphi, \frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}^{(\mathbf{i}/n)}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right\} \right]. \end{aligned}$$

By utilizing Lemma A.1 along with the inequality (2.12) and considering Assumption 1, it becomes evident that

$$\begin{aligned} &\sup_{\mathcal{F}_m \mathcal{R}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |Q_4(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})| \\ &\leq \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left| \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \right\} \right| \\ &\quad \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| \left| \left\{ r^{(m)}(\varphi, \frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}^{(\mathbf{i}/n)}) - r^{(m)}(\varphi, \boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right\} \right| \right] \\ &\lesssim \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left| \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \right\} \right| \\ &\quad \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| \left(d_{\mathcal{H}^m}(X_{\mathbf{i}, n}^{(\mathbf{i}/n)}, \mathbf{x}) + \|\mathbf{u} - \frac{\mathbf{i}}{n}\| \right)^\alpha \right] \\ &\lesssim \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left| \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \right| \\ &\quad - \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m K_1 \left(\frac{(u_k - v_k)}{h_n} \right) dv_k \left| \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| h_n^\alpha \right] \right. \\ &\quad \left. + \frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m K_1 \left(\frac{(u_k - v_k)}{h_n} \right) dv_k \right. \\ &\quad \left. \times \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| h_n^\alpha \right] \right. \\ &\lesssim O \left(\frac{1}{n^m h_n^{2m}} \right) h_n^\alpha + h_n^\alpha. \quad (8.18) \end{aligned}$$

Keep in mind

$$(n)^{-m} h_n^{\alpha-2m} \lesssim h_n^{2m} \phi^m(h_n) \ll h_n^{2m},$$

we deduce that

$$\sup_{\mathcal{F}_m \mathcal{R}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |Q_4(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})| \ll h_n^{2m} + h_n^\alpha.$$

Certainly, within the framework of our assumptions, the approximation error can be considered as

$$O\left(\frac{1}{n^m \phi(h_n) h_n^{m-(1 \wedge \alpha)}}\right) \ll h_n^{2m \wedge \alpha}. \quad (8.19)$$

With this inequality, the proof is concluded. \square

Proof of Theorem 4.1. The goal is to establish weak convergence, encompassing finite-dimensional convergence and asymptotic equicontinuity, for the stochastic U -process:

$$\sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} D_n(\varphi) = \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \left(\tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta) - B_n(\theta, \mathbf{u}, \mathbf{x}) \right),$$

across all the specified function classes within the framework. According to [64, Section 4.2], finite-dimensional convergence asserts that every finite set of functions f_1, \dots, f_q in L_2

$$\left(\sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} D_n(f_1), \dots, \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} D_n(f_q) \right) \quad (8.20)$$

converges to the corresponding finite-dimensional distributions of the process G_p . By leveraging Cramér-Wold and considering the countability of various classes, we can simplify the scenario from the weak convergence of the U -process to the weak convergence of U -statistics with the kernel f_r for all $r \in \{1, \dots, q\}$. As the U -process is a linear operator, our focus narrows down to demonstrating the convergence of $\sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} D_n(f_r)$ towards a Gaussian distribution. Therefore, for a fixed kernel, we have:

$$\begin{aligned} & \tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta) \\ &= \frac{1}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta)} \left(\widehat{g}_1(\theta, \mathbf{u}, \mathbf{x}) + \widehat{g}_2(\theta, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta) \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta; h_n) \right) \\ &= \frac{1}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta)} \left(\widehat{g}_1(\theta, \mathbf{u}, \mathbf{x}) + \widehat{G}(\theta, \mathbf{u}, \mathbf{x}) \right), \end{aligned} \quad (8.21)$$

where

$$\widehat{G}(\theta, \mathbf{u}, \mathbf{x}) = \widehat{g}_2(\theta, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta) \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta; h_n).$$

We begin treating each term. For this sake, we will calculate the variance of $\widehat{G}(\theta, \mathbf{u}, \mathbf{x})$. Take

$$\Delta_{i,n}(\theta, \mathbf{u}, \mathbf{x}) = \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \left[r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{i,n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta) \right].$$

Observe that

$$\begin{aligned} \text{Var}(\widehat{G}(\theta, \mathbf{u}, \mathbf{x})) &= \text{Var}\left(\widehat{g}_2(\theta, \mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta) \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \theta; h_n)\right) \\ &= \text{Var}\left(\frac{(n-m)!}{n! h^m \phi^m(h_n)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h_n}\right) \right\} \Delta_{i,n}(\theta, \mathbf{u}, \mathbf{x})\right) \\ &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \text{Var}(\Delta_{i,n}(\theta, \mathbf{u}, \mathbf{x})) \end{aligned}$$

$$\begin{aligned}
& + \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\substack{i_1 \neq \dots \neq i_m, \\ i'_1 \neq \dots \neq i'_m, \\ \exists j/i_j \neq i'_j}} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \prod_{k=1}^m K_1^2 \left(\frac{u_k - i'_k/n}{h_n} \right) \\
& \text{Cov}(\Delta_{i,n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \Delta_{i',n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) \\
& := V_1 + V_2. \tag{8.22}
\end{aligned}$$

Looking on V_1 , we have

$$\begin{aligned}
|V_1| &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \text{Var}(\Delta_{i,n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) \\
&= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \left[\mathbb{E}(\Delta_{i,n}^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) - (\mathbb{E}(\Delta_{i,n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})))^2 \right] \\
&\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \mathbb{E}(\Delta_{i,n}^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) \\
&\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right. \\
&\quad \left. \left[r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{i,n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) \right]^2 \right) \\
&\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right. \\
&\quad \left. - \prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) + \prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right. \\
&\quad \left. \times \left[r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{i,n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) \right]^2 \right) \\
&\leq \frac{((n-m)!)^2 h^{2\alpha}}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \mathbb{E} \left[\prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right. \\
&\quad \left. - \prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right] + \mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right). \tag{8.23}
\end{aligned}$$

The last part of (8.23) follows from the smoothness assumption on $r^{(m)}$ in 3 i)

$$\left| r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{i,n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) \right|^2 \lesssim h^{2\alpha}. \tag{8.24}$$

Combining the latter inequalities with Eqs (8.60) and (8.58) from Lemma 8.1 where:

$$\mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) - \prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right)$$

$$\lesssim \mathbb{E} \left(\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right),$$

to get

$$\begin{aligned} |V_1| &\lesssim \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h_n} \right) \\ &\lesssim \frac{h^{2\alpha}(h_n)((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \left[\frac{m\phi^{m-1}(h_n)}{nh_n} + \phi^m(h_n) \right] \ll \frac{1}{nh^m \phi(h_n)}. \end{aligned} \quad (8.25)$$

A deep sight into the work of [10], specially Lemma 2, makes us see V_2 as follows:

$$\begin{aligned} |V_2| &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \\ &\quad \times \text{Cov} \left(\Delta_{i_{\sigma(1)}, \dots, i_{\sigma(2m)}, n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}), \Delta_{i_{\sigma(m+1)}, \dots, i_{\sigma(2m)}, n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right) \\ &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \\ &\quad \left| \mathbb{E} \left[\Delta_{i_{\sigma(1)}, \dots, i_{\sigma(2m)}, n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \Delta_{i_{\sigma(1)}, \dots, i_{\sigma(2m)}, n}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right] \right| \\ &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \times cM^2 \left(1 + \sum_{k=1}^{n-1} k^{m-1} \beta_k^{(p-2)/p} \right) \\ &\lesssim \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \times M^2 \ll \frac{1}{nh^m \phi(h_n)}, \end{aligned} \quad (8.26)$$

where

$$M := \sup_{1 \leq i_1 < \dots < i_m < \infty} \mathbb{E} \left[\left| \Delta_{i, n} \right|^\zeta \right]^{1/\zeta},$$

and $j_1 = i_2 - i_1$, $j_l = \min(i_{2l-1} - i_{2l-2}, i_{2l} - i_{2l-1})$ for $2 \leq l \leq m-1$, and $j_m = i_{2m} - i_{2m-1}$. If we designate $j_1 = \max(j_1, \dots, j_m)$, we can align the original sequence $\{X_1, \dots, X_n\}$ with another sequence characterized by independent blocks $\{i_1, i_2, \dots, i_{2m}\}$ while preserving the identical block distribution. It is straightforward to demonstrate now that

$$\text{Var}(\widehat{G}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) \leq |V_1| + |V_2| = o\left(\frac{1}{nh_n^m \phi(h_n)}\right). \quad (8.27)$$

This indicates the quadratic-mean convergence of $\widehat{G}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})$ with the specified rate as follows:

$$\widehat{G}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) - \mathbb{E}(\widehat{G}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})) = o\left(\frac{1}{nh_n^m \phi(h_n)}\right) \text{ in probability.} \quad (8.28)$$

Let's remember that

$$\begin{aligned}\tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) &= \frac{1}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta})} \left(\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + \widehat{G}(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \right), \\ B_n(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) &= \mathbb{E}[\widehat{g}^B(u, x)] / \mathbb{E}[\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta})], \\ \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) &= \mathbb{E}[\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta})] + o_{\mathbb{P}}(1), \\ \lim_{n \rightarrow \infty} \mathbb{E}[\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta})] &> 0,\end{aligned}$$

then

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}) = \frac{\widehat{g}_1(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x})}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta})} + B_n(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}\left(\frac{1}{nh_n^m \phi(h_n)}\right).$$

In the next step, we will consider the first part of the last equation.

$$\begin{aligned}& \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} \text{Var}(\widehat{g}_1(x)) \\ &= \frac{\sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \text{Var} \left[\prod_{k=1}^m K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \boldsymbol{\varepsilon}_{\mathbf{i}, n} \right] \\ &= \frac{\sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \boldsymbol{\varepsilon}_{\mathbf{i}, n}^2 \right] \\ &= \frac{\sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \sigma^2\left(\boldsymbol{\theta}, \frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}\right) \boldsymbol{\varepsilon}_{\mathbf{i}}^2 \right] \\ & \quad (\{\boldsymbol{\varepsilon}_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}} \text{ is a sequence i.i.d r.v.'s, independent of } \{X_{i, n}\}_{i=1}^n) \\ &= \frac{\mathbb{E}(\boldsymbol{\varepsilon}_{\mathbf{i}}^2) \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right] \left[\sigma^2\left(\boldsymbol{\theta}, \frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}\right) \right] \\ &= \frac{\mathbb{E}(\boldsymbol{\varepsilon}_{\mathbf{i}}^2) \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \\ & \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{i_k/n})}{h_n}\right) \right] \left[\sigma^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + o(1) \right] \text{ (According to Assumption 3 [ii]- iii)- iv)} \\ & \quad + \frac{\mathbb{E}(\boldsymbol{\varepsilon}_{\mathbf{i}}^2) o(\phi^m(h_n)) \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \left[\sigma^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + o(1) \right] \\ &= \frac{\mathbb{E}(\boldsymbol{\varepsilon}_{\mathbf{i}}^2) \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \\ & \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{i_k/n})}{h_n}\right) \right] \left[\sigma^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + o(1) \right] + o(1)\end{aligned}$$

$$\begin{aligned}
& \sim \frac{\mathbb{E}(\varepsilon_{\mathbf{i}}^2)(\sigma^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) + o(1)) \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h_n}\right) \\
& \sim \frac{1}{\sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)}} \mathbb{E}(\varepsilon_{\mathbf{i}}^2) \sigma^2(\boldsymbol{\theta}, \mathbf{u}, \mathbf{x}) \int_{[0, h]^m} \prod_{k=1}^m K_1^2(\mathbf{z}) d\mathbb{P}(z_1, \dots, z_m). \tag{8.29}
\end{aligned}$$

The intricate relationship between weak convergence and \widehat{g}_1 is further substantiated. To elucidate this connection, we undertake the following steps:

- (1) Truncation of the function \widehat{g}_1 is performed, given the unbounded nature of the function class.
- (2) The convergence of the remainder term resulting from truncation to zero is established.
- (3) Hoeffding's decomposition is applied to the truncated part.
- (4) The convergence to zero of the non-linear term in this decomposition is validated.
- (5) The weak convergence to the linear term is established by demonstrating finite-dimensional convergence and asymptotic equicontinuity.

These steps closely parallel the proof strategy employed in Theorem 4.2. Consequently, the proof is concluded. \square

Proof of Theorem 4.2. Keep in mind that

$$\widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h_n}\right) K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right\} \varphi(Y_{\mathbf{i}, n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h_n}\right) K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right\}}. \tag{8.30}$$

Let's define

$$\mathfrak{G}_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) := \frac{\prod_{k=1}^m \left\{ K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right\} \varphi(Y_{\mathbf{i}, n})}{\mathbb{E} \prod_{k=1}^m \left\{ K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n}\right) \right\}} \text{ for } \mathbf{x} \in \mathcal{H}^m, \mathbf{y} \in \mathcal{Y}^m,$$

$$\mathcal{G} := \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}(\cdot, \cdot) : \varphi \in \mathcal{F}_m, \mathbf{i} = (i_1, \dots, i_m) \right\},$$

$$\mathcal{G}^{(k)} := \left\{ \pi_{k, m} \mathfrak{G}_{\varphi, \mathbf{i}}(\cdot, \cdot) : \varphi \in \mathcal{F}_m \right\},$$

$$\mathfrak{U}_n(\varphi, \mathbf{i}) = \mathfrak{U}_n^{(m)}(\mathfrak{G}_{\varphi, \mathbf{i}}) := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \xi_{i_k} \mathfrak{G}_{\varphi, \mathbf{i}}(\mathbf{X}_i, \mathbf{Y}_i),$$

and the U -empirical process is defined to be

$$\mu_n(\varphi, \mathbf{i}) := \sqrt{nh^m \phi_{\mathbf{x}, \boldsymbol{\theta}}^{1/m}(h_n)} \{ \mathfrak{U}_n(\varphi, \mathbf{i}) - \mathbb{E}(\mathfrak{U}_n(\varphi, \mathbf{i})) \}.$$

Subsequently, we have

$$\widetilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}, \boldsymbol{\theta}; h_n) = \frac{\mathfrak{U}_n(\varphi, \mathbf{i})}{\mathfrak{U}_n(1, \mathbf{i})}.$$

To ensure the weak convergence of our estimator, it is essential to first establish it for $\mu_n(\varphi, \mathbf{i})$. As mentioned earlier, dealing with unbounded function classes necessitates truncating the function $\mathfrak{G}_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y})$. Specifically, for $\lambda_n = n^{1/\zeta}$, where $\zeta > 2$, we have:

$$\begin{aligned}\mathfrak{G}_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) &= \mathfrak{G}_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y})\mathbb{1}_{\{F(\mathbf{y}) \leq \lambda_n\}} + \mathfrak{G}_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y})\mathbb{1}_{\{F(\mathbf{y}) > \lambda_n\}} \\ &:= \mathfrak{G}_{\varphi, \mathbf{i}}^{(T)}(\mathbf{x}, \mathbf{y}) + \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}(\mathbf{x}, \mathbf{y}).\end{aligned}$$

We can write the U -statistic as follows :

$$\begin{aligned}\mu_n(\varphi, \mathbf{i}) &= \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \left\{ \mathfrak{U}_n^{(m)}(\mathfrak{G}_{\varphi, \mathbf{i}}^{(T)}) - \mathbb{E} \left(\mathfrak{U}_n^{(m)}(\mathfrak{G}_{\varphi, \mathbf{i}}^{(T)}) \right) \right\} \\ &\quad + \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \left\{ \mathfrak{U}_n^{(m)}(\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}) - \mathbb{E} \left(\mathfrak{U}_n^{(m)}(\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}) \right) \right\} \\ &:= \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \left\{ \mathfrak{U}_n^{(T)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathfrak{U}_n^{(T)}(\varphi, \mathbf{i}) \right) \right\} \\ &\quad + \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \left\{ \mathfrak{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathfrak{U}_n^{(R)}(\varphi, \mathbf{i}) \right) \right\} \\ &:= \mu_n^{(T)}(\varphi, \mathbf{i}) + \mu_n^{(R)}(\varphi, \mathbf{i}).\end{aligned}\tag{8.31}$$

The first term is the truncated part and the second is the remaining one. We have to prove that:

- (1) $\mu_n^{(T)}(\varphi, \mathbf{i})$ converges to a Gaussian process.
- (2) The remainder part is negligible, in the sense that

$$\left\| \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \left\{ \mathfrak{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathfrak{U}_n^{(R)}(\varphi, \mathbf{i}) \right) \right\} \right\|_{\mathcal{F}_m, \mathfrak{S}_m^{\mathfrak{G}}} \xrightarrow{\mathbb{P}} 0.$$

For the initial step, we will utilize the Hoeffding decomposition, akin to the one presented in the previous Subsection 3.1, with the sole modification of replacing $\mathfrak{B}_{\mathbf{i}, n}$ with $\varphi(Y_{\mathbf{i}, n})$

$$\mathfrak{U}_n^{(T)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathfrak{U}_n^{(T)}(\varphi, \mathbf{i}) \right) := \mathfrak{U}_{1, n}(\varphi, \mathbf{i}) + \mathfrak{U}_{2, n}(\varphi, \mathbf{i}),$$

where

$$\mathfrak{U}_{1, n}(\varphi, \mathbf{i}) := \frac{1}{n} \sum_{i=1}^n \widehat{H}_{1, i}(\mathbf{u}, \mathbf{x}, \theta, \varphi),\tag{8.32}$$

$$\mathfrak{U}_{2, n}(\varphi, \mathbf{i}) := \frac{(n-m)!}{(n)!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H_{2, \mathbf{i}}(\mathbf{z}).\tag{8.33}$$

The convergence of $\mathfrak{U}_{2, n}(\varphi, \mathbf{i})$ to zero in probability has been established by Lemma 8.2. Therefore, it suffices to demonstrate the weak convergence of $\mathfrak{U}_{1, n}(\varphi, \mathbf{i})$ to a Gaussian process denoted as $\mathfrak{G}(\varphi)$. To accomplish this, we will proceed with finite-dimensional convergence and equicontinuity. Finite-dimensional convergence asserts that for every finite set of functions f_1, \dots, f_q in L_2 , where $\tilde{\mathfrak{U}}$ represents the centered form of \mathfrak{U} :

$$\left(\sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \tilde{\mathfrak{U}}_{1, n}(f_1, \mathbf{i}), \dots, \sqrt{nh^m \phi_{\mathbf{x}, \theta}^{1/m}(h_n)} \tilde{\mathfrak{U}}_{1, n}(f_q, \mathbf{i}) \right)\tag{8.34}$$

convergences to the corresponding finite-dimensional distributions of the process $\mathbb{G}(\varphi)$. We only need to demonstrate that for every fixed collection $(a_1, \dots, a_k) \in \mathbb{R}$, we have

$$\sum_{j=1}^q a_j \tilde{\mathfrak{U}}_{1,n}(f_j, \mathbf{i}) \rightarrow N(0, \sigma^2),$$

where

$$\sigma^2 = \sum_{j=1}^q a_j^2 \text{Var}(\tilde{\mathfrak{U}}_{1,n}(f_j, \mathbf{i})) + \sum_{s \neq r} a_s a_r \text{Cov}(\tilde{\mathfrak{U}}_{1,n}(f_s, \mathbf{i}), \tilde{\mathfrak{U}}_{1,n}(f_r, \mathbf{i})). \quad (8.35)$$

Take

$$\Psi(\cdot) = \sum_{j=1}^q a_j f_j(\cdot).$$

By linearity of $\Psi(\cdot)$, we have to see that $\tilde{\mathfrak{U}}_{1,n}(\Psi, \mathbf{i}) \rightarrow \mathbb{G}(\Psi)$. Let us denote

$$N = \mathbb{E} \left\{ \prod_{k=1}^m \left(K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right) \right\}.$$

We have

$$\begin{aligned} \tilde{\mathfrak{U}}_{1,n}(h_n, \mathbf{i}) &= N^{-1} \times \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{\substack{m-1 \\ n-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \\ &\quad \times \int h(y_1, \dots, y_{\ell-1}, Y_i, y_\ell, \dots, y_{m-1}) \prod_{\substack{k=1 \\ k \neq i}}^{m-1} \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_k}(x_k, v_k)}{h_n} \right) \\ &\quad \mathbb{P}(d(v_1, y_1), \dots, d(v_{\ell-1}, y_{\ell-1}), d(v_\ell, y_\ell), \dots, d(v_{m-1}, y_{m-1})), \\ &= N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \hat{\mathbb{H}}(Y_i). \end{aligned}$$

Now, we shall employ the blocking procedure for this empirical process. We intend to partition the set $\{1, \dots, n\}$ into $2v_n + 1$ subsets, each containing small and large blocks. In alignment with the notation used in Lemma 8.2, we denote the size of large blocks as a_n and the size of small blocks as b_n , satisfying:

$$v_n := \left\lfloor \frac{n}{a_n + b_n} \right\rfloor, \quad \frac{b_n}{a_n} \rightarrow 0, \quad \frac{a_n}{n} \rightarrow 0, \quad \frac{n}{a_n} \beta(b_n) \rightarrow 0. \quad (8.36)$$

In this case, we can see that :

$$\begin{aligned} \tilde{\mathfrak{U}}_{1,n}(h, \mathbf{j}) &= \sum_{j=1}^{v_n-1} \widehat{\mathfrak{U}}_{j,n}^{(1)} + \sum_{j=1}^{v_n-1} \widehat{\mathfrak{U}}_{j,n}^{(2)} + \widehat{\mathfrak{U}}_{j,n}^{(3)} \\ &:= \tilde{\mathfrak{U}}_{1,n}^{(1)} + \tilde{\mathfrak{U}}_{1,n}^{(2)} + \tilde{\mathfrak{U}}_{1,n}^{(3)}, \end{aligned} \quad (8.37)$$

where

$$\widehat{\mathfrak{U}}_{j,n}^{(1)} = N^{-1} \sum_{i=j(a_n+b_n)+1}^{j(a_n+b_n)+a_n} \xi_i \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \hat{\mathbb{H}}(Y_i), \quad (8.38)$$

$$\widehat{\mathfrak{U}}_{j,n}^{(2)} = N^{-1} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \frac{1}{\xi_i \phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i), \quad (8.39)$$

$$\widehat{\mathfrak{U}}_{j,n}^{(3)} = N^{-1} \sum_{i=v(a_n+b_n)+1}^n \xi_i \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i). \quad (8.40)$$

First, we aim to prove that

$$\frac{1}{n} \mathbb{E}(\tilde{\mathfrak{U}}_{1,n}^{(2)})^2 \rightarrow 0 \quad \text{and} \quad \frac{1}{n} \mathbb{E}(\tilde{\mathfrak{U}}_{1,n}^{(3)})^2 \rightarrow 0$$

to show that the case of summation over the small blocks and the summation over the last one is asymptotically negligible. Hence, we infer

$$\mathbb{E}(\tilde{\mathfrak{U}}_{1,n}^{(2)})^2 = \text{Var} \left(\sum_{j=1}^{v_n-1} \widehat{\mathfrak{U}}_{j,n}^{(2)} \right) = \sum_{j=1}^{v_n-1} \text{Var} \left(\widehat{\mathfrak{U}}_{j,n}^{(2)} \right) + \sum_{\substack{j=1 \\ j \neq k}}^{v_n-1} \sum_{k=1}^{v_n-1} \text{Cov} \left(\widehat{\mathfrak{U}}_{j,n}^{(2)}, \widehat{\mathfrak{U}}_{k,n}^{(2)} \right).$$

We have :

$$\begin{aligned} \text{Var} \left(\widehat{\mathfrak{U}}_{j,n}^{(2)} \right) &= \text{Var} \left(N^{-1} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \xi_i \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i) \right) \\ &= N^{-2} \frac{1}{\phi^2(h_n)} \frac{a_n}{a_n} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \xi_i^2 \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i) \right) \\ &\lesssim \frac{b_n}{\phi^2(h_n) (m\phi^{m-1}(h_n)/nh_n + \phi^m(h_n))^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{\omega}{h_n} \right) d\omega \right] \\ &\quad \times \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i) \right). \quad (\text{Using Lemma 8.1 ii}) \end{aligned}$$

Thus, we have

$$\begin{aligned} \sum_{j=1}^{v_n-1} \text{Var} \left(\widehat{\mathfrak{U}}_{j,n}^{(2)} \right) &\lesssim v_n b_n \frac{1}{\phi^{2(m+1)}(h_n)} \left[\int_{[0,h]} K_1^2 \left(\frac{\omega}{h_n} \right) d\omega \right] \times \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i) \right) \\ &\sim \frac{n}{a_n + b_n} b_n \sim \frac{nb_n}{a_n} = o_{\mathbb{P}}(n), \quad (\text{by (8.36)}). \end{aligned} \quad (8.41)$$

and

$$\begin{aligned} \sum_{\substack{j=1 \\ j \neq k}}^{v_n-1} \sum_{k=1}^{v_n-1} \text{Cov} \left(\widehat{\mathfrak{U}}_{j,n}^{(2)}, \widehat{\mathfrak{U}}_{k,n}^{(2)} \right) &= \sum_{\substack{j=1 \\ j \neq k}}^{v_n-1} \sum_{k=1}^{v_n-1} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \sum_{i'=k(a_n+b_n)+a_n+1}^{(k+1)(a_n+b_n)} \frac{N^{-2}}{\phi(h_n)} \xi_i \xi_{i'} \\ &\text{Cov} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}(Y_i), K_2 \left(\frac{d_{\theta_{i'}}(x_{i'}, X_{i'})}{h_n} \right) \tilde{\mathfrak{H}}(Y_{i'}) \right) \\ &= \sum_{\substack{j=1 \\ j \neq k}}^{v_n-1} \sum_{k=1}^{v_n-1} \sum_{l_1=1}^{b_n} \sum_{l_2=1}^{b_n} \frac{1}{N^2 \phi^2(h_n)} \xi_{\lambda_i+l_1} \xi_{\lambda_{i'}+l_2} \end{aligned}$$

$$\text{Cov} \left(K_2 \left(\frac{d_{\theta_{\lambda_i+l_1}}(x_{\lambda_i+l_1}, X_{\lambda_i+l_1})}{h_n} \right) \tilde{\mathbb{H}}(Y_{\lambda_i+l_1}), K_2 \left(\frac{d_{\theta_{\lambda_{i'}+l_2}}(x_{\lambda_{i'}+l_2}, X_{\lambda_{i'}+l_2})}{h_n} \right) \tilde{\mathbb{H}}(Y_{\lambda_{i'}+l_2}) \right),$$

where $\lambda_i = j(a_n + b_n) + a_n$, but for $j \neq k$, $|\lambda_i - \lambda_{i'} + l_1 - l_2| \geq b_n$, then

$$\left| \sum_{\substack{j=1 \\ j \neq k}}^{v_n-1} \sum_{k=1}^{v_n-1} \text{Cov}(\tilde{\mathbb{U}}_{j,n}^{(2)}, \tilde{\mathbb{U}}_{k,n}^{(2)}) \right| \leq \sum_{\substack{j=1 \\ |j-k| \geq b_n}}^{v_n-1} \sum_{k=1}^{v_n-1} \frac{1}{N^2 \phi^2(h_n)} \xi_j \xi_k \left| \text{Cov} \left(K_2 \left(\frac{d_{\theta_j}(x_j, X_j)}{h_n} \right) \tilde{\mathbb{H}}(Y_j), K_2 \left(\frac{d_{\theta_k}(x_k, X_k)}{h_n} \right) \tilde{\mathbb{H}}(Y_k) \right) \right|,$$

here, the use of Davydov's lemma (Lemma A.7) is necessary, we have

$$\begin{aligned} & \left| \text{Cov} \left(K_2 \left(\frac{d_{\theta_j}(x_j, X_j)}{h_n} \right) \tilde{\mathbb{H}}(Y_j), K_2 \left(\frac{d_{\theta_k}(x_k, X_k)}{h_n} \right) \tilde{\mathbb{H}}(Y_k) \right) \right| \\ & \leq 8 \left(\mathbb{E} \left| K_2 \left(\frac{d_{\theta_j}(x_j, X_j)}{h_n} \right) \right|^p \right)^{1/p} \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} \beta(|i-j|)^{1-2/p} \\ & \lesssim \phi(h_n) \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} \beta(|i-j|)^{1-2/p}, \end{aligned}$$

it follows that

$$\begin{aligned} \left| \sum_{\substack{j=1 \\ j \neq k}}^{v_n-1} \sum_{k=1}^{v_n-1} \text{Cov}(\tilde{\mathbb{U}}_{j,n}^{(2)}, \tilde{\mathbb{U}}_{k,n}^{(2)}) \right| & \lesssim \sum_{\substack{j=1 \\ |j-k| \geq b_n}}^{v_n-1} \sum_{k=1}^{v_n-1} \frac{1}{N^2 \phi^2(h_n)} \xi_j \xi_k \phi(h_n) \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} \beta(|i-j|)^{1-2/p} \\ & \lesssim \frac{1}{b_n^2 N^2 \phi^2(h_n)} \phi(h_n) \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} \sum_{l=b_n+1}^{\infty} l^p \beta(l)^{1-2/p} \\ & \lesssim \frac{1}{b_n^2 N^2 \phi^2(h_n)} \phi(h_n) \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} n \varrho = o_{\mathbb{P}}(n), \end{aligned} \quad (8.42)$$

where the last inequality follows also from (8.36) and the size of b_n . Then, (8.41) and (8.42) shows us that

$$\frac{1}{n} \mathbb{E}(\tilde{\mathbb{U}}_{1,n}^{(2)})^2 \rightarrow 0.$$

Using the same footsteps, we find that

$$\begin{aligned} \text{Var}(\tilde{\mathbb{U}}_{1,n}^{(3)}) & = \text{Var} \left(N^{-1} \sum_{i=v(a_n+b_n)+1}^n \xi_i \frac{1}{\phi(h_n)} K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \\ & = N^{-2} \sum_{i=v(a_n+b_n)+1}^n \xi_i^2 \frac{1}{\phi^2(h_n)} \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \\ & \quad + \frac{1}{N^2 \phi^2(h_n)} \sum_{i=v(a_n+b_n)+1}^n \sum_{\substack{j=v(a_n+b_n)+1 \\ |i-j| > 0}}^n \xi_i \xi_j \text{Cov} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i), K_2 \left(\frac{d_{\theta_j}(x_j, X_j)}{h_n} \right) \tilde{\mathbb{H}}(Y_j) \right) \\ & = N^{-2} \sum_{i=v(a_n+b_n)+1}^n \xi_i^2 \frac{1}{\phi^2(h_n)} \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{N^2 \phi^2(h_n)} \sum_{l_1=1}^{n-\nu(a_n+b_n)} \sum_{\substack{l_2=1 \\ |i-j|>0}}^{n-\nu(a_n+b_n)} \xi_i \xi_j \\
 & \quad \text{Cov} \left(K_2 \left(\frac{d_{\theta_{\lambda_i+l_1}}(x_{\lambda_i+l_1}, X_{\lambda_i+l_1})}{h_n} \right) \tilde{\mathbb{H}}(Y_{\lambda_i+l_1}), K_2 \left(\frac{d_{\theta_{\lambda_i+l_2}}(x_{\lambda_i+l_2}, X_{\lambda_i+l_2})}{h_n} \right) \tilde{\mathbb{H}}(Y_{\lambda_i+l_2}) \right) \\
 & \hspace{20em} \text{(For } \lambda_i := n - \nu(a_n + b_n)) \\
 \lesssim & \frac{n - \nu(a_n + b_n)}{\phi^2(h_n) (m\phi^{m-1}(h_n)/nh_n + \phi^m(h_n))^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{t}{h_n} \right) dt \right] \times \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \\
 & + \frac{1}{(n - \nu(a_n + b_n))^{\rho} (m\phi^{m-1}(h_n)/nh_n + \phi^m(h_n))^2 \phi^2(h_n)} \phi(h_n) \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} \\
 & \times \sum_{l=(n-\nu(a_n+b_n))+1}^{\infty} l^{\rho} \beta(l)^{1-2/p} \\
 & \hspace{15em} \text{(Using Lemma 8.1 ii) and Lemma A.7)} \\
 \lesssim & \frac{n - \nu_n(a_n + b_n)}{\phi^2(h_n) (m\phi^{m-1}(h_n)/nh_n + \phi^m(h_n))^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{t}{h_n} \right) dt \right] \times \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \\
 & + \frac{1}{(n - \nu(a_n + b_n))^{\rho} (m\phi^{m-1}(h_n)/nh_n + \phi^m(h_n))^2 \phi^2(h_n)} \phi(h_n) \mathbb{E}(|\tilde{\mathbb{H}}(Y_j)|^p)^{1/p} n \rho. \tag{8.43}
 \end{aligned}$$

By (8.36), we find that

$$\frac{1}{n} \text{Var}(\tilde{\mathfrak{U}}_{i,n}^{(3)}) \rightarrow 0.$$

Now, we need to establish that the summands in $\tilde{\mathfrak{U}}_{1,n}^{(1)}$ are asymptotically independent, allowing us to apply the conditions of Lindeberg-Feller for asymptotic finite normality. We can utilize Lemma A.8, where $\tilde{\mathfrak{U}}_{i_a,n}^{(1)}$ is $\mathcal{F}_{i_a}^{j_a}$ -measurable with $i_a = a(a_n + b_n) + 1$ and $j_a = a(a_n + b_n) + a_n$, giving us

$$\left| \mathbb{E} \left(\exp(itn^{-1/2} \tilde{\mathfrak{U}}_{1,n}^{(1)}) - \prod_{i=0}^{\nu_n-1} \mathbb{E} \left(\exp(itn^{-1/2} \tilde{\mathfrak{U}}_{i,n}^{(1)}) \right) \right| \leq 16\nu_n \beta(b_n + 1), \tag{8.44}$$

which tends to zero using (8.36), then the asymptotic independence is achieved. We can see also that

$$\begin{aligned}
 \frac{1}{n} \text{Var}(\tilde{\mathfrak{U}}_{1,n}^{(1)}) & \lesssim \frac{\nu_n a_n}{n\phi^2(h_n)N^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{t}{h_n} \right) dt \right] \times \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \\
 & \rightarrow \frac{1}{\phi^2(h_n)N^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{t}{h_n} \right) dt \right] \times \text{Var} \left(K_2 \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathbb{H}}(Y_i) \right) \text{ (since } \nu_n a_n/n \rightarrow 1). \\
 & := \mathbb{V}(X, Y). \tag{8.45}
 \end{aligned}$$

Up to this point, we have addressed the final condition for finite-dimensional convergence. It is important to observe that, for sufficiently large n , the set

$$\{|\tilde{\mathfrak{U}}_{i,n}^{(1)}| > \epsilon \mathbb{V}(X, Y) \sqrt{n}\}$$

is empty, thus

$$\frac{1}{n} \sum_{i=0}^{\nu_n-1} \mathbb{E} \left(\tilde{\mathfrak{U}}_{i,n}^{(1)2} \mathbb{1}_{\{|\tilde{\mathfrak{U}}_{i,n}^{(1)}| > \epsilon \mathbb{V}(X, Y) \sqrt{n}\}} \right) \rightarrow 0. \tag{8.46}$$

Therefore, we conclude the proof of finite-dimensional convergence. Now, we shift our focus to asymptotic equicontinuity, aiming to demonstrate that:

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sqrt{nh^m \phi_{x,\theta}^{1/m}(h_n)} \left\| \tilde{\mathfrak{U}}_{1,n}(h_n, \mathbf{i}) \right\|_{\mathcal{FK}(\delta, \|\cdot\|_{\zeta})} > \epsilon \right\} = 0, \quad (8.47)$$

where

$$\begin{aligned} \mathcal{FK}(\delta, \|\cdot\|_{\zeta}) &:= \left\{ \tilde{\mathfrak{U}}'_{1,n}(h_n, \mathbf{i}) - \tilde{\mathfrak{U}}''_{1,n}(h_n, \mathbf{i}) : \right. \\ &\quad \left. \left\| \tilde{\mathfrak{U}}'_{1,n}(h_n, \mathbf{i}) - \tilde{\mathfrak{U}}''_{1,n}(h_n, \mathbf{i}) \right\| < \delta, \tilde{\mathfrak{U}}'_{1,n}(h_n, \mathbf{i}), \tilde{\mathfrak{U}}''_{1,n}(h_n, \mathbf{i}) \in \mathcal{FK} \right\}, \end{aligned} \quad (8.48)$$

for

$$\begin{aligned} \tilde{\mathfrak{U}}'_{1,n}(h_n, \mathbf{i}) &= N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h_n)} K_{2,1} \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}_1(Y_i) - \mathbb{E}(\mathfrak{U}'_{1,n}(h_n, \mathbf{i})), \\ \tilde{\mathfrak{U}}''_{1,n}(h_n, \mathbf{i}) &= N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h_n)} K_{2,2} \left(\frac{d_{\theta_i}(x_i, X_i)}{h_n} \right) \tilde{\mathfrak{H}}_2(Y_i) - \mathbb{E}(\mathfrak{U}''_{1,n}(h_n, \mathbf{i})). \end{aligned} \quad (8.49)$$

Now, we will employ the chaining technique from [13] and follow the approach outlined in [39] for the conditional setting. The fundamental idea is to decompose a sequence (X_1, \dots, X_n) into $2\nu_n$ equal-sized blocks, each of length a_n , and a residual block of length $n - 2\nu_n a_n$, where, for $1 \leq j \leq \nu_n$:

$$\begin{aligned} H_j &= \{i : 2(j-1)a_n + 1 \leq i \leq (2j-1)a_n\}, \\ T_j &= \{i : (2j-1)a_n + 1 \leq i \leq 2ja_n\}, \\ R &= \{i : (2\nu_n a_n + 1 \leq i \leq n)\}. \end{aligned}$$

The values of ν_n, a_n are given in the following. Another ingredient is essential, in this proof, that is a sequence of independent blocks $(\zeta_1, \dots, \zeta_n)$ such as:

$$\mathcal{L}(\zeta_1, \dots, \zeta_n) = \mathcal{L}(X_1, \dots, X_{a_n}) \times \mathcal{L}(X_{a_n+1}, \dots, X_{2a_n}) \times \dots.$$

In the same line as [39], the results of the work of [73] on β -mixing are applied, and get, for any measurable set A :

$$\begin{aligned} &\left| \mathbb{P} \left\{ (\zeta_1, \dots, \zeta_{a_n}, \zeta_{2a_n+1}, \dots, \zeta_{3a_n}, \dots, \zeta_{2(\nu_n-1)a_n+1}, \dots, \zeta_{2\nu_n a_n}) \in A \right\} \right. \\ &\quad \left. - \mathbb{P} \left\{ (X_1, \dots, X_{a_n}, X_{2a_n+1}, \dots, X_{3a_n}, \dots, X_{2(\nu_n-1)a_n+1}, \dots, X_{2\nu_n a_n}) \in A \right\} \right| \\ &\leq 2(\nu_n - 1)\beta_{a_n}. \end{aligned} \quad (8.50)$$

We will focus solely on the independent blocks represented as $\zeta_i = (\eta_i, \varsigma_i)$ sequences, rather than dealing with the dependent variables. We will utilize a strategy akin to the one employed in Lemma 8.2 to transition from the sequence of locally stationary random variables to the stationary one:

$$\mathbb{P} \left\{ \left\| (n\phi(h_n))^{-1/2} h_n^{m/2} N^{-1} \sum_{j=1}^n \left(\xi_j K_2 \left(\frac{d_{\theta_j}(x_j, X_j)}{h_n} \right) \tilde{h}(Y_j) - \mathbb{E}(\mathfrak{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{FK}(b, \|\cdot\|_{\zeta})} > \epsilon \right\}$$

$$\leq 2\mathbb{P} \left\{ \left\| (n\phi(h_n))^{-1/2} h_n^{m/2} N^{-1} \sum_{j=1}^{v_n} \sum_{i \in H_j} \left(\xi_i K_2 \left(\frac{d_{\theta_i}(x_i, \eta_i)}{h_n} \right) \tilde{h}(\zeta_i) - \mathbb{E}(\mathfrak{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{F}, \mathcal{H}(b, \|\cdot\|_{\zeta})} > \epsilon' \right\} + 2(v_n - 1)\beta_{a_n} + o_{\mathbb{P}}(1). \quad (8.51)$$

We choose

$$a_n = [(\log n)^{-1} (n^{p-2} \phi^p(h_n))^{1/2(p-1)}]$$

and

$$v_n = \left\lfloor \frac{n}{2a_n} \right\rfloor - 1.$$

Exploiting condition (v) from Assumption 4, we deduce that $(v_n - 1)\beta_{a_n} \rightarrow 0$ as $n \rightarrow \infty$. This primarily pertains to the first term on the right-hand side of (8.51). Given the independence of the blocks, we symmetrize the sequence using a set $\{\epsilon_j\}_{j \in \mathbb{N}^*}$ of i.i.d. Rademacher variables, where

$$\mathbb{P}(\epsilon_j = 1) = \mathbb{P}(\epsilon_j = -1) = 1/2.$$

It is noteworthy that the sequence $\{\epsilon_j\}_{j \in \mathbb{N}^*}$ is independent of the sequence $\{\xi_i = (\zeta_i, \zeta_i)\}_{i \in \mathbb{N}^*}$. Now, our objective is to establish, for all $\epsilon > 0$,

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left\| (n\phi(h_n))^{-1/2} h_n^{m/2} N^{-1} \sum_{j=1}^{v_n} \sum_{i \in H_j} \left(\xi_i K_2 \left(\frac{d_{\theta_i}(x_i, \eta_i)}{h_n} \right) \tilde{h}(\zeta_i) - \mathbb{E}(\mathfrak{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{F}, \mathcal{H}(b, \|\cdot\|_{\zeta})} > \epsilon \right\} \quad (8.52)$$

Define the semi-norm:

$$\begin{aligned} \tilde{d}_{n\phi,2} &:= \left((n\phi(h_n))^{-1/2} h_n^{m/2} N^{-1} \sum_{j=1}^{v_n} \sum_{i \in H_j} \left| \left(\xi_i K_{2,1} \left(\frac{d_{\theta_i}(x_i, \eta_i)}{h_n} \right) \tilde{h}_1(\zeta_i) - \mathbb{E}(\mathfrak{U}'_{1,n}(h_n, \mathbf{i})) \right) \right. \right. \\ &\quad \left. \left. - \left(\xi_i K_{2,2} \left(\frac{d_{\theta_i}(x_i, \eta_i)}{h_n} \right) \tilde{h}_2(\zeta_i) - \mathbb{E}(\mathfrak{U}''_{1,n}(h_n, \mathbf{i})) \right) \right|^2 \right)^{1/2}, \end{aligned} \quad (8.53)$$

and the covering number defined for any class of functions \mathcal{E} by :

$$\tilde{N}_{n\phi,2}(u, \mathcal{E}) := N_{n\phi,2}(u, \mathcal{E}, \tilde{d}_{n\phi,2}).$$

Building upon the preceding discussion, we can bound (8.52), with further elaboration available in [39]. Similarly, by following the methodology in [39] and referring back to [13], the independence between the blocks, coupled with Assumption 6 ii) and the utilization of [89, Lemma 5.2], ensures equicontinuity, thus laying the foundation for weak convergence. Now, our task is to establish that:

$$\mathbb{P} \left\{ \left\| \mu_n^{(R)}(\varphi, \mathbf{t}) \right\|_{\mathcal{F}_m, \mathcal{S}_{\Theta}^m} > \lambda \right\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In this proof, for clarity, we present the case where $m = 2$, with different sizes for a_n and b_n , where b_n denotes the size of the alternate blocks. Both a_n and b_n satisfy:

$$b_n \ll a_n, \quad (v_n - 1)(a_n + b_n) < n \leq v_n(a_n + b_n), \quad (8.54)$$

and set, for $1 \leq j \leq v_n - 1$:

$$\begin{aligned} \mathbf{H}_j^{(U)} &= \{i : (j-1)(a_n + b_n) + 1 \leq i \leq (j-1)(a_n + b_n) + a_n\}, \\ \mathbf{T}_j^{(U)} &= \{i : (j-1)(a_n + b_n) + a_n + 1 \leq i \leq (j-1)(a_n + b_n) + a_n + b_n\}, \\ \mathbf{H}_{v_n}^{(U)} &= \{i : (v_n - 1)(a_n + b_n) + 1 \leq i \leq n \wedge (v_n - 1)(a_n + b_n) + a_n\}, \\ \mathbf{T}_{v_n}^{(U)} &= \{i : (v_n - 1)(a_n + b_n) + a_n + 1 \leq i \leq n\}. \end{aligned}$$

We have:

$$\begin{aligned} \mu_n^{(R)}(\varphi, \mathbf{i}) &= \sqrt{n\phi_{\mathbf{x},\theta}^{1/m}(h_n)} \left\{ \mathfrak{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathfrak{U}_n^{(R)}(\varphi, \mathbf{i}) \right) \right\} \\ &= \frac{\sqrt{n\phi_{\mathbf{x},\theta}^{1/m}(h_n)}}{n(n-1)} \sum_{i_1 \neq i_2}^n \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left(((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &\leq \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in \mathbf{H}_p^{(U)}} \sum_{i_2 \in \mathbf{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &\quad + \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2, i_1, i_2 \in \mathbf{H}_p^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &\quad + 2 \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in \mathbf{H}_p^{(U)}} \sum_{q: |q-p| \geq 2} \sum_{i_2 \in \mathbf{T}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &\quad + 2 \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in \mathbf{H}_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in \mathbf{T}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &\quad + \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in \mathbf{T}_p^{(U)}} \sum_{i_2 \in \mathbf{T}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &\quad + \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2, i_1, i_2 \in \mathbf{T}_p^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \\ &\quad \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{t}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \\ &=: \text{I}' + \text{II}' + \text{III}' + \text{IV}' + \text{V}' + \text{VI}'. \end{aligned}$$

We will employ blocking arguments and address the resulting terms. Let's begin by examining the first term I' . We have:

$$\begin{aligned} & \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right. \right. \right. \\ & \qquad \qquad \qquad \left. \left. \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{K}^2} > \delta \right\} \\ & \leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right. \right. \right. \\ & \qquad \qquad \qquad \left. \left. \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{K}^2} > \delta \right\} + 2\nu_n \beta(b_n). \end{aligned}$$

Notice that $\nu_n \beta_{b_n} \rightarrow 0$ and recall that for all $\varphi \in \mathcal{F}_m$, and :

$$\theta, \mathbf{x} \in \mathcal{H}^2, \mathbf{y} \in \mathcal{Y}^2 : \mathbb{1}_{\{d_\theta(\mathbf{x}, X_{i,n}) \leq h\}} F(\mathbf{y}) \geq \varphi(\mathbf{y}) K_2 \left(\frac{d_\theta(\mathbf{x}, X_{i,n})}{h_n} \right).$$

Hence, by the symmetry of $F(\cdot)$:

$$\begin{aligned} & \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)}((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{t}}^{(R)}((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \\ & \lesssim \left| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right] \right\} \right|. \end{aligned} \tag{8.55}$$

We are going to use Chebyshev's inequality, Hoeffding's trick and inequality, respectively to obtain:

$$\begin{aligned} & \mathbb{P} \left\{ \left| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right] \right\} \right| > \delta \right\} \\ & \lesssim \delta^{-2} n^{-1} \phi^{-1}(h_n) \text{Var} \left(\sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right) \\ & \lesssim c_2 \nu_n \delta^{-2} n^{-1} \phi^{-1}(h_n) \text{Var} \left(\sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta'_j) \mathbb{1}_{\{F > \lambda_n\}} \right) \\ & \lesssim 2c_2 \nu_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \mathbb{E} \left[(F(\zeta_1, \zeta_2))^2 \mathbb{1}_{\{F > \lambda_n\}} \right]. \end{aligned} \tag{8.56}$$

Under Assumption 6 iii), we have for each $\lambda > 0$:

$$c_2 \nu_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \mathbb{E} \left[(F(\zeta_1, \zeta_2))^2 \mathbb{1}_{\{F > \lambda_n\}} \right]$$

$$\begin{aligned}
 &= c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_0^\infty \mathbb{P} \left\{ (F(\zeta_1, \zeta_2))^2 \mathbb{1}_{\{F > \lambda_n\}} \geq t \right\} dt \\
 &= c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_0^{\lambda_n} \mathbb{P} \{ F > \lambda_n \} dt \\
 &\quad + c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_{\lambda_n}^\infty \mathbb{P} \{ (F)^2 > t \} dt,
 \end{aligned}$$

which tends to 0 as $n \rightarrow \infty$. Terms II', V', and VI' will be handled similarly to the previous term. However, II' and VI' deviate from this pattern due to the variables $\{\zeta_i, \zeta_j\}_{i,j \in H_p^{(U)}}$ (or $\{\zeta_i, \zeta_j\}_{i,j \in T_p^{(U)}}$ for VI') belonging to the same blocks. Regarding term IV', its analysis can be inferred from the study of terms I' and III'. Considering term III', we have:

$$\begin{aligned}
 &\mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) \right. \right. \right. \\
 &\quad \left. \left. \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\| \right\|_{\mathcal{F}_2, \mathcal{H}^2} > \delta \right\} \\
 &\leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right. \right. \right. \\
 &\quad \left. \left. \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right] \right\| \right\|_{\mathcal{F}_2, \mathcal{H}^2} > \delta \right\} + \frac{v_n^2 a_n b_n \beta(a_n)}{\sqrt{n\phi(h_n)}}. \tag{8.57}
 \end{aligned}$$

We also have

$$\begin{aligned}
 &\mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right. \right. \right. \\
 &\quad \left. \left. \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right] \right\| \right\|_{\mathcal{F}_2, \mathcal{H}^2} > \delta \right\} \\
 &\leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ \mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right. \right. \right. \\
 &\quad \left. \left. \left. - \mathbb{E} \left[\mathfrak{G}_{\varphi, \mathbf{i}}^{(R)} \left((S_{i_1}, S_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right] \right\| \right\|_{\mathcal{F}_2, \mathcal{H}^2} > \delta \right\}.
 \end{aligned}$$

Since the Eq (8.55) is still satisfied, the problem is reduced to

$$\mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right. \right. \right.$$

$$\begin{aligned}
& \left. -\mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right] \right\} > \delta \Bigg\} \\
& \lesssim \delta^{-2} n^{-1} \phi(h_n) \text{Var} \left(\sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right),
\end{aligned}$$

we follow the same procedure as in (8.56). The rest has just been shown to be asymptotically negligible. Finally, with

$$|r^{(m)}(\varphi, \theta, \mathbf{u}, \mathbf{x}) - \mathbb{E}(\mathfrak{U}_n(\varphi, \mathbf{i}))| \rightarrow 0,$$

and for

$$(\mathfrak{U}_n(1, \mathbf{i})) \xrightarrow{\mathbb{P}} 1,$$

the weak convergence of our estimator is accomplished. \square

Technical lemmas

The forthcoming proof relies on the arguments delineated in [39, 40, 165], extended to the single index model framework.

Lemma 8.1. *Let $K_2(\cdot)$ denote one dimensional kernel function satisfying Assumption 2 part i), if Assumption 1, then:*

$$i) \quad \mathbb{E} \left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| \lesssim \frac{m\phi^{m-1}(h_n)}{nh_n}; \quad (8.58)$$

$$ii) \quad \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right] \lesssim \frac{m\phi^{m-1}(h_n)}{nh_n} + \phi^m(h_n); \quad (8.59)$$

$$iii) \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right] \sim \phi^m(h_n). \quad (8.60)$$

Proof of Lemma 8.1. For the first inequality *i)*, by assuming that the kernel function $K_2(\cdot)$ is an asymmetrical triangle kernel, that is, $K_2(x) = (1-x)\mathbb{1}_{(x \in [0,1])}$, we have

$$\begin{aligned}
& \mathbb{E} \left| \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| \\
& = \mathbb{E} \left[\left| \sum_{k=1}^m \left(K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right) \right| \right] \\
& \quad \times \left| \prod_{i=1}^{k-1} K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right| \times \left| \prod_{j=k+1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})}{h_n} \right) \right| \\
& \hspace{15em} \text{(Using a telescoping argument)}
\end{aligned}$$

$$\begin{aligned}
&\leq \left\{ \mathbb{E} \left[\left| \sum_{k=1}^m \left(K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) - K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right) \right|^3 \right] \right\}^{1/3} \\
&\quad \times \left\{ \mathbb{E} \left[\prod_{i=1}^{k-1} K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right]^3 \right\}^{1/3} \times \left\{ \mathbb{E} \left[\prod_{j=k+1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right]^3 \right\}^{1/3} \\
&\hspace{20em} \text{(By Hölder's inequality)} \\
&\leq \left\{ \mathbb{E} \left[\sum_{k=1}^m \left| \left(K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) - K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right) \right|^3 \right] \right\}^{1/3} \\
&\quad \times \left\{ \prod_{i=1}^{k-1} \left(\mathbb{E} \left| K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right|^{3p_i} \right)^{1/p_i} \right\}^{1/3} \\
&\quad \times \left\{ \prod_{j=k+1}^m \mathbb{E} \left(\left| K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right|^{3q_j} \right)^{1/q_j} \right\}^{1/3} \quad \text{(By Hölder's inequality)} \\
&\lesssim \left\{ \sum_{k=1}^m \mathbb{E} \left| \frac{1}{nh_n} U_{i_k,n}^{(i_k/n)} \right|^3 \right\}^{1/3} \times \left\{ \prod_{i=1}^{k-1} \left(\mathbb{E} \left| \mathbb{1}_{\{d_{\theta_k}(x_k, X_{i_k,n}) \leq h\}} \right|^{3p_i} \right)^{1/p_i} \right\} \\
&\quad \times \left\{ \prod_{j=k+1}^m \left(\mathbb{E} \left| \mathbb{1}_{\{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)}) \leq h\}} \right|^{3q_j} \right)^{1/q_j} \right\}^{1/3} \quad \text{(By Assumption 1)} \\
&\lesssim \left\{ \sum_{k=1}^m \frac{1}{n^3 h^3} \mathbb{E} |U_{i_k,n}^{(i_k/n)}|^3 \right\}^{1/3} \times \left\{ \prod_{i=1}^{k-1} (F^{3p_i}(h, x_k))^{1/p_i} \prod_{j=k+1}^m (F^{3q_j}(h, x_k))^{1/q_j} \right\}^{1/3} \\
&\lesssim \frac{1}{nh_n} \left\{ \sum_{k=1}^m \mathbb{E} |U_{i_k,n}^{(i_k/n)}|^3 \right\}^{1/3} \times \left\{ \prod_{i=1}^{k-1} C_d \phi^3(h_n) f_1^3(x_k) \times \prod_{j=k+1}^m C_d \phi(h_n)^3 f_1^3(x_k) \right\}^{1/3} \\
&\lesssim \frac{m\phi^{m-1}(h_n)}{nh_n}. \tag{8.61}
\end{aligned}$$

For the second inequality *ii*), we have:

$$\begin{aligned}
&\mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right] \\
&= \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) - \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) + \prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right].
\end{aligned}$$

By linearity of the expectation, inequality *i*) and for

$$\mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i_k/n)})}{h_n} \right) \right] \lesssim \phi^{m-1}(h_n),$$

using Assumption 1 part iv), the proof of this inequality holds. Now, we consider the last one. Set

$$\widetilde{K}_2^2(x_k) := K_2^2(d_{\theta_k}(x_k, X_{i_k, n}^{(i_k/n)})).$$

We have

$$\begin{aligned} \mathbb{E} \left[\left(\prod_{k=1}^m \widetilde{K}_2^2 \left(\frac{x_k}{h_n} \right) \right) \right] &= \int_0^h \cdots \int_0^h \prod_{k=1}^m \widetilde{K}_2^2 \left(\frac{y_k}{h_n} \right) \mathbb{P}(dy_1, \dots, dy_m) \\ &= -\frac{2}{h_n} \int_0^h \cdots \int_0^h \prod_{j=1, j \neq k}^m \widetilde{K}_2(y_j) \\ &\quad \times \int_0^h \widetilde{K}_2(y_\ell) \widetilde{K}_2'(y_\ell) \mathbb{P}(dy_1, \dots, dy_{\ell-1}, y_\ell, dy_{\ell+1}, \dots, y_m) dy_\ell \quad (\text{Integration by parts}) \\ &= \frac{(-2)^m}{h^m} \int_0^h \cdots \int_0^h \prod_{k=1}^m \widetilde{K}_2(y_k) \widetilde{K}_2'(y_k) \mathbb{P}(y_1, \dots, y_m) dy_1 \cdots dy_m \\ &\sim \frac{2^m}{h^m} \int_0^h \cdots \int_0^h \prod_{k=1}^m \widetilde{K}_2(y_k) \widetilde{K}_2'(y_k) \phi_k(y_k) d\mathbb{P}(y_1, \dots, y_m) \\ &= \frac{2^m}{h^m} \int_0^h \cdots \int_0^h \prod_{k=1}^m \left(1 - \frac{y_k}{h_n} \right) \phi_k(y_k) d\mathbb{P}(y_1, \dots, y_m) \\ &\quad (\text{Using Assumption 2 ii) and } K_2(x) = (1-x)I(x \in [0, 1])) \\ &= \frac{2^m}{h^{2m}} \int_0^h \cdots \int_0^h \left(\int_0^y \cdots \int_0^y \prod_{k=1}^m \phi_k(\epsilon_k) d(\epsilon_1, \dots, \epsilon_m) \right) d\mathbb{P}(y_1, \dots, y_m) \\ &\quad (\text{By an integration by parts}) \\ &\sim \frac{2^m}{h^{2m}} \int_0^h \cdots \int_0^h \prod_{k=1}^m y_k \phi_k(y_k) d\mathbb{P}(y_1, \dots, y_m) \sim \frac{1}{h^{2m}} \phi^m(h_n) h^{2m} \\ &\sim \phi^m(h_n). \end{aligned}$$

The final result is established by utilizing the small-ball lower bound provided in (2.11). Consequently, inequality (8.60) follows. \square

In the ensuing discussion, we will present a lemma that can be regarded as a technical result in the proof of our proposition.

Lemma 8.2. Consider $\mathcal{F}_m \mathfrak{R}_\Theta^m$ as a uniformly bounded class of measurable canonical functions, where $m \geq 2$. Suppose there exist finite constants \mathbf{a} and \mathbf{b} such that the $\mathcal{F}_m \mathfrak{R}_\Theta^m$ covering number satisfies:

$$N(\epsilon, \mathcal{F}_m \mathfrak{R}_\Theta^m, \|\cdot\|_{L_2(Q)}) \leq \mathbf{a} \epsilon^{-\mathbf{b}}, \quad (8.62)$$

for every $\epsilon > 0$ and every probability measure Q . If the mixing coefficients β of the local stationary sequence $\{Z_i = (X_{i,n}, \mathfrak{B}_{i,n})\}_{i \in \mathbb{N}^*}$ satisfy:

$$\beta(k)k^r \rightarrow 0, \quad \text{as } k \rightarrow \infty, \quad (8.63)$$

for some $r > 1$, then:

$$\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in \mathcal{B}^m} \mathbb{P} \left[h_n^{m/2} \phi^{m/2}(h_n) n^{-m+1/2} \sum_{i \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H(Z_{i_1}, \dots, Z_{i_m}) \right] \rightarrow 0. \quad (8.64)$$

Remark 8.3. As mentioned before, $\mathfrak{B}_{i,n}$ will be equal to 1 or $\varepsilon_{i,n} = \sigma\left(\frac{i}{n}, X_{i,n}\right) \varepsilon_i$. In the proof of the previous Lemma, $\mathfrak{B}_{i,n}$ will be equal $\varepsilon_{i,n} = \sigma\left(\frac{i}{n}, X_{i,n}\right) \varepsilon_i$, and we will use the notation $\mathfrak{B}_{i,n}^{(u)}$ to indicate $\sigma(\mathbf{u}, \mathbf{x}) \varepsilon_i$

Proof of Lemma 8.2. The proof of this lemma relies on the blocking method, specifically drawing upon techniques introduced by [13]. The central idea involves partitioning the strictly stationary sequence (Z_1, \dots, Z_n) into $2n$ blocks, each of length a_n , along with a residual block of length $n - 2v_n a_n$. This approach, known as Bernstein's method and discussed in [22], facilitates the application of symmetrization and various techniques designed for i.i.d. random variables. To establish the independence between the blocks, the smaller blocks are placed between two consecutive larger blocks, and their contribution should be asymptotically negligible. Next, introduce the sequence of independent blocks (η_1, \dots, η_n) such that:

$$\mathcal{L}(\eta_1, \dots, \eta_n) = \mathcal{L}(Z_1, \dots, Z_{a_n}) \times \mathcal{L}(Z_{a_n+1}, \dots, Z_{2a_n}) \times \cdots$$

An application of the result of [73] implies that for any measurable set A :

$$\begin{aligned} & \left| \mathbb{P}\{\{\eta_1, \dots, \eta_{a_n}, \eta_{2a_n+1}, \dots, \eta_{3a_n}, \dots, \eta_{2(v_n-1)a_n+1}, \dots, \eta_{2v_n a_n}\} \in A\} \right. \\ & \quad \left. - \mathbb{P}\{(Z_1, \dots, Z_{a_n}, Z_{2a_n+1}, \dots, Z_{3a_n}, \dots, Z_{2(v_n-1)a_n+1}, \dots, Z_{2v_n a_n}) \in A\} \right| \\ & \leq 2(v_n - 1)\beta(a_n). \end{aligned} \quad (8.65)$$

Since we are working with a locally stationary sequence (X_1, \dots, X_n) , the sequence of independent blocks used subsequently is denoted by $\{\eta_i\}_{i \in \mathbb{N}^*}$. We decompose the process based on the distribution of these blocks:

$$\begin{aligned} & \sum_{i_1 \neq i_2}^n \frac{1}{h^2 \phi^2(h_n)} \prod_{k=1}^2 \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \mathfrak{B}_{i_1, i_2, \varphi, n} \\ & = \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathcal{H}_p^{(U)}} \sum_{i_2 \in \mathcal{H}_q^{(U)}} \frac{\mathfrak{B}_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h_n)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \\ & \quad + \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in \mathcal{H}_p^{(U)}} \frac{\mathfrak{B}_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h_n)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \\ & \quad + 2 \sum_{p=1}^{v_n} \sum_{i_1 \in \mathcal{H}_p^{(U)}} \sum_{q: |q-p| \geq 2}^{v_n} \sum_{i_2 \in \mathcal{T}_q^{(U)}} \frac{\mathfrak{B}_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h_n)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \\ & \quad + 2 \sum_{p=1}^{v_n} \sum_{i_1 \in \mathcal{H}_p^{(U)}} \sum_{q: |q-p| \leq 1}^{v_n} \sum_{i_2 \in \mathcal{T}_q^{(U)}} \frac{\mathfrak{B}_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h_n)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n} \right) \right\} \end{aligned}$$

$$\begin{aligned}
& + \sum_{p \neq q} \sum_{i_1 \in \mathbb{T}_p^{(U)}} \sum_{i_2 \in \mathbb{T}_q^{(U)}} \frac{\mathfrak{W}_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h_n)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\} \\
& + \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2} \sum_{i_1, i_2 \in \mathbb{T}_p^{(U)}} \frac{\mathfrak{W}_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h_n)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right\} \\
& := \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI}.
\end{aligned} \tag{8.66}$$

(I): The same type of block but not the same block: Assume that the sequence of independent blocks $\{\eta_i\}_{i \in \mathbb{N}^*}$ is of size a_n . An application of (8.65) shows that:

$$\begin{aligned}
& \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
& \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right] \mathfrak{W}_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
& + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \left[\mathfrak{W}_{i_1, i_2, \varphi, n} - \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right] \right] > \delta \right) \\
& + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) \\
& \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{S}_{\Theta}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\
& \quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) + 2\nu_n \beta(b_n) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1).
\end{aligned}$$

By the fact that

$$\begin{aligned}
& \mathbb{E} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \left[\prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right] \mathfrak{W}_{i_1, i_2, \varphi, n} \right| \\
& = n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \left| \mathbb{E} \left[\prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right] \mathfrak{W}_{i_1, i_2, \varphi, n} \right|
\end{aligned}$$

$$\begin{aligned}
&= n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \\
&\quad \mathbb{E} \left[\left| \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right| \sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) \varepsilon_{\mathbf{i}} \right] \\
&= n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) \mathbb{E} \left[\left| \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) \right. \right. \\
&\quad \left. \left. - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right| \left| \sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) - \sigma(\mathbf{u}, X_{i, n}) + \sigma(\mathbf{u}, X_{i, n}) \right| \right] \\
&\lesssim n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (\sigma(\mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}(1)) \\
&\quad \mathbb{E} \left[\left| \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \right| \right] \\
&\lesssim n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (\sigma(\mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}(1)) \left[\frac{m \phi^{m-1}(h_n)}{n h_n} \right]
\end{aligned}$$

(where $m = 2$ and using Lemma 8.1 Equation (8.58))

$\sim o_{\mathbb{P}}(1)$,

and

$$\begin{aligned}
&\mathbb{E} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) [\mathfrak{W}_{i_1, i_2, \varphi, n} - \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)}] \right| \\
&= n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E} \left[\left| \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \left[\sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) \varepsilon_{\mathbf{i}} - \sigma(\mathbf{u}, \mathbf{x}) \varepsilon_{\mathbf{i}} \right] \right| \right] \\
&= n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) \mathbb{E} \left[\left| \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n}^{(i/n)})}{h_n} \right) \left[\sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) - \sigma(\mathbf{u}, \mathbf{x}) \right] \right| \right] \\
&\lesssim n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (o_{\mathbb{P}}(1)) \int_0^h \prod_{k=1}^m K_2 \left(\frac{y_k}{h_n} \right) dF_{i_k/n}(y_k, x_k) \\
&\lesssim n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (o_{\mathbb{P}}(1)) (\phi^2(h_n)) \\
&\sim o_{\mathbb{P}}(1). \tag{8.67}
\end{aligned}$$

We keep the choice of b_n and v_n such that

$$v_n b_n^r \leq 1, \tag{8.68}$$

which implies that $2\nu_n\alpha_{b_n} \rightarrow 0$ as $n \rightarrow \infty$, so the term to consider is the second summand. For the second part of the inequality, we turn to the work of [11] in the non-fixed kernel setting. Specifically, we define

$$f_{i_1, \dots, i_m} = \prod_{k=1}^m \xi_{i_k} \times H,$$

and $\mathcal{F}_{i_1, \dots, i_m}$ represents a collection of kernels and the corresponding class of functions associated with this kernel. Subsequently, we will apply [64, Theorem 3.1.1 and Remarks 3.5.4 part 2] for decoupling and randomization. Given our assumption that $m = 2$, we can observe that:

$$\begin{aligned} & \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, \eta_{i_k})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ &= \mathbb{E} \left\| n^{-3/2} h \phi(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \\ &\leq c_2 \mathbb{E} \left\| n^{-3/2} h \phi(h_n) \sum_{p \neq q} \epsilon_p \epsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \\ &\leq c_2 \mathbb{E} \int_0^{D_{nh}^{(U_1)}} N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(1)}) dt, \quad (\text{by Lemma A.5 and Proposition A.6}) \end{aligned} \tag{8.69}$$

where $D_{nh}^{(U_1)}$ is the diameter of \mathcal{F}_{i_1, i_2} according to the distance $\tilde{d}_{nh, 2}^{(1)}$, respectively defined as

$$\begin{aligned} D_{nh}^{(U_1)} &:= \left\| \mathbb{E}_\epsilon \left\| n^{-3/2} h \phi(h_n) \sum_{p \neq q} \epsilon_p \epsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \right\| \\ &= \left\| \mathbb{E}_\epsilon \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \epsilon_p \epsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, \eta_{i_k})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \right\|, \end{aligned}$$

and :

$$\begin{aligned} & \tilde{d}_{nh, 2}^{(1)} \left(\xi_1, K_{2,1} \mathfrak{W}^{(u)}, \xi_2, K_{2,2} \mathfrak{W}''^{(u)} \right) \\ &:= \mathbb{E}_\epsilon \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \epsilon_p \epsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \left[\xi_{1i_1} \xi_{1i_2} \prod_{k=1}^2 K_{1,2} \left(\frac{d_{\theta_k}(x_k, \eta_{i_k})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right. \right. \\ & \quad \left. \left. - \xi_{2i_1} \xi_{2i_2} \prod_{k=1}^2 K_{2,2} \left(\frac{d_{\theta_k}(x_k, \eta_{i_k})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}''^{(u)} \right] \right\|. \end{aligned}$$

Let's consider another semi-norm $\tilde{d}_{nh, 2}^{(2)}$:

$$\tilde{d}_{nh, 2}^{(2)} \left(\xi_1, K_{2,1} \mathfrak{W}^{(u)}, \xi_2, K_{2,2} \mathfrak{W}''^{(u)} \right) = \frac{1}{nh^2 \phi^2(h_n)} \left[\sum_{i \neq j} \left(\xi_{1i_1} \xi_{1i_2} \prod_{k=1}^2 K_{1,2} \left(\frac{d_{\theta_k}(x_k, \eta_{i_k})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right. \right.$$

$$- \xi_{2i_1} \xi_{2i_2} \prod_{k=1}^2 K_{2,2} \left(\frac{d_{\theta_k}(x_k, \eta_{ik})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \Big)^2 \Big]^{1/2}.$$

One can see that

$$\widetilde{d}_{nh,2}^{(1)}(\xi_1.K_{2,1}\mathfrak{W}^{(u)}, \xi_2.K_{2,2}\mathfrak{W}^{(u)}) \leq a_n n^{-1/2} h \phi(h_n) \widetilde{d}_{nh,2}^{(2)}(\xi_1.K_{2,1}\mathfrak{W}^{(u)}, \xi_2.K_{2,2}\mathfrak{W}^{(u)}).$$

We readily infer that

$$\begin{aligned} & \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, \eta_{ik})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ & \leq c_2 \mathbb{E} \int_0^{D_{nh}^{(U_1)}} N(t a_n^{-1} n^{1/2}, \mathcal{F}_{i,j}, \widetilde{d}_{nh,2}^{(2)}) dt \\ & \leq c_2 a_n n^{-1/2} \mathbb{P} \{ D_{nh}^{(U_1)} a_n^{-1} n^{1/2} \geq \lambda_n \} + c_m a_n n^{-1/2} \int_0^{\lambda_n} \log t^{-1} dt, \end{aligned} \tag{8.70}$$

where $\lambda_n \rightarrow 0$. We have

$$\frac{\left(\int_0^{\lambda_n} \log t^{-1} dt \right)}{\left(\lambda_n \log \lambda_n^{-1} \right)} \rightarrow 0,$$

where a_n and λ_n must be chosen in such a way that the following relation will be achieved

$$a_n \lambda_n n^{-1/2} \log \lambda_n^{-1} \rightarrow 0. \tag{8.71}$$

Utilizing the triangle inequality along with Hoeffding’s trick, we readily obtain that

$$\begin{aligned} & a_n n^{-1/2} \mathbb{P} \{ D_{nh}^{(U_1)} \geq \lambda_n a_n n^{-1/2} \} \\ & \leq \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h_n) \mathbb{E} \left\| \sum_{p \neq q} \left[\sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ & \leq c_2 v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h_n) \mathbb{E} \left\| \sum_{p=1}^{v_n} \left[\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{H}^2}, \end{aligned} \tag{8.72}$$

where $\{\eta'_i\}_{i \in \mathbb{N}^*}$ are independent copies of $(\eta_i)_{i \in \mathbb{N}^*}$. By imposing:

$$\lambda_n^{-2} a_n^{1-r} n^{-1/2} \rightarrow 0, \tag{8.73}$$

we readily infer that

$$\left\| v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h_n) \mathbb{E} \sum_{p=1}^{v_n} \left[\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, \eta_{ik})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{H}^2} \leq O(\lambda_n^{-2} a_n^{1-r} n^{-1/2}).$$

Symmetrizing the last inequality in (8.72) and subsequently applying Proposition A.6 from the Appendix yields

$$\begin{aligned} & \nu_n \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h_n) \mathbb{E} \left\| \sum_{p=1}^{\nu_n} \left[\sum_{i_1, i_2 \in H_p^{(U)}} \epsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ & \leq c_2 \mathbb{E} \left(\int_0^{D_{nh}^{(U_2)}} (\log N(u, \mathcal{F}_{i,j}, \tilde{d}'_{nh,2}))^{1/2} \right), \end{aligned} \tag{8.74}$$

where

$$\begin{aligned} D_{nh}^{(U_2)} &= \left\| \mathbb{E}_\epsilon \left[\nu_n \lambda_n^{-2} a_n^{-1} n^{-5/2} \phi^{-1}(h_n) \right. \right. \\ & \quad \left. \left. \sum_{p=1}^{\nu_n} \epsilon_p \left[\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right] \right\|_{\mathcal{F}_2 \mathcal{H}^2}. \end{aligned}$$

and for $\xi_{1, K_{2,1}} \mathfrak{W}'$, $\xi_{2, K_{2,2}} \mathfrak{W}'' \in \mathcal{F}_{ij}$:

$$\begin{aligned} & \tilde{d}'_{nh,2} (\xi_{1, K_{2,1}} \mathfrak{W}'^{(u)}, \xi_{2, K_{2,2}} \mathfrak{W}''^{(u)}) \\ & := \mathbb{E}_\epsilon \left[\nu_n \lambda_n^{-2} a_n^{-1} n^{-5/2} \phi^{-1}(h_n) \sum_{p=1}^{\nu_n} \epsilon_p \left[\left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{1i_1} \xi_{1i_2} K_{2,1} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,1} \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}'_{i_1, i_2, \varphi, n} \right)^2 \right. \right. \\ & \quad \left. \left. - \left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,2} \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}''_{i_1, i_2, \varphi, n} \right)^2 \right] \right]. \end{aligned}$$

By the fact that :

$$\begin{aligned} & \mathbb{E}_\epsilon \left[\nu_n \lambda_n^{-2} a_n^{-1} n^{-5/2} \phi^{-1}(h_n) \sum_{p=1}^{\nu_n} \epsilon_p \left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right] \\ & \leq a_n^{3/2} \lambda_n^{-2} n^{-1} \left[\nu_n^{-1} a_n^{-2} \phi^{-2}(h_n) \sum_{p=1}^{\nu_n} \sum_{i_1, i_2 \in H_p^{(U)}} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta'_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right)^4 \right]^{1/2}, \end{aligned}$$

so:

$$a_n^{3/2} \lambda_n^{-2} n^{-1} \rightarrow 0, \tag{8.75}$$

we have the convergence of (8.74) to zero. For the choice of a_n , b_n and ν_n , it should be noted that all the values satisfying (8.54), (8.68), (8.71), (8.73), and (8.74) are suitable.

(II): The same block:

$$\mathbb{P} \left(\sup_{\mathcal{F}_m \mathfrak{R}_\Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi(h_n) \sum_{p=1}^{\nu_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \frac{1}{\phi^2(h_n)} \xi_{i_1} \xi_{i_2} \right. \right.$$

$$\begin{aligned}
& K_2\left(\frac{d_{\theta_1}(x_1, X_{i_1,n})}{h_n}\right) K_2\left(\frac{d_{\theta_2}(x_2, X_{i_2,n})}{h_n}\right) \mathfrak{B}_{i_1, i_2, \varphi, n} \Big| > \delta \Big) \\
\leq & \mathbb{P}\left(\sup_{\mathcal{F}_m, \mathcal{S}_\emptyset^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left\| n^{-3/2} h \phi(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \frac{1}{\phi^2(h_n)} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \left. \left. \left[\prod_{k=1}^2 K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n})}{h_n}\right) - \prod_{k=1}^2 K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i/n)})}{h_n}\right) \right] \mathfrak{B}_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
+ & \mathbb{P}\left(\sup_{\mathcal{F}_m, \mathcal{S}_\emptyset^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \left. \left. \prod_{k=1}^2 K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i/n)})}{h_n}\right) \left[\mathfrak{B}_{i_1, i_2, \varphi, n} - \mathfrak{B}_{i_1, i_2, \varphi, n}^{(u)} \right] \right| > \delta \right) \\
+ & \mathbb{P}\left(\sup_{\mathcal{F}_m, \mathcal{S}_\emptyset^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \left. \left. \prod_{k=1}^2 K_2\left(\frac{d_{\theta_k}(x_k, X_{i_k,n}^{(i/n)})}{h_n}\right) \mathfrak{B}_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) \\
\leq & \mathbb{P}\left(\sup_{\mathcal{F}_m, \mathcal{S}_\emptyset^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2\left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n}\right) \right. \right. \\
& \left. \left. K_2\left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n}\right) \mathfrak{B}_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) + 2v_n \alpha_{b_n}. \tag{8.76}
\end{aligned}$$

Similar to I, we can show that both the first and second terms in the previous inequality are of order $o_{\mathbb{P}}(1)$. Therefore, as in the preceding proof, it is enough to establish

$$\mathbb{E}\left(\left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2\left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n}\right) K_2\left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n}\right) \mathfrak{B}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \right) \rightarrow 0.$$

Notice that when we consider a uniformly bounded class of functions, we obtain uniformity in $B^m \times \mathcal{F}_2 \mathcal{K}^2$

$$\mathbb{E}\left(\sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2\left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n}\right) K_2\left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n}\right) \mathfrak{B}_{i_1, i_2, \varphi, n}\right) = O(a_n).$$

This implies that we have to prove that, for $\mathbf{u} \in B^m$

$$\mathbb{E}\left(\left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \left[\xi_{i_1} \xi_{i_2} K_2\left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n}\right) K_2\left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n}\right) \mathfrak{B}_{i_1, i_2, \varphi, n}^{(u)} \right] \right\| \right)$$

$$- \mathbb{E} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n} \right) \Big\|_{\mathcal{F}_2, \mathcal{K}^2} \Big\| \rightarrow 0.$$

As for empirical processes, to prove (8.77), it suffices to symmetrize and show that

$$\mathbb{E} \left(\left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \epsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \right) \rightarrow 0.$$

In a similar way as in (8.69), we infer that :

$$\begin{aligned} & \mathbb{E} \left(\left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \epsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \right) \\ & \leq \mathbb{E} \left(\int_0^{D_{nh}^{(U_3)}} (\log N(u, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh,2}^{(3)}))^{1/2} du \right), \end{aligned}$$

where

$$\begin{aligned} D_{nh}^{(U_3)} &= \left\| \mathbb{E}_\epsilon \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \epsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \right\|, \end{aligned} \quad (8.77)$$

and the semi-metric $\tilde{d}_{nh,2}^{(3)}$ is defined by

$$\begin{aligned} & \tilde{d}_{nh,2}^{(3)} \left(\xi_{1 \cdot} K_{2,1} \mathfrak{W}^{(u)}, \xi_{2 \cdot} K_{2,2} \mathfrak{W}^{(u)} \right) \\ &= \mathbb{E}_\epsilon \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \left(\xi_{1i_1} \xi_{1i_2} K_{2,1} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,1} \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \right. \right. \\ & \quad \left. \left. \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} - \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,2} \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right) \right|. \end{aligned}$$

Since we are trading uniformly bounded classes of functions, we infer that

$$\begin{aligned} & \mathbb{E}_\epsilon \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right| \\ & \leq a_n^{3/2} (n)^{-1} h \phi^{-1}(h_n) \left[\frac{1}{v_n a_n^2} \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right]^{1/2} \leq O \left(a_n^{3/2} (n)^{-1} \phi^{-1}(h_n) \right). \end{aligned}$$

Since $a_n^{3/2} (n)^{-1} \phi^{-1}(h_n) \rightarrow 0$, $D_{nh}^{(U_3)} \rightarrow 0$, we obtain $\text{II} \rightarrow 0$ as $n \rightarrow \infty$.

(III): Different types of blocks:

$$\begin{aligned}
& \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \frac{1}{\phi^2(h_n)} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) \right. \right. \\
& \quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
& \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \frac{1}{\phi^2(h_n)} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i_k, n})}{h_n} \right) - \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i, n}^{(i/n)})}{h_n} \right) \right] \mathfrak{W}_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
& \quad + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \quad \left. \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i, n}^{(i/n)})}{h_n} \right) \left[\mathfrak{W}_{i_1, i_2, \varphi, n} - \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right] \right| > \delta \right) \\
& \quad + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
& \quad \left. \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i, n}^{(i/n)})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right).
\end{aligned}$$

As mentioned earlier, we have addressed the first and second summands in the previous inequality. What remains is the last summation, where the application of (8.65) reveals that

$$\begin{aligned}
& \sum_{p=1}^{v_n} \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d_{\theta_k}(x_k, X_{i, n}^{(i/n)})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\
& \leq \sum_{p=1}^{v_n} \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \\
& \quad \left. K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} + n^{-3/2} h \phi^{-1}(h_n) v_n^2 a_n b_n \beta(a_n),
\end{aligned}$$

we have

$$n^{-3/2} \phi^{-1}(h_n) v_n^2 a_n b_n \beta(a_n) \rightarrow 0,$$

using Condition (8.63) and the choice of a_n , b_n and v_n . For $p = 1$ and $p = v_n$:

$$\mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2}$$

$$= \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2}.$$

For $2 \leq p \leq v_n - 1$, we obtain :

$$\begin{aligned} & \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &= \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=4}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &\leq \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2}, \end{aligned}$$

therefore it suffices to treat the convergence:

$$\mathbb{E} \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \rightarrow 0. \tag{8.78}$$

Using similar arguments as in [13], we apply the standard symmetrization and:

$$\begin{aligned} & \mathbb{E} \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &\leq 2 \mathbb{E} \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \epsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &= 2 \mathbb{E} \left\{ \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \epsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ &\quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \mathbb{1}_{\{D_{nh}^{(U_4)} \leq \gamma_n\}} \right\} \\ &\quad + 2 \mathbb{E} \left\{ \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \epsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ &\quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \mathbb{1}_{\{D_{nh}^{(U_4)} > \gamma_n\}} \right\} \\ &= 2\text{III}_1 + 2\text{III}_2, \end{aligned} \tag{8.79}$$

where

$$D_{nh}^{(U_4)} = \left\| \nu_n n^{-3/2} h \phi^{-1}(h_n) \left[\sum_{q=3}^{\nu_n} \left(\sum_{i_2 \in T_q^{(U)}} \sum_{i_1 \in H_1^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right]^{1/2} \right\|_{\mathcal{F}_2, \mathcal{K}^2}. \quad (8.80)$$

In a similar way as in (8.69), we infer that

$$\text{III}_1 \leq c_2 \int_0^{\gamma_n} (\log N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh,2}^{(4)}))^{1/2} dt, \quad (8.81)$$

where

$$\begin{aligned} \tilde{d}_{nh,2}^{(4)}(\xi_1, K_{2,1} \mathfrak{W}^{(u)}, \xi_2, K_{2,2} \mathfrak{W}''^{(u)}) &:= \mathbb{E}_\epsilon \left| \nu_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{\nu_n} \sum_{i_2 \in T_q^{(U)}} \epsilon_q \left[\xi_{i_1} \xi_{i_2} K_{2,1} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ &\times K_{2,1} \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} - \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,2} \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}''^{(u)} \left. \right] \Big|. \end{aligned}$$

Since we have

$$\begin{aligned} &\mathbb{E}_\epsilon \left| \nu_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{\nu_n} \sum_{i_2 \in T_q^{(U)}} \epsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right| \\ &\leq a_n^{-1/2} b_n h^2 \phi(h_n) \left(\frac{1}{a_n b_n \nu_n h^2 \phi^4(h_n)} \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{\nu_n} \sum_{i_2 \in T_q^{(U)}} \left[\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ &\left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right)^{1/2}, \end{aligned}$$

and considering the semi-metric

$$\begin{aligned} &\tilde{d}_{nh,2}^{(5)}(\xi_1, K_{2,1} \mathfrak{W}^{(u)}, \xi_2, K_{2,2} \mathfrak{W}''^{(u)}) \\ &:= \left(\frac{1}{a_n b_n \nu_n h^2 \phi^4(h_n)} \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{\nu_n} \sum_{i_2 \in T_q^{(U)}} \left[\xi_{i_1} \xi_{i_2} K_{2,1} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,1} \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right. \right. \\ &\left. \left. - \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_{2,2} \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}''^{(u)} \right]^2 \right)^{1/2}. \end{aligned}$$

We show that the expression in (8.81) is bounded as follows

$$\nu_n^{1/2} b_n n^{-1/2} h^2 \phi(h_n) \int_0^{\nu_n^{-1/2} b_n^{-1} n^{1/2} h^2 \gamma_n} (\log N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh,2}^{(5)}))^{1/2} dt,$$

by choosing $\gamma_n = n^{-\alpha}$ for some $\alpha > (17r - 26)/60r$, we get the convergence to zero of the previous quantity. To bound the second term on the right-hand side of (8.79), we observe that

$$\begin{aligned} \text{III}_2 &= \mathbb{E} \left\{ \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \epsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) \right. \right. \\ &\quad \left. \left. K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{H}^2} \mathbb{1}_{\{D_{nh}^{(U_4)} > \gamma_n\}} \right\} \\ &\leq a_n^{-1} b_n n^{1/2} h \phi^{-1}(h_n) \mathbb{P} \left\{ \left\| v_n^2 n^{-3} h^2 \phi^{-2}(h_n) \sum_{q=3}^{v_n} \left(\sum_{i_2 \in T_q^{(U)}} \sum_{i_1 \in H_1^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \right. \\ &\quad \left. \left. \left. K_2 \left(\frac{d_{\theta_1}(x_1, \eta_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, \eta_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right\|_{\mathcal{F}_2, \mathcal{H}^2} \geq \gamma_n^2 \right\}. \end{aligned} \tag{8.82}$$

Now, we apply the square root trick to the last expression conditional on H_1^U . Denoting \mathbb{E}_T as the expectation with respect to $\sigma\{\eta_{i_2} : i_2 \in T_q, q \geq 3\}$, we assume that any class of functions \mathcal{F}_m is unbounded, and its envelope function satisfies, for some $p > 2$:

$$\theta_p := \sup_{\mathbf{t} \in \mathcal{S}^m} \mathbb{E}(F^p(\mathbf{Y}) | \mathbf{X} = \mathbf{t}) < \infty, \tag{8.83}$$

for $2r/(r - 1) < s < \infty$, (in the notation in of [89, Lemma 5.2]).

$$M_n = v_n^{1/2} \mathbb{E}_T \left(\sum_{j \in T_q^{(U)}} \sum_{i \in H_1^{(U)}} \xi_i \xi_j K_2 \left(\frac{d_{\theta_1}(x_1, X_i)}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_j)}{h_n} \right) \mathfrak{W}_{i, j, \varphi, n}^{(u)} \right)^2,$$

where

$$t = \gamma_n^2 a_n^{5/2} n^{1/2} h \phi^{-1}(h_n), \quad \rho = \lambda = 2^{-4} \gamma_n a_n^{5/4} n^{1/4} h^{1/2} \phi^{-1/2}(h_n),$$

and

$$m = \exp(\gamma_n^2 n h^2 \phi^{-2}(h_n) b_n^{-2}).$$

Nevertheless, since we require $t > 8M_n$ and $m \rightarrow \infty$, using similar arguments as in [13, page 69], we achieve the convergence of (8.81) and (8.82) to zero.

(IV): Different types of blocks: The target here is to prove that:

$$\mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{S}_\Theta^m} \sup_{\theta \in \Theta^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n} \right| > \delta \right) \rightarrow 0.$$

We have

$$\begin{aligned} &\left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d_{\theta_1}(x_1, X_{i_1}^{(i_1/n)})}{h_n} \right) K_2 \left(\frac{d_{\theta_2}(x_2, X_{i_2}^{(i_2/n)})}{h_n} \right) \mathfrak{W}_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{H}^2} \\ &\leq c_2 v_n a_n b_n n^{-3/2} h \phi^{-1}(h_n) \rightarrow 0. \end{aligned}$$

Hence the proof of the lemma is complete. □

Use of AI tools declaration

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

Acknowledgements

The authors would like to thank the Editor-in-Chief, an Associate-Editor, and three referees for their extremely helpful remarks, which resulted in a substantial improvement of the original form of the work and a presentation that was more sharply focused.

Conflict of interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. J. Abrevaya, W. Jiang, A nonparametric approach to measuring and testing curvature, *J. Bus. Econom. Statist.*, **23** (2005), 1–19. <https://doi.org/10.1198/073500104000000316>
2. 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>
3. I. M. Almanjahie, S. Bouzebda, Z. Chikr Elmezouar, A. Laksaci, The functional k NN estimator of the conditional expectile: uniform consistency in number of neighbors, *Stat. Risk Model.*, **38** (2022a), 47–63. <https://doi.org/10.1515/strm-2019-0029>
4. 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>
5. I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, The local linear functional k NN estimator of the conditional expectile: uniform consistency in number of neighbors, *Metrika*, **34** (2024), 1–29. <https://doi.org/10.1007/s00184-023-00942-0>
6. N. T. Andersen, The central limit theorem for non-separable valued functions, *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **70** (1985), 445–455.
7. P. K. Andersen, O. R. Borgan, R. D. Gill, N. Keiding, *Statistical Models Based on Counting Processes*, New York: Springer, 1993.
8. G. Aneiros, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, *J. Multivariate Anal.*, **170** (2019), 3–9. <https://doi.org/10.1016/j.jmva.2018.11.007>
9. A. Araujo, E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables*, New York: John Wiley & Sons, 1980.
10. M. A. Arcones, The law of large numbers for U -statistics under absolute regularity, *Electron. Comm. Probab.*, **3** (1998), 13–19.

11. M. A. Arcones, E. Giné, Limit theorems for U -processes, *Ann. Probab.*, **21** (1993), 1494–1542.
12. M. A. Arcones, E. Giné, On the law of the iterated logarithm for canonical U -statistics and processes, *Stochast. Process. Appl.*, **58** (1995), 217–245. [https://doi.org/10.1016/0304-4149\(94\)00023-M](https://doi.org/10.1016/0304-4149(94)00023-M)
13. M. A. Arcones, B. Yu, Central limit theorems for empirical and U -processes of stationary mixing sequences, *J. Theor. Probab.*, **7** (1994), 47–71. <https://doi.org/10.1007/BF02213360>
14. M. A. Arcones, Z. Chen, E. Giné, Estimators related to U -processes with applications to multivariate medians: asymptotic normality, *Ann. Statist.*, **22** (1994), 1460–1477.
15. 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>
16. S. Attaoui, B. Bentat, 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 Math.*, **9** (2024), 7340–7371. <http://dx.doi.org/10.3934/math.2024356>
17. A. K. Basu, A. Kundu, Limit distribution for conditional U -statistics for dependent processes, *Calcutta Statist. Assoc. Bull.*, **52** (2002), 381–407. <https://doi.org/10.1177/0008068320020522>
18. A. Bellet, A. Habrard, Robustness and generalization for metric learning, *Neurocomputing*, **151** (2015), 259–267. <https://doi.org/10.1016/j.neucom.2014.09.044>
19. A. Bellet, A. Habrard, M. Sebban, A survey on metric learning for feature vectors and structured data, preprint paper, 2013. <https://doi.org/10.48550/arXiv.1306.6709>
20. 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>
21. W. Bergsma, A. Dassios, A consistent test of independence based on a sign covariance related to Kendall's tau, *Bernoulli*, **20** (2014), 1006–1028.
22. S. Bernstein, Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes, *Math. Ann.*, **97** (1927), 1–59. <https://doi.org/10.1007/BF01447859>
23. S. Bhattacharjee, H. G. Müller, Single index Fréchet regression, *Ann. Statist.*, **51** (2023), 1770–1798. <https://doi.org/10.1214/23-AOS2307>
24. J. R. Blum, J. Kiefer, M. Rosenblatt, Distribution free tests of independence based on the sample distribution function, *Ann. Math. Statist.*, **32** (1961), 485–498.
25. V. I. Bogachev, Gaussian measures, In: *Mathematical Surveys and Monographs*, Providence: American Mathematical Society, 1998.
26. E. G. Bongiorno, A. Goia, Classification methods for Hilbert data based on surrogate density, *Comput. Statist. Data Anal.*, **99** (2016), 204–222. <https://doi.org/10.1016/j.csda.2016.01.019>
27. E. G. Bongiorno, A. Goia, Some insights about the small-ball probability factorization for Hilbert random elements, *Statist. Sinica*, **27** (2017), 1949–1965.
28. E. G. Bongiorno, A. Goia, P. Vieu, Evaluating the complexity of some families of functional data, *SORT*, **42** (2018), 27–44.

29. S. Borovkova, R. Burton, H. Dehling, Consistency of the Takens estimator for the correlation dimension, *Ann. Appl. Probab.*, **9** (1999), 376–390.
30. S. Borovkova, R. Burton, H. Dehling, Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation, *Trans. Amer. Math. Soc.*, **353** (2001), 4261–4318.
31. Y. V. Borovskikh, *U-Statistics in Banach Spaces*, Utrecht: VSP, 1996.
32. D. Bosq, Linear processes in function spaces, In: *Lecture Notes in Statistics*, New York: Springer, 2000.
33. 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>
34. S. Bouzebda, On the weak convergence and the uniform-in-bandwidth consistency of the general conditional U -processes based on the copula representation: multivariate setting, *Hacet. J. Math. Stat.*, **52** (2023), 1303–1348.
35. S. Bouzebda, M. Cherfi, General bootstrap for dual ϕ -divergence estimates, *J. Probab. Stat.*, **2012** (2012), 834107. <https://doi.org/10.1155/2012/834107>
36. 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>
37. 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>
38. S. Bouzebda, A. Keziou, A semiparametric maximum likelihood ratio test for the change point in copula models, *Stat. Methodol.*, **14** (2013), 39–61. <https://doi.org/10.1016/j.stamet.2013.02.003>
39. S. Bouzebda, B. Nemouchi, Central limit theorems for conditional empirical and conditional U -processes of stationary mixing sequences, *Math. Meth. Stat.*, **28** (2019), 169–207. <https://doi.org/10.3103/S1066530719030013>
40. 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>
41. 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 Math.*, **9** (2024), 4427–4550. <https://doi.org/10.3934/math.2024218>
42. S. Bouzebda, Soukarieh, Non-parametric conditional U -processes for locally stationary functional random fields under stochastic sampling design, *Mathematics*, **11** (2023), 16. <https://doi.org/10.3390/math11010016>
43. S. Bouzebda, I. Soukarieh, Limit theorems for a class of processes generalizing the U -empirical process, *Stochastics*, **96** (2024), 799–845. <https://doi.org/10.1080/17442508.2024.2320402>
44. S. Bouzebda, N. Taachouche, On the variable bandwidth kernel estimation of conditional U -statistics at optimal rates in sup-norm, *Phys. A Stat. Mechan. Appl.*, **625** (2023), 129000. <https://doi.org/10.1016/j.physa.2023.129000>

45. S. Bouzebda, N. Taachouche, Rates of the strong uniform consistency for the kernel-type regression function estimators with general kernels on manifolds, *Math. Meth. Stat.*, **32** (2023), 27–80. <https://doi.org/10.3103/S1066530723010027>
46. S. Bouzebda, N. Taachouche, Rates of the strong uniform consistency with rates for conditional U -statistics estimators with general kernels on manifolds, *Math. Meth. Stat.*, in press, 2023.
47. S. Bouzebda, S. Didi, L. El Hajj, Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: asymptotic results, *Math. Meth. Stat.*, **24** (2015), 163–199. <https://doi.org/10.3103/S1066530715030011>
48. S. Bouzebda, I. Elhattab, B. Nemouchi, On the uniform-in-bandwidth consistency of the general conditional U -statistics based on the copula representation, *J. Nonparametr. Stat.*, **33** (2021), 321–358. <https://doi.org/10.1080/10485252.2021.1937621>
49. 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>
50. Q. Cao, Z. C. Guo, Y. Ying, Generalization bounds for metric and similarity learning, *Mach. Learn.*, **102** (2016), 115–132. <https://doi.org/10.1007/s10994-015-5499-7>
51. A. Carbonez, L. Györfi, E. C. van der Meulen, Partitioning-estimates of a regression function under random censoring, *Stat. Risk Model.*, **13** (1995), 21–37. <https://doi.org/10.1524/strm.1995.13.1.21>
52. D. Chen, P. Hall, H. G. Müller, Single and multiple index functional regression models with nonparametric link, *Ann. Statist.*, **39** (2011), 1720–1747. <https://doi.org/10.1214/11-AOS882>
53. Y. Chen, S. Datta, Adjustments of multi-sample U -statistics to right censored data and confounding covariates, *Comput. Statist. Data Anal.*, **135** (2019), 1–14. <https://doi.org/10.1016/j.csda.2019.01.012>
54. S. Cléménçon, I. Colin, A. Bellet, Scaling-up empirical risk minimization: optimization of incomplete U -statistics, *J. Mach. Learn. Res.*, **17** (2016), 1–36.
55. G. B. Cybis, M. Valk, S. R. C. Lopes, Clustering and classification problems in genetics through U -statistics, *J. Stat. Comput. Simul.*, **88** (2018), 1882–1902. <https://doi.org/10.1080/00949655.2017.1374387>
56. 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)
57. R. Dahlhaus, Fitting time series models to non-stationary processes, *Ann. Statist.*, **25** (1997), 1–37. <https://doi.org/10.1214/aos/1034276620>
58. 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>
59. R. Dahlhaus, W. Polonik, Empirical spectral processes for locally stationary time series, *Bernoulli*, **15** (2009), 1–39. <https://doi.org/10.3150/08-BEJ137>
60. S. Datta, D. Bandyopadhyay, G. A. Satten, Inverse probability of censoring weighted U -statistics for right-censored data with an application to testing hypotheses, *Scand. J. Stat.*, **37** (2010), 680–700. <https://doi.org/10.1111/j.1467-9469.2010.00697.x>

61. J. A. Davydov, Convergence of distributions generated by stationary stochastic processes, *Theory Probab. Appl.*, **13** (1968), 691–696. <https://doi.org/10.1137/1113086>
62. J. A. Davydov, Mixing conditions for Markov chains, *Theory Probab. Appl.*, **18** (1973), 312–328. <https://doi.org/10.1137/1118033>
63. V. H. de la Peña, Decoupling and Khintchine’s inequalities for U -statistics, *Ann. Probab.*, **20** (1992), 1877–1892.
64. V. H. de la Peña, E. Giné, Decoupling, In: *Probability and its Applications*, New York: Springer, 1999.
65. P. Deheuvels, One bootstrap suffices to generate sharp uniform bounds in functional estimation, *Kybernetika*, **47** (2011), 855–865.
66. M. Denker, G. Keller, On U -statistics and v. Mises’ statistics for weakly dependent processes, *Z. Wahrsch. Verw. Gebiete*, **64** (1983), 505–522. <https://doi.org/10.1007/BF00534953>
67. L. Devroye, G. Lugosi, *Combinatorial Methods in Density Estimation*, New York: Springer, 2001.
68. S. Didi, S. Bouzebda, Wavelet density and regression estimators for continuous time functional stationary and ergodic processes, *Mathematics*, **10** (2022), 4356. <https://doi.org/10.3390/math10224356>
69. S. Didi, A. Al Harby, S. Bouzebda, Wavelet density and regression estimators for functional stationary and ergodic data: discrete time, *Mathematics*, **10** (2022), 3433. <https://doi.org/10.3390/math10193433>
70. R. M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, *J. Funct. Anal.*, **1** (1967), 290–330.
71. R. M. Dudley, An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions, In: *Probability in Banach Spaces V. Lecture Notes in Mathematics*, Springer, **1153** (1985), 141–178. <https://doi.org/10.1007/BFb0074949>
72. R. M. Dudley, Uniform central limit theorems, In: *Cambridge Studies in Advanced Mathematics*, 2 Eds., New York: Cambridge University Press, 2014.
73. E. Eberlein, Weak convergence of partial sums of absolutely regular sequences, *Statist. Probab. Lett.*, **2** (1984), 291–293. [https://doi.org/10.1016/0167-7152\(84\)90067-1](https://doi.org/10.1016/0167-7152(84)90067-1)
74. P. P. B. Eggermont, V. N. LaRiccia, *Maximum Penalized Likelihood Estimation*, New York: Springer, 2001.
75. L. Faivishevsky, J. Goldberger, ICA based on a smooth estimation of the differential entropy, In: *Advances in Neural Information Processing Systems*, Inc: Curran Associates, 2008.
76. 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>
77. F. Ferraty, P. Vieu, Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination, In: *The International Conference on Recent Trends and Directions in Nonparametric Statistics*, *J. Nonparametr. Stat.*, **16** (2004), 111–125. <https://doi.org/10.1080/10485250310001622686>
78. F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis*, New York: Springer, 2006.

79. F. Ferraty, A. Peuch, P. Vieu, Modèle à indice fonctionnel simple, *C. R. Math. Acad. Sci. Paris*, **336** (2003), 1025–1028. [https://doi.org/10.1016/S1631-073X\(03\)00239-5](https://doi.org/10.1016/S1631-073X(03)00239-5)
80. F. Ferraty, A. Laksaci, P. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, *Stat. Infer. Stoch. Process.*, **9** (2006), 47–76. <https://doi.org/10.1007/s11203-004-3561-3>
81. F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: inference and practical aspects, *Aust. N.Z. J. Stat.*, **49** (2007), 267–286. <https://doi.org/10.1111/j.1467-842X.2007.00480.x>
82. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Rate of uniform consistency for nonparametric estimates with functional variables, *J. Statist. Plann. Inference*, **140** (2010), 335–352. <https://doi.org/10.1016/j.jspi.2009.07.019>
83. F. Ferraty, N. Kudraszow, P. Vieu, Nonparametric estimation of a surrogate density function in infinite-dimensional spaces, *J. Nonparametr. Stat.*, **24** (2012), 447–464. <https://doi.org/10.1080/10485252.2012.671943>
84. A. Földes, L. Rejtő, A LIL type result for the product limit estimator, *Z. Wahrsch. Verw. Gebiete*, **56** (1981), 75–86. <https://doi.org/10.1007/BF00531975>
85. E. W. Frees, Infinite order U -statistics, *Scand. J. Statist.*, **16** (1989), 29–45.
86. K. A. Fu, An application of U -statistics to nonparametric functional data analysis, *Commun. Stat. Theory Meth.*, **41** (2012), 1532–1542. <https://doi.org/10.1080/03610926.2010.526747>
87. 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>
88. S. Ghosal, A. Sen, A. W. van der Vaart, Testing monotonicity of regression, *Ann. Statist.*, **28** (2000), 1054–1082.
89. E. Giné, J. Zinn, Some limit theorems for empirical processes, *Ann. Probab.*, **12** (1984), 929–998.
90. A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivar. Anal.*, **146** (2016), 1–6. <https://doi.org/10.1016/j.jmva.2015.12.001>
91. L. Gu, L. Yang, Oracally efficient estimation for single-index link function with simultaneous confidence band, *Electron. J. Stat.*, **9** (2015), 1540–1561. <https://doi.org/10.1214/15-EJS1051>
92. P. Hall, Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function, *Z. Wahrsch. Verw. Gebiete*, **67** (1984), 175–196. <https://doi.org/10.1007/BF00535267>
93. P. R. Halmos, The theory of unbiased estimation, *Ann. Math. Statist.*, **17** (1946), 34–43. <https://doi.org/10.1214/aoms/1177731020>
94. F. Han, T. Qian, On inference validity of weighted U -statistics under data heterogeneity, *Electron. J. Statist.*, **12** (2018), 2637–2708. <https://doi.org/10.1214/18-EJS1462>
95. W. Härdle, Applied nonparametric regression, In: *Econometric Society Monographs*, Cambridge: Cambridge University Press, 1990.
96. W. Härdle, J. S. Marron, Optimal bandwidth selection in nonparametric regression function estimation, *Ann. Statist.*, **13** (1985), 1465–1481.

97. M. Harel, M. L. Puri, Conditional U -statistics for dependent random variables, *J. Multivar. Anal.*, **57** (1996), 84–100. <https://doi.org/10.1006/jmva.1996.0023>
98. C. Heilig, D. Nolan, Limit theorems for the infinite-degree U -process, *Statist. Sinica*, **11** (2001), 289–302.
99. W. Hoeffding, A class of statistics with asymptotically normal distribution, *Ann. Math. Stat.*, **19** (1948), 293–325.
100. J. Hoffmann-Jørgensen, Stochastic processes on Polish spaces, In: *Various Publications Series*, Aarhus: Aarhus Universitet, Matematisk Institut, 1991.
101. M. Hollander, F. Proschan, Testing whether new is better than used, *Ann. Math. Statist.*, **43** (1972), 1136–1146. <https://doi.org/10.1214/aoms/1177692466>
102. L. Horváth, P. Kokoszka, *Inference for Functional Data with Applications*, New York: Springer, 2012.
103. I. A. Ibragimov, V. N. Solev, A certain condition for the regularity of Gaussian stationary sequence, *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, **12** (1969), 113–125.
104. S. Jadhav, S. Ma, An association test for functional data based on Kendall's tau, *J. Multivar. Anal.*, **184** (2021), 104740. <https://doi.org/10.1016/j.jmva.2021.104740>
105. C. Jentsch, S. Subba Rao, A test for second order stationarity of a multivariate time series, *J. Econometrics*, **185** (2015), 124–161. <https://doi.org/10.1016/j.jeconom.2014.09.010>
106. 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>
107. R. Jin, S. Wang, Y. Zhou, Regularized distance metric learning: theory and algorithm, In: *Advances in Neural Information Processing Systems*, Inc: Curran Associates, 2009.
108. E. L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.*, **53** (1958), 457–481.
109. L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu, Uniform in bandwidth consistency for various kernel estimators involving functional data, *J. Nonparametr. Stat.*, **29** (2017), 85–107. <https://doi.org/10.1080/10485252.2016.1254780>
110. M. G. Kendall, A new measure of rank correlation, *Biometrika*, **30** (1938), 81–93.
111. M. Kohler, K. Máthé, M. Pintér, Prediction from randomly right censored data, *J. Multivar. Anal.*, **80** (2002), 73–100. <https://doi.org/10.1006/jmva.2000.1973>
112. A. N. Kolmogorov, V. M. Tihomirov, ε -entropy and ε -capacity of sets in function spaces, *Uspehi. Mat. Nauk.*, **14** (1959), 3–86.
113. V. S. Koroljuk, Y. V. Borovskich, Theory of U -statistics, In: *Mathematics and its Applications*, Dordrecht: Kluwer Academic Publishers Group, 1994.
114. M. R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer, 2008.
115. J. P. Kreiss, E. Paparoditis, Bootstrapping locally stationary processes, *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **77** (2015), 267–290. <https://doi.org/10.1111/rssb.12068>

- 116.D. Kurisu, Nonparametric regression for locally stationary functional time series, *Electron. J. Statist.*, **16** (2022), 3973–3995. <https://doi.org/10.1214/22-EJS2041>
- 117.A. J. Lee, *U*-statistics, In: *Statistics: Textbooks and Monographs*, New York: Marcel Dekker, 1990.
- 118.S. Lee, O. Linton, Y. J. Whang, Testing for stochastic monotonicity, *Econometrica*, **77** (2009), 585–602. <https://doi.org/10.3982/ECTA7145>
- 119.A. Leucht, Degenerate *U*- and *V*-statistics under weak dependence: asymptotic theory and bootstrap consistency, *Bernoulli*, **18** (2012), 552–585. <https://doi.org/10.3150/11-BEJ354>
- 120.A. Leucht, M. H. Neumann, Degenerate *U*- and *V*-statistics under ergodicity: asymptotics, bootstrap and applications in statistics, *Ann. Inst. Stat. Math.*, **65** (2013), 349–386. <https://doi.org/10.1007/s10463-012-0374-9>
- 121.J. Li, C. Huang, Z. Hongtu, A functional varying-coefficient single-index model for functional response data, *J. Amer. Stat. Assoc.*, **112** (2017), 1169–1181. <https://doi.org/10.1080/01621459.2016.1195742>
- 122.W. V. Li, Q. M. Shao, Gaussian processes: inequalities, small-ball probabilities and applications, *Handbook Stat.*, **19** (2001), 533–597. [https://doi.org/10.1016/S0169-7161\(01\)19019-X](https://doi.org/10.1016/S0169-7161(01)19019-X)
- 123.H. Liang, X. Liu, R. Li, C. L. Tsai, Estimation and testing for partially linear single-index models, *Ann. Stat.*, **38** (2010), 3811–3836. <https://doi.org/10.1214/10-AOS835>
- 124.E. Liescher, Strong convergence of sums of α -mixing random variables with applications to density estimation, *Stochast. 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)
- 125.F. Lim, V. M. Stojanovic, On *U*-statistics and compressed sensing I: non-asymptotic average-case analysis, *IEEE T. Signal Process.*, **61** (2013), 2473–2485. <https://doi.org/10.1109/TSP.2013.2247598>
- 126.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>
- 127.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. IWFOs 2020. Contributions to Statistics*. Springer, Cham., 2020.
- 128.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>
- 129.Q. Liu, J. Lee, M. Jordan, A kernelized stein discrepancy for goodness-of-fit tests, In: *Proceedings of The 33rd International Conference on Machine Learning, PMLR*, **48** (2016), 276–284.
- 130.B. Maillot, V. Viallon, Uniform limit laws of the logarithm for nonparametric estimators of the regression function in presence of censored data, *Math. Meth. Stat.*, **18** (2009), 159–184. <https://doi.org/10.3103/S1066530709020045>
- 131.T. Masak, S. Sarkar, V. M. Panaretos, Principal separable component analysis via the partial inner product, *Stat. Theory*, 2020.

- 132.D. M. Mason, Proving consistency of non-standard kernel estimators, *Stat. Inference Stoch. Process.*, **15** (2012), 151–176. <https://doi.org/10.1007/s11203-012-9068-4>
- 133.E. Masry, Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stochast. Process. Appl.*, **115** (2005), 155–177. <https://doi.org/10.1016/j.spa.2004.07.006>
- 134.U. Mayer, H. Zähle, Z. Zhou, Functional weak limit theorem for a local empirical process of non-stationary time series and its application, *Bernoulli*, **26** (2020), 1891–1911. <https://doi.org/10.3150/19-BEJ1174>
- 135.E. Mayer-Wolf, O. Zeitouni, The probability of small Gaussian ellipsoids and associated conditional moments, *Ann. Probab.*, **21** (1993), 14–24.
- 136.M. Mohammedi, S. Bouzebda, A. Laksaci, The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data, *J. Multivar. Anal.*, **181** (2021), 104673. <https://doi.org/10.1016/j.jmva.2020.104673>
- 137.M. Mohammedi, S. Bouzebda, A. Laksaci, O. Bouanani, Asymptotic normality of the k-NN single index regression estimator for functional weak dependence data, *Commun. Stat. Theory Meth.*, **53** (2024), 3143–3168. <https://doi.org/10.1080/03610926.2022.2150823>
- 138.J. S. Morris, Functional regression, *Annu. Rev. Stat. Appl.*, **2** (2015), 321–359. <https://doi.org/10.1146/annurev-statistics-010814-020413>
- 139.E. A. Nadaraja, On a regression estimate, *Teor. Veroyatnost. Primenen.*, **9** (1964), 157–159.
- 140.E. A. Nadaraya, Nonparametric estimation of probability densities and regression curves, In: *Mathematics and its Applications (Soviet Series)*, Dordrecht: Kluwer Academic Publishers Group, 1989.
- 141.G. P. Nason, R. von Sachs, G. Kroisandt, Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **62** (2000), 271–292. <https://doi.org/10.1111/1467-9868.00231>
- 142.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>
- 143.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>
- 144.D. Nolan, D. Pollard, U -processes: rates of convergence, *Ann. Statist.*, **15** (1987), 780–799.
- 145.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>
- 146.W. Peng, T. Coleman, L. Mentch, Rates of convergence for random forests via generalized U -statistics, *Electron. J. Stat.*, **16** (2022), 232–292. <https://doi.org/10.1214/21-EJS1958>
- 147.N. Phandoidaen, S. Richter, Empirical process theory for locally stationary processes, *Bernoulli*, **28** (2022), 453–480. <https://doi.org/10.3150/21-BEJ1351>
- 148.B. L. S. Prakasa Rao, A. Sen, Limit distributions of conditional U -statistics, *J. Theoret. Probab.*, **8** (1995), 261–301. <https://doi.org/10.1007/BF02212880>

149. M. B. Priestley, Evolutionary spectra and non-stationary processes, *J. Roy. Statist. Soc. Ser. B*, **27** (1965), 204–237. <https://doi.org/10.1111/j.2517-6161.1965.tb01488.x>
150. 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>
151. J. O. Ramsay, B. W. Silverman, *Applied Functional Data Analysis*, New York: Springer, 2002.
152. G. Rempala, A. Gupta, Weak limits of U -statistics of infinite order, *Random Oper. Stoch. Equ.*, **7** (1999), 39–52. <https://doi.org/10.1515/rose.1999.7.1.39>
153. K. Sakiyama, M. Taniguchi, Discriminant analysis for locally stationary processes, *J. Multivar. Anal.*, **90** (2004), 282–300. <https://doi.org/10.1016/j.jmva.2003.08.002>
154. A. Sen, Uniform strong consistency rates for conditional U -statistics, *Sankhyā Ind. J. Stat. Ser. A*, **56** (1994), 179–194.
155. R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons, 1980.
156. H. L. Shang, Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density, *J. Nonparametr. Stat.*, **26** (2014), 599–615. <https://doi.org/10.1080/10485252.2014.916806>
157. R. P. Sherman, The limiting distribution of the maximum rank correlation estimator, *Econometrica*, **61** (1993), 123–137.
158. R. P. Sherman, Maximal inequalities for degenerate U -processes with applications to optimization estimators, *Ann. Statist.*, **22** (1994), 439–459. <https://doi.org/10.1214/aos/1176325377>
159. B. W. Silverman, Distances on circles, toruses and spheres, *J. Appl. Probab.*, **15** (1978), 136–143. <https://doi.org/10.2307/3213243>
160. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall, 1986.
161. R. A. Silverman, Locally stationary random processes, *IRE T. Inform. Theory*, **3** (1957), 182–187. <https://doi.org/10.1109/TIT.1957.1057413>
162. Y. Song, X. Chen, K. Kato, Approximating high-dimensional infinite-order U -statistics: statistical and computational guarantees, *Electron. J. Stat.*, **13** (2019), 4794–4848. <https://doi.org/10.1214/19-EJS1643>
163. I. Soukariéh, S. Bouzebda, Exchangeably weighted bootstraps of general Markov U -process, *Mathematics*, **10** (2022), 3745. <https://doi.org/10.3390/math10203745>
164. I. Soukariéh, S. Bouzebda, Renewal type bootstrap for increasing degree U -process of a Markov chain, *J. Multivar. Anal.*, **195** (2023), 105143. <https://doi.org/10.1016/j.jmva.2022.105143>
165. I. Soukariéh, S. Bouzebda, Weak convergence of the conditional U -statistics for locally stationary functional time series, *Stat. Inference Stoch. Process.*, **17** (2024), 227–304. <https://doi.org/10.1007/s11203-023-09305-y>
166. W. Stute, Conditional U -statistics, *Ann. Probab.*, **19** (1991), 812–825.

167. W. Stute, L^p -convergence of conditional U -statistics, *J. Multivar. Anal.*, **51** (1994), 71–82. <https://doi.org/10.1006/jmva.1994.1050>
168. W. Stute, Universally consistent conditional U -statistics, *Ann. Statist.*, **22** (1994), 460–473. <https://doi.org/10.1214/aos/1176325378>
169. W. Stute, Symmetrized NN-conditional U -statistics. In: *Research Developments in Probability and Statistics*, 231–237, 1996.
170. W. Stute, W. and Wang, Multi-sample U -statistics for censored data, *Scand. J. Statist.*, **20** (1993), 369–374.
171. W. Stute, L. X. Zhu, Nonparametric checks for single-index models, *Ann. Statist.*, **33** (2005), <https://doi.org/10.1214/009053605000000020> 1048–1083.
172. K. K. Sudheesh, S. Anjana, M. Xie, U -statistics for left truncated and right censored data, *Statistics*, **57** (2023), 900–917. <https://doi.org/10.1080/02331888.2023.2217314>
173. Q. Tang, L. Kong, D. Rupper, R. J. Karunamuni, Partial functional partially linear single-index models, *Statist. Sinica*, **31** (2021), 107–133.
174. W. Y. Tsai, N. P. Jewell, M. C. Wang, A note on the product-limit estimator under right censoring and left truncation, *Biometrika*, **74** (1987), 883–886. <https://doi.org/10.1093/biomet/74.4.883>
175. A. van Delft, H. Dette, A general framework to quantify deviations from structural assumptions in the analysis of non-stationary function-valued processes, preprint paper, 2022. <https://doi.org/10.48550/arXiv.2208.10158>
176. A. van Delft, M. Eichler, Locally stationary functional time series, *Electron. J. Stat.*, **12** (2018), 107–170. <https://doi.org/10.1214/17-EJS1384>
177. A. van der Vaart, New donsker classes, *Ann. Probab.*, **24** (1996), 2128–2140. <https://doi.org/10.1214/aop/1041903221>
178. A. W. van der Vaart, J. A. Wellner, *Weak Convergence and Empirical Processes*, New York: Springer, 1996.
179. M. Vogt, Nonparametric regression for locally stationary time series, *Ann. Statist.*, **40** (2012), 2601–2633. <https://doi.org/10.1214/12-AOS1043>
180. V. A. Volkonskiui, Y. A. Rozanov, Some limit theorems for random functions I, *Theory Probab. Appl.*, **4** (1959), 178–197. <https://doi.org/10.1137/1104015>
181. R. von Mises, On the asymptotic distribution of differentiable statistical functions, *Ann. Math. Stat.*, **18** (1947), 309–348.
182. M. P. Wand, M. C. Jones, Kernel smoothing, In: *Monographs on Statistics and Applied Probability*, London: Chapman and Hall, 1995.
183. J. L. Wang, J. M. Chiou, H. G. Müller, Functional data analysis, *Annu. Rev. Stat. Appl.*, **3** (2016), 257–295. <https://doi.org/10.1146/annurev-statistics-041715-033624>
184. G. S. Watson, Smooth regression analysis, *Sankhyā Ind. J. Stat. Ser. A*, **26** (1964), 359–372.
185. J. Yang, Z. Zhou, Spectral inference under complex temporal dynamics, *J. Amer. Statist. Assoc.*, **117** (2022), 133–155. <https://doi.org/10.1080/01621459.2020.1764365>

- 186.A. Yuan, M. Giurcanu, G. Luta, M. T. Tan, U-statistics with conditional kernels for incomplete data models, *Ann. Inst. Statist. Math.*, **69** (2017), 271–302. <https://doi.org/10.1007/s10463-015-0537-6>
- 187.Y. Zhou, P. S. F. Yip, A strong representation of the product-limit estimator for left truncated and right censored data, *J. Multivar. Anal.*, **69** (1999), 261–280. <https://doi.org/10.1006/jmva.1998.1806>
- 188.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>

Appendix

A. Additional information

This appendix includes additional information that is integral to achieving a more comprehensive understanding of the paper.

Lemma A.1 ([165]). *Let $I_h = [C_1h, 1 - C_1h]$. Suppose that kernel K_1 satisfies Assumption 2 part(i). Then, for $q = 0, 1, 2$ and $m > 1$:*

$$\sup_{\mathbf{u} \in I_h^m} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) \left(\frac{u_k - i_k/n}{h_n} \right)^q - \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - v_k}{h_n} \right) \left(\frac{u_k - v_k}{h_n} \right)^q \right\} \prod_{k=1}^m dv_k \right| = O \left(\frac{1}{nh^{m+1}} \right).$$

Lemma A.2 ([165]). *Suppose that kernel K_1 satisfies Assumption 2 part (i) and let $g : [0, 1] \times \mathcal{H} \rightarrow \mathbb{R}$, $(\mathbf{u}, \mathbf{x}) \mapsto g(\mathbf{u}, \mathbf{x})$ be continuously differentiable with respect to \mathbf{u} . Then,*

$$\sup_{\mathbf{u} \in I_h} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h_n} \right) g \left(\frac{i_k}{n}, x_k \right) - \prod_{k=1}^m g(u_k, x_k) \right| = O \left(\frac{1}{nh^{m+1}} \right) + o(h_n). \quad (\text{A.1})$$

Lemma A.3 ([124]). *Let $\{Z_{i,n}\}$ be a zero-mean triangular array such that $|Z_{i,n}| \leq b_n$ with α -mixing coefficients $\alpha(k)$. Then, for any $\varepsilon > 0$ and $S_n \leq n$ with $\varepsilon > 4S_n b_n$,*

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_{i,i}(u, x) \right| \geq \varepsilon \right) \leq 4 \exp \left(- \frac{\varepsilon^2}{64\sigma_{S_n, n}^2 \frac{n}{S_n} + \frac{8}{3}\varepsilon b_n S_n} \right) + 4 \frac{n}{S_n} \alpha(S_n). \quad (\text{A.2})$$

Lemma A.4. *Let $\{Z_{i,n}\}$ be a zero-mean triangular array such that $|Z_{i,n}| \leq b_n$ with β -mixing coefficients $\beta(k)$. Then, for any $\varepsilon > 0$ and $S_n \leq n$ with $\varepsilon > 4S_n b_n$,*

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_{i,i}(u, x) \right| \geq \varepsilon \right) \leq 4 \exp \left(- \frac{\varepsilon^2}{64\sigma_{S_n, n}^2 \frac{n}{S_n} + \frac{8}{3}\varepsilon b_n S_n} \right) + 4 \frac{n}{S_n} \beta(S_n). \quad (\text{A.3})$$

Proof of Lemma A.4. Using Lemma A.3 and the fact that for any σ -algebra \mathcal{A} and \mathcal{B} , $\alpha(\mathcal{A}, \mathcal{B}) \subseteq \beta(\mathcal{A}, \mathcal{B})$, Lemma A.7 holds. \square

Lemma A.5 ([63]). Let X_1, \dots, X_n be a sequence of independent random elements taking values in a Banach space $(B, \|\cdot\|)$ with $\mathbb{E}X_i = 0$ for all i . Let $\{\varepsilon_i\}$ be a sequence of independent Bernoulli random variables independent of $\{X_i\}$. Then, for any convex increasing function Φ ,

$$\mathbb{E}\Phi\left(\frac{1}{2}\left\|\sum_{i=1}^n X_i\varepsilon_i\right\|\right) \leq \mathbb{E}\Phi\left(\left\|\sum_{i=1}^n X_i\right\|\right) \leq \mathbb{E}\Phi\left(2\left\|\sum_{i=1}^n X_i\varepsilon_i\right\|\right).$$

Proposition A.6 ([11]). Let $\{X_i : i \in n\}$ be a process satisfying, for $m \geq 1$:

$$\left(\mathbb{E}\|X_i - X_j\|^p\right)^{1/p} \leq \left(\frac{p-1}{q-1}\right)^{m/2} \left(\mathbb{E}\|X_i - X_j\|^q\right)^{1/q}, \quad 1 < q < p < \infty,$$

and the semi-metric :

$$\rho(j, i) = \left(\mathbb{E}\|X_i - X_j\|^2\right)^{1/2}.$$

There exists a constant $K = K(m)$ such that :

$$\mathbb{E} \sup_{i, j \in n} \|X_i - X_j\| \leq K \int_0^D [\log N(\epsilon, n, \rho)]^{m/2} d\epsilon,$$

where D is the ρ -diameter of n .

Lemma A.7 ([61]). 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 $p > 0$,

$$q > 1, p^{-1} + q^{-1} < 1.$$

Then,

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

Proof of Lemma A.7. This Lemma follows directly using Lemma A.7 and the fact that for any σ -algebra \mathcal{A} and \mathcal{B} , $\alpha(\mathcal{A}, \mathcal{B}) \subseteq \beta(\mathcal{A}, \mathcal{B})$. \square

Lemma A.8 ([180]). Let V_1, \dots, V_L be strongly mixing random variables measurable with respect to the σ -algebras $\mathcal{F}_{i_1}^{j_1}, \dots, \mathcal{F}_{i_L}^{j_L}$ respectively with $1 \leq i_1 < j_1 < i_2 < \dots < j_L \leq n$, $i_{l+1} - j_l \geq w \geq 1$ and $|V_j| \leq 1$ for $j = 1, \dots, L$. Then,

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

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

A1. Examples of classes of functions

Example A.9. The set \mathcal{F} of all indicator functions $\mathbb{I}_{(-\infty, t]}$ of cells in \mathbb{R} satisfies :

$$N(\epsilon, \mathcal{F}, d_{\mathbb{P}}^{(2)}) \leq \frac{2}{\epsilon^2},$$

for any probability measure \mathbb{P} and $\epsilon \leq 1$. Notice that :

$$\int_0^1 \sqrt{\log\left(\frac{1}{\epsilon}\right)} d\epsilon \leq \int_0^\infty u^{1/2} \exp(-u) du \leq 1.$$

For more details and discussion on this example refer to Example 2.5.4 of [178] and [114, p. 157]. The covering numbers of the class of cells $(-\infty, t]$ in higher dimension satisfy a similar bound, but with higher power of $(1/\epsilon)$, see Theorem 9.19 of [114].

Example A.10. (Classes of functions that are Lipschitz in a parameter, Section 2.7.4 in [178]). Let \mathcal{F} be the class of functions $x \mapsto \varphi(t, x)$ that are Lipschitz in the index parameter $t \in T$. Suppose that:

$$|\varphi(t_1, x) - \varphi(t_2, x)| \leq d(t_1, t_2)\kappa(x)$$

for some metric d on the index set T , the function $\kappa(\cdot)$ defined on the sample space \mathcal{X} , and all x . According to Theorem 2.7.11 of [178] and Lemma 9.18 of [114], it follows, for any norm $\|\cdot\|_{\mathcal{F}}$ on \mathcal{F} , that :

$$N(\epsilon \|F\|_{\mathcal{F}}, \mathcal{F}, \|\cdot\|_{\mathcal{F}}) \leq N(\epsilon/2, T, d).$$

Hence if (T, d) satisfy

$$J(\infty, T, d) = \int_0^\infty \sqrt{\log N(\epsilon, T, d)} d\epsilon < \infty,$$

then the conclusions holds for \mathcal{F} .

Example A.11. Let us consider as an example the classes of functions that are smooth up to order α defined as follows, see Section 2.7.1 of [178] and Section 2 of [177]. For $0 < \alpha < \infty$ let $[\alpha]$ be the greatest integer strictly smaller than α . For any vector $k = (k_1, \dots, k_d)$ of d integers define the differential operator

$$D^k := \frac{\partial^k}{\partial^{k_1} \dots \partial^{k_d}}, \text{ where } k := \sum_{i=1}^d k_i.$$

Then, for a function $f : \mathcal{X} \rightarrow \mathbb{R}$, let

$$\|f\|_\alpha := \max_{k \leq [\alpha]} \sup_x |D^k f(x)| + \max_{k = [\alpha]} \sup_x \frac{D^k f(x) - D^k f(y)}{\|x - y\|^{\alpha - [\alpha]}},$$

where the suprema are taken over all x, y in the interior of \mathcal{X} with $x \neq y$. Let $C_M^\alpha(\mathcal{X})$ be the set of all continuous functions $f : \mathcal{X} \rightarrow \mathbb{R}$ with

$$\|f\|_\alpha \leq M.$$

Note that for $\alpha \leq 1$ this class consists of bounded functions f that satisfy a Lipschitz condition. [112] computed the entropy of the classes of $C_M^\alpha(\mathcal{X})$ for the uniform norm. As a consequence of their

results, [177] shows that there exists a constant K depending only on α, d and the diameter of \mathcal{X} such that for every measure γ and every $\epsilon > 0$,

$$\log \mathcal{N}_{[\cdot]}(\epsilon M \gamma(\mathcal{X}), C_M^\alpha(\mathcal{X}), L_2(\gamma)) \leq K \left(\frac{1}{\epsilon}\right)^{d/\alpha},$$

$\mathcal{N}_{[\cdot]}$ is the bracketing number, refer to Definition 2.1.6 of [178] and we refer to Theorem 2.7.1 of [178] for a variant of the last inequality. By Lemma 9.18 of [114], we have

$$\log \mathcal{N}(\epsilon M \gamma(\mathcal{X}), C_M^\alpha(\mathcal{X}), L_2(\gamma)) \leq K \left(\frac{1}{2\epsilon}\right)^{d/\alpha}.$$

A2. Examples of U -kernels

In this section, we present some classical U -kernels.

Example A.12. [99] introduced the parameter

$$\Delta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D^2(y_1, y_2) dF(y_1, y_2),$$

where $D(y_1, y_2) = F(y_1, y_2) - F(y_1, \infty)F(\infty, y_2)$ and $F(\cdot, \cdot)$ is the distribution function of Y_1 and Y_2 . The parameter Δ has the property that $\Delta = 0$ if and only if Y_1 and Y_2 are independent. From [117], an alternative expression for Δ can be developed by introducing the functions

$$\psi(y_1, y_2, y_3) = \begin{cases} 1 & \text{if } y_2 \leq y_1 < y_3, \\ 0 & \text{if } y_1 < y_2, y_3 \text{ or } y_1 \geq y_2, y_3, \\ -1 & \text{if } y_3 \leq y_1 < y_2, \end{cases}$$

and

$$h(y_{1,1}, y_{1,2}, \dots, y_{5,1}, y_{5,2}) = \frac{1}{4} \psi(y_{1,1}, y_{1,2}, y_{1,3}) \psi(y_{1,1}, y_{1,4}, y_{1,5}) \psi(y_{1,2}, y_{2,2}, y_{3,2}) \psi(y_{1,2}, y_{4,2}, y_{5,2}).$$

We have

$$\Delta = \int \dots \int h(y_{1,1}, y_{1,2}, \dots, y_{5,1}, y_{5,2}) dF(y_{1,1}, y_{1,2}) \dots dF(y_{1,5}, y_{2,5}).$$

Example A.13. (Hoeffding's D). From the symmetric kernel,

$$\begin{aligned} h_D(z_1, \dots, z_5) &:= \frac{1}{16} \sum_{(i_1, \dots, i_5) \in \mathcal{P}_5} [\{\mathbb{1}(z_{i_1,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_2,1} \leq z_{i_5,1})\} \{\mathbb{1}(z_{i_3,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_4,1} \leq z_{i_5,1})\}] \\ &\quad \times [\{\mathbb{1}(z_{i_1,2} \leq z_{i_5,2}) - \mathbb{1}(z_{i_2,2} \leq z_{i_5,2})\} \{\mathbb{1}(z_{i_3,2} \leq z_{i_5,2}) - \mathbb{1}(z_{i_4,2} \leq z_{i_5,2})\}], \end{aligned}$$

we recover Hoeffding's D statistic, a rank-based U -statistic of order 5, and gives rise to Hoeffding's D correlation measure $\mathbb{E}h_D$.

Example A.14. (Blum-Kiefer-Rosenblatt's R). The symmetric kernel

$$\begin{aligned} h_R(z_1, \dots, z_6) \\ := \frac{1}{32} \sum_{(i_1, \dots, i_6) \in \mathcal{P}_6} [\{\mathbb{1}(z_{i_1,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_2,1} \leq z_{i_5,1})\} \{\mathbb{1}(z_{i_3,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_4,1} \leq z_{i_5,1})\}] \\ \times [\{\mathbb{1}(z_{i_1,2} \leq z_{i_6,2}) - \mathbb{1}(z_{i_2,2} \leq z_{i_6,2})\} \{\mathbb{1}(z_{i_3,2} \leq z_{i_6,2}) - \mathbb{1}(z_{i_4,2} \leq z_{i_6,2})\}] \end{aligned}$$

yields Blum-Kiefer-Rosenblatt's R statistic ([24]),

Example A.15. (Bergsma-Dassios-Yanagimoto's τ^*). [21] introduced a rank correlation statistic as a U -statistic of order 4 with the symmetric kernel

$$\begin{aligned} h_{\tau^*}(z_1, \dots, z_4) \\ := \frac{1}{16} \sum_{(i_1, \dots, i_4) \in \mathcal{P}_4} \{\mathbb{1}(z_{i_1,1}, z_{i_3,1} < z_{i_2,1}, z_{i_4,1}) + \mathbb{1}(z_{i_2,1}, z_{i_4,1} < z_{i_1,1}, z_{i_3,1}) \\ - \mathbb{1}(z_{i_1,1}, z_{i_4,1} < z_{i_2,1}, z_{i_3,1}) - \mathbb{1}(z_{i_2,1}, z_{i_3,1} < z_{i_1,1}, z_{i_4,1})\} \\ \times \{\mathbb{1}(z_{i_1,2}, z_{i_3,2} < z_{i_2,2}, z_{i_4,2}) + \mathbb{1}(z_{i_2,2}, z_{i_4,2} < z_{i_1,2}, z_{i_3,2}) \\ - \mathbb{1}(z_{i_1,2}, z_{i_4,2} < z_{i_2,2}, z_{i_3,2}) - \mathbb{1}(z_{i_2,2}, z_{i_3,2} < z_{i_1,2}, z_{i_4,2})\} \end{aligned}$$

Here, $\mathbb{1}(y_1, y_2 < y_3, y_4) := \mathbb{1}(y_1 < y_3) \mathbb{1}(y_1 < y_4) \mathbb{1}(y_2 < y_3) \mathbb{1}(y_2 < y_4)$.

Example A.16. The Wilcoxon Statistic. Suppose that $E \subset \mathbb{R}$ is symmetric around zero. As an estimate of the quantity

$$\int_{(x,y) \in E^2} \{2\mathbb{1}_{\{x+y>0\}} - 1\} dF(x)dF(y),$$

it is pertinent to consider the statistic

$$W_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \{2 \cdot \mathbb{1}_{\{X_i + X_j > 0\}} - 1\},$$

which is relevant for testing whether or not μ is located at zero.

Example A.17. The Takens estimator. Denote by $\|\cdot\|$ the usual Euclidean norm on \mathbb{R}^d . In [29], the following estimate of the correlation integral,

$$C_F(r) = \int \mathbb{I}_{\{\|x-x'\| \leq r\}} dF(x)dF(x'), \quad r > 0,$$

is considered:

$$C_n(r) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathbb{I}_{\{\|X_i - X_j\| \leq r\}}.$$

In the case where a scaling law holds for the correlation integral, i.e., when there exists $(\alpha, r_0, c) \in \mathbb{R}_+^3$ such that $C_F(r) = c \cdot r^{-\alpha}$ for $0 < r \leq r_0$, the U -statistic

$$T_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \log \left(\frac{\|X_i - X_j\|}{r_0} \right),$$

is used in order to build the Takens estimator $\hat{\alpha}_n = -T_n^{-1}$ of the correlation dimension α .

Example A.18. Let $\widehat{Y_1 Y_2}$ denote the oriented angle between $Y_1, Y_2 \in T$, T is the circle of radius 1 and center 0 in \mathbb{R}^2 . Let:

$$h_t(Y_1, Y_2) = \mathbb{1}\{\widehat{Y_1 Y_2} \leq t\} - t/\pi, \text{ for } t \in [0, \pi).$$

Reference [159] has used this kernel in order to propose a U -process to test uniformity on the circle.

Example A.19. For $m = 3$, let :

$$\varphi(Y_1, Y_2, Y_3) = \mathbb{1}\{Y_1 - Y_2 - Y_3 > 0\},$$

We have

$$r^{(3)}(\varphi, t_1, t_2, t_3) = \mathbb{P}(Y_1 > Y_2 + Y_3 \mid X_1 = X_2 = X_3 = t)$$

and the corresponding conditional U -Statistic can be considered a conditional analog of the Hollander-Proschan test-statistic ([101]). It may be used to test the hypothesis that the conditional distribution of Y_1 given $X_1 = t$, is exponential, against the alternative that it is of the New-Better than-Used-type.

Example A.20. The Gini mean difference. The Gini index provides another popular measure of dispersion. It corresponds to the case where $E \subset \mathbb{R}$ and $h(x, y) = |x - y|$:

$$G_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} |X_i - X_j|.$$

Example A.21 ([113]). Let the sample central moments of any order $m = 2, 3, \dots$ be given by

$$\theta_m(F) = \mathbb{E}(X_1 - \mathbb{E}X_1)^m = \int (x - \mathbb{E}X_1)^m dF(x).$$

In this case, the U -statistic has a symmetric kernel

$$h(x_1, \dots, x_m) = \frac{1}{m!} \sum \left[x_{i_1}^m - \binom{m}{1} x_{i_1}^{m-1} x_{i_2} + \binom{m}{2} x_{i_1}^{m-2} x_{i_2} x_{i_3} - \dots + (-1)^{m-1} \left(\binom{m}{m-1} - 1 \right) x_{i_1} x_{i_2} \dots x_{i_m} \right],$$

where summation is carried out over all permutations (i_1, \dots, i_m) of the numbers $(1, \dots, m)$. In particular, if $m = 3$, then

$$h(x_1, x_2, x_3) = \frac{1}{3} (x_1^3 + x_2^3 + x_3^3) - \frac{1}{2} (x_1^2 x_2 + x_2^2 x_1 + x_1^2 x_3 + x_3^2 x_1 + x_2^2 x_3 + x_3^2 x_2) + 2x_1 x_2 x_3.$$

In the case of $m = 2$,

$$\theta_2(F) = \mathbb{E}(X_1 - \mathbb{E}X_1)^2 = \int (x - \mathbb{E}X_1)^2 dF(x).$$

For the kernel

$$h(x_1, x_2) = \frac{x_1^2 + x_2^2 - 2x_1 x_2}{2} = \frac{1}{2} (x_1 - x_2)^2.$$

the corresponding U -statistic is the sample variance

$$\begin{aligned} U_n(h) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n \left\{ \frac{1}{n} \sum_{i=1}^n X_i \right\}^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right), \end{aligned}$$

we refer also to [155].



AIMS Press

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