ISSN 1066-5307, Mathematical Methods of Statistics, 2019, Vol. 28, No. 3, pp. 169–207. © Allerton Press, Inc., 2019.

Central Limit Theorems for Conditional Empirical and Conditional U-Processes of Stationary Mixing Sequences

S. Bouzebda[1]* and B. Nemouchi[1]**

[1] *Sorbonne Univ. de Technologie de Compiègne, L.M.A.C., Compiègne, France*

Received August 6, 2018; in final form, July 7, 2019

Abstract—In this paper we are concerned with the weak convergence to Gaussian processes of conditional empirical processes and conditional U-processes from stationary β-mixing sequences indexed by classes of functions satisfying some entropy conditions. We obtain uniform central limit theorems for conditional empirical processes and conditional U-processes when the classes of functions are uniformly bounded or unbounded with envelope functions satisfying some moment conditions. We apply our results to introduce statistical tests for conditional independence that are multivariate conditional versions of the Kendall statistics.

Keywords: conditional empirical processes, conditional U-processes, uniform central limit theorems, VC-classes, stationary sequence, absolutely regular sequences.

AMS 2010 Subject Classification: 60F05, 60G15, 60G10, 62G08, 62G07.

DOI: 10.3103/S1066530719030013

1. INTRODUCTION

Motivated by numerous applications, the theory of U-statistics (introduced in a seminal work by Hoeffding [32]) and U-processes has received considerable attention in the past decades. U-processes are useful for solving complex statistical problems. For example, for density estimation, nonparametric regression tests and goodness-of-fit tests. More precisely, U-processes appear in statistics in many different instances, e.g. as the components of higher order terms in von Mises expansions. In particular, U-statistics are used in the analysis of estimators (including function estimators) with varying degrees of smoothness. For example, Stute [64] applies a.s. uniform bounds for $\mathbb{P}$-canonical U-processes to the analysis of the product limit estimator for truncated data. Arcones and Wang [5] present two new tests for normality based on U-processes. Making use of the results of [26, 27], Schick et al. [55] introduced new tests for normality using weighted L_1-distances between the standard normal density and local U-statistics based on standardized observations. Joly and Lugosi [36] discussed the estimation of the mean of multivariate functions in case of possibly heavy-tailed distributions and introduce the median-of-means that is based on U-statistics. U-processes are important tools for a broad range of statistical applications such as testing for qualitative features of functions in nonparametric statistics [43, 25, 1], cross-validation for density estimation [40], and establishing limiting distributions of M-estimators (see, e.g., [3, 58, 59, 12]), Halmos [29], von Mises [68] and Hoeffding [32], who provided (among others) the first asymptotic results for the case that the underlying random variables are independent and identically distributed. Under weak dependence assumptions asymptotic results are for instance shown in [8], in [15], or more recently in [44] and in more general setting in [45]. For excellent resource of references on the U-statistics and U-processes the interested reader may refer to [9, 39, 42, 4, 7, 3]. For the U-statistics with random kernels of diverging orders we refer to [24, 53, 31, 61]. Infinite-order U-statistics are a useful tool for constructing simultaneous prediction intervals that quantify the uncertainty of ensemble methods such as subbagging and random forests. Song et al. (2019) (unpublished preprint) provide the following important example.

*E-mail: salim.bouzebda@utc.fr

**E-mail: boutheina.nemouchi@utc.fr

Example 1.0.1. Simultaneous prediction intervals for random forests [61]. Consider a training dataset of size n, $\{(Y_1,Z_1),\dots,(Y_n,Z_n)\}=\{X_1,\dots,X_n\}=X_1^n$, where $Y_i\in\mathcal{Y}$ is a vector of features and $Z_i\in\mathbb{R}$ is a response. Let h be a deterministic prediction rule that takes as input a sub-sample $\{X_{i_1},\dots,X_{i_m}\}$ with $1\le m\le n$ and outputs predictions at d testing points $(y_1^*,\dots,y_d^*)$ in the feature space $\mathcal{Y}$. For random forests, the tree-based prediction rule is constructed on each sub-sample with additional randomness. Specifically, let $\{W_\iota\colon \iota\in I(m,n)\}$ be a collection of i.i.d. random variables taking values in a measurable space $(S',\mathcal{S}')$ that are independent of the data X_1^n. Let $H\colon S^m\times S'\to\mathbb{R}^d$ be an $\mathcal{S}^m\otimes\mathcal{S}'$-measurable function such that $\mathbb{E}[H(x_1,\dots,x_m,W)]=h(x_1,\dots,x_m)$. Then predictions of random forests are given by a d-dimensional U-statistic with random kernel H:

$$\widehat{U}_n:=\frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I(m,n)}H(X_{i_1},\dots,X_{i_m},W_{\mathbf{i}}). \tag{1.1}$$

where the random kernel H varies with m and

$$I(m,n)=\{\mathbf{i}=(i_1,\dots,i_m)\colon 0\le i_j\le n \quad\text{and}\quad i_j\neq i_r \quad\text{if}\quad j\neq r\}$$

is the set of all m-tuples of different integers between 1 and n.

A very deep insight into the theory of U-processes is given in [12]. In this paper we consider the so-called conditional U-statistics introduced in [63]. These statistics may be viewed as generalizations of the Nadaraya–Watson ([46], [69]) estimates of a regression function. To be more precise, let us consider the strictly stationary sequence of random elements $\{(\mathbf{X}_i,\mathbf{Y}_i), i\in\mathbb{N}^*\}$ defined on the probability space $(\Omega,\mathcal{A},\mathbb{P})$ with $\mathbf{X}_i\in\mathbb{R}^d$ and $\mathbf{Y}_i\in\mathbb{R}^d$, and a measurable function $\varphi\colon\mathbb{R}^{dm}\to\mathbb{R}$. In this paper, we are primarily concerned with the estimation of the conditional expectation, or the regression function,

$$r^{(m)}(\varphi,\mathbf{t})=\mathbb{E}\left(\varphi(\mathbf{Y}_1,\dots,\mathbf{Y}_m)\mid(\mathbf{X}_1,\dots,\mathbf{X}_m)=\mathbf{t}\right)\quad\text{for}\quad \mathbf{t}\in\mathbb{R}^{dm},$$

whenever it exists, i.e., $\mathbb{E}\left(|\varphi(\mathbf{Y}_1,\dots,\mathbf{Y}_m)|\right)<\infty$. We now introduce a kernel function $K:\mathbb{R}^d\to\mathbb{R}$ with support contained in $[-B,B]^d$, $0<B<\infty$, satisfying

$$\sup_{\mathbf{x}\in\mathbb{R}^d}|K(\mathbf{x})|=:\kappa<\infty\quad\text{and}\quad\int K(\mathbf{x})\,d\mathbf{x}=1. \tag{1.2}$$

Stute [63] introduced a class of estimators for $r^{(m)}(\varphi,\mathbf{t})$, called conditional U-statistics, which is defined for each $\mathbf{t}\in\mathbb{R}^{dm}$ by

$$\widehat{r}_n^{(m)}(\varphi,\mathbf{t};h_n)=\frac{\sum_{(i_1,\dots,i_m)\in I(m,n)}\varphi(\mathbf{Y}_{i_1},\dots,\mathbf{Y}_{i_m})K\big(\frac{\mathbf{X}_{i_1}-\mathbf{t}_1}{h_n}\big)\cdots K\big(\frac{\mathbf{X}_{i_m}-\mathbf{t}_m}{h_n}\big)}{\sum_{(i_1,\dots,i_m)\in I(m,n)}K\big(\frac{\mathbf{X}_{i_1}-\mathbf{t}_1}{h_n}\big)\cdots K\big(\frac{\mathbf{X}_{i_m}-\mathbf{t}_m}{h_n}\big)}, \tag{1.3}$$

where $\{h=h_n\}_{n\ge1}$ is a sequence of positive constants converging to zero at the rate $\big(nh^{dm}=nh_n^{dm}\big)\to\infty$. In the particular case $m=1$, the $r^{(m)}(\varphi,\mathbf{t})$ is reduced to $r^{(1)}(\varphi,\mathbf{t})=\mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X}=\mathbf{t})$ and Stute's estimator becomes the Nadaraya–Watson estimator of $r^{(1)}(\varphi,\mathbf{t})$ given by

$$\widehat{r}_n^{(1)}(\varphi,\mathbf{t},h_n)=\sum_{i=1}^n\varphi(\mathbf{Y}_i)K\left(\frac{\mathbf{X}_i-\mathbf{t}}{h_n}\right)\Big/\sum_{i=1}^nK\left(\frac{\mathbf{X}_i-\mathbf{t}}{h_n}\right).$$

The work of Sen [56] was devoted to estimate the rate of the uniform convergence in $\mathbf{t}$ of $\widehat{r}_n^{(m)}(\varphi,\mathbf{t};h_n)$ to $r^{(m)}(\varphi,\mathbf{t})$. In [52], the limit distributions of $\widehat{r}_n^{(m)}(\varphi,\mathbf{t};h_n)$ are discussed and compared with those obtained by Stute. Harel and Puri [30] extend the results of Stute [63], under appropriate mixing conditions, to weakly dependent data and apply their findings to verify the Bayes risk consistency of the corresponding discrimination rules. Stute [65] proposed symmetrized nearest neighbor conditional U-statistics as alternatives to the usual kernel-type estimators. An important contribution is given in [16], where a much stronger form of consistency holds, namely, uniform in $\mathbf{t}$ and in bandwidth consistency (i.e., $h_n\in[a_n,b_n]$, where $0<a_n<b_n\to0$ at some specific rate) of $\widehat{r}_n^{(m)}(\varphi,\mathbf{t};h_n)$. In addition, uniform consistency is also established over $\varphi\in\mathscr{F}$ for a suitably restricted class of functions $\mathscr{F}$. The main tool in

their result is the use of local conditional U-processes investigated in [26]. Let $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ be a stationary sequence of random variables on some probability space $(\Omega, \mathcal{D}, P)$ and let σ_i^j be the σ-field generated by $\mathbf{Z}_i, \dots, \mathbf{Z}_j$, for $i, j \geqslant 1$. The sequence $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ is said β-mixing or absolute regular if

$$\beta(k) := \mathbb{E} \sup_{l \geqslant 1} \left\{ |P(A|\sigma_1^l) - P(A)| \colon A \in \sigma_{l+k}^{\infty} \right\} \longrightarrow 0 \quad \text{as} \quad k \to \infty.$$

A more general definition is introduced in the Appendix in addition to other measures of dependence. Throughout the sequel, we assume tacitly that sequence of random elements $\{(\mathbf{X}_i, \mathbf{Y}_i), i \in \mathbb{N}^*\}$ is absolutely regular.

In this paper, we are mainly interested in extensions of the results of [6] to the conditional U-processes. More precisely, we investigate the weak convergence of the conditional empirical process indexed by a suitable class of functions and of conditional U-processes. We treat the weak convergence in both cases when the class of functions is bounded or unbounded satisfying some moment conditions. It is important to notice that in [6] the weak convergence of the U-processes is obtained for uniformly bounded class of functions. To the best of our knowledge, these results have not yet been investigated, and this give the main motivation of this study. Although Arcones and Yu [6] have established some weak convergence, their results are not directly applicable in our framework since we are interested in some local empirical processes. However their results will be essential to obtain our main theorems. We shall not discuss removing the bias in general. The bias is not probabilistic and can always be studied by adding enough smoothness to the kernel and the regression function. In order to obtain our general results we make use and combine several techniques: blocking, chaining argument method, square root trick, Hoeffding's trick, etc. Several inequalities, Eberlein's inequality, Bernstein's inequality, Hoeffding's inequality and others, are used in nontrivial ways to obtain some desired bounds.

The layout of the present article is as follows. Section 2 is devoted to the weak convergence of the conditional empirical processes indexed by a function in the spirit of [51] and [50]. In Section 3, we investigate weak convergence of the conditional U-processes indexed by a class of functions satisfying some entropy conditions. In Section 4, we provide an application of our results to testing the conditional independence. In Section 5, we collect some examples of classes of functions and some U-statistics. Some concluding remarks and possible future developments are relegated to Section 6. To avoid interrupting the flow of our presentation, all mathematical developments are given in Section 7. We recall some facts and technical results that we need in our proof in the Appendix.

2. CONDITIONAL EMPIRICAL PROCESSES

Let $\mathscr{F}_m = \{\varphi \colon \mathbb{R}^{dm} \to \mathbb{R}\}$ denote a class of real-valued symmetric measurable functions on $\mathbb{R}^{dm}$ with a measurable envelope function

$$F(\mathbf{y}) \geq \sup_{\varphi \in \mathscr{F}_m} |\varphi(\mathbf{y})| \quad \text{for} \quad \mathbf{y} \in \mathbb{R}^{dm}. \tag{2.1}$$

For a kernel function $K(\cdot)$ and a measurable set $\mathbb{I} \subset \mathbb{R}^d$, we define the class of functions

$$\mathscr{K}^m := \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m K\left(\frac{x_i - \mathbf{t}_i}{h_n}\right), (\mathbf{t}_1, \dots, \mathbf{t}_m) \in \mathbb{I}^m \right\}.$$

$\mathbb{I}$ is chosen so that $(x_i - t)/h_n$ be contained in $[-B, B]$, the support of K, and the classes $\mathscr{F}_m$, $\mathscr{K}^m$ are assumed to be pointwise measurable classes. From now on, to stress the role of the class $\mathscr{K}^m$, we shall write $\widehat{r}_n^{(m)}(\psi, \mathbf{t}; h_n)$ or $\widehat{r}_n^{(m)}(\psi, \mathbf{t})$ for the estimator of the regression function defined in (1.3), where $\psi = \varphi \prod_{i=1}^m K\left(\frac{\cdot - \mathbf{t}_i}{h_n}\right) \in \mathscr{F}_m \mathscr{K}^m$, and by misuse of notation we write $r^{(m)}(\psi, \mathbf{t})$ for $r^{(m)}(\varphi, \mathbf{t})$ and $\nu_n(\psi \mid \mathbf{t})$ to denote $\nu_n(\varphi \mid \mathbf{t})$ that we will define later. Nevertheless, in some cases we keep the notation $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \widehat{r}_n^{(m)}(\varphi, \mathbf{t})$. Let us introduce the following class of functions on $\mathbb{R}^d \times \mathbb{R}^d$:

$$\mathscr{F}\mathscr{K} = \mathscr{F}_1 \mathscr{K}^1 := \left\{ (\cdot, \cdot) \mapsto \varphi(\cdot) K\left(\frac{\cdot - \mathbf{t}}{h_n}\right) \colon \varphi \in \mathscr{F}_1, \mathbf{t} \in \mathbb{I} \right\}.$$

The main objective of this section is to investigate the uniform central limit theorems for the conditional empirical process defined by

$$\left\{\nu_n(\psi \mid \mathbf{t}) = \sqrt{nh^d}\,(\widehat{r}_n^{(1)}(\psi, \mathbf{t}; h_n) - r^{(1)}(\psi, \mathbf{t}),\ \psi \in \mathscr{F}\mathscr{K},\ \mathbf{t} \in \mathbb{I}\right\}.$$

If, for $\mathbb{P}\psi = \int \psi\, d\mathbb{P}$, where $\mathbb{P}$ is a probability measure and for each $(\mathbf{x}, \mathbf{y})$

$$\sup_{\psi \in \mathscr{F}\mathscr{K}} |\psi(\mathbf{x}, \mathbf{y}) - \mathbb{P}\psi| < \infty,$$

then $\left\{\nu_n(\psi \mid \mathbf{t})\colon \psi \in \mathscr{F}\mathscr{K}, \mathbf{t} \in \mathbb{I}\right\}$ is a random element with values in $l_\infty(\mathscr{F}\mathscr{K})$ consisting of all functionals ν_∞ on $\mathscr{F}\mathscr{K}$ such that

$$\sup_{\psi \in \mathscr{F}\mathscr{K}} |\nu_\infty(\psi)| < \infty.$$

Then it will be important to investigate the following weak convergence:

$$\{\nu_n(\psi \mid \mathbf{t})\colon \psi \in \mathscr{F}\mathscr{K}, \mathbf{t} \in \mathbb{I}\} \xrightarrow{w} \{\mathbb{G}(\psi) : \psi \in \mathscr{F}\mathscr{K}\} \quad \text{in} \quad l_\infty(\mathscr{F}\mathscr{K}),$$

in the sense of [33], refer to Definition 1.3.3 in [67], where $\{\mathbb{G}(\psi) : \psi \in \mathscr{F}\mathscr{K}\}$ is the Gaussian process indexed by the class of functions $\mathscr{F}\mathscr{K}$. It is known that the weak convergence to a Gaussian limit with a version of uniformly bounded and uniformly continuous paths (with respect to the $\|\cdot\|_2$) is equivalent to the finite-dimensional convergence and the existence of pseudo-metric $d_{\mathrm{p.m}}$ on $\mathscr{F}\mathscr{K}$ such that $(\mathscr{F}\mathscr{K}, d_{\mathrm{p.m}})$ is totally bounded pseudo-metric space and

$$\lim_{r\to 0} \limsup_{n\to\infty} \mathbb{P}^*\left\{ \sup_{d_{\mathrm{p.m}}(\psi_1,\psi_2)\le r} |\nu_n\left((\psi_1 - \psi_2) \mid \mathbf{t}\right)| > \varepsilon \right\} = 0 \tag{2.2}$$

for all $\varepsilon > 0$, where $\mathbb{P}^*$ is the outer probability, see p. 6 in [67] for definition. As mentioned earlier, Poryvai [51] provided finite-dimensional convergence results for dependent sequences, we will borrow some techniques from this work, which will adapted to our framework in connection with the methodology of [6]. We will show henceforth that the Eq. (2.2) holds for the VC subgraph classes and classes satisfying some entropy conditions.

2.1. Convergence of Finite-Dimensional Distributions

Below, we write $Z \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2, $\stackrel{d}{\to}$ denotes convergence in distribution and $\stackrel{\mathbb{P}}{\to}$ convergence in probability. To unburden our notation a bit, we assume the $\{X_i\}_{i\in\mathbb{N}^*}$ to be real-valued ($d = 1$) and suppose that the distribution of the r.v. X_1 admits a density $\rho(\cdot)$ with respect to the Lebesgue measure. In what follows, we assume that $\rho(t) > 0$. We will remark that our results extend with obvious changes of notation and modifications of assumptions to the case when $\mathbf{X}$ takes values in $\mathbb{R}^d$, for $d > 1$. The proofs carry over exactly as before after some additional but cumbersome indexing. Let us introduce

$$\nu_{ni} = \sqrt{\frac{h}{n}}\,\breve{\varphi}(Y_i) K_h(X_i - t), \quad \forall i = 1, \ldots, n, \quad \text{for} \quad t \in \mathbb{R},$$

where

$$\breve{\varphi} = \varphi - r^{(1)}(\varphi, t) \quad \text{and} \quad K_h(t) = h^{-1} K\left(h^{-1} t\right).$$

In this article, the finite-dimensional convergence requires the following assumptions:

(C.1) The functions $\rho(z)$ and $r^{(1)}(\varphi, z)$ (for all $\varphi \in \mathscr{F}$) are twice continuously differentiable with respect to z in some neighborhood of the point t.

(C.2) The function $K\colon \mathbb{R} \to \mathbb{R}$ with support contained in $[-B, B]$, $0 < B < \infty$, satisfies

$$\int_{\mathbb{R}} K(z)\, dz = 1, \qquad \int_{\mathbb{R}} K^2(z)\, dz < \infty, \qquad K(-z) = K(z) \quad \text{for} \quad z \in \mathbb{R},$$

(C.3) We assume that $h = h_n \to 0$ as $n \to \infty$ and is such that $nh^{5/2} \searrow c = cte < \infty$ with $nh \to \infty$ as $n \to \infty$.

(C.4) The new sequence $\{\nu_{ni}, 1 \leq i \leq n\}$ is formed by strictly stationary β-mixing r.v.'s and the joint distribution of (X_1, X_j) for $j > 1$ admits a density $\rho_j(z_1, z_j)$ with respect to the Lebesgue measure such that the following conditions hold for all $(z_1, z_j) \in \mathbb{R}^2$:

$$|\rho_j(z_1, z_j) - \rho(z_1)\rho(z_j)| \leq \theta(j)\rho(z_1)\rho(z_j) \quad \text{for} \quad \theta(j) < 1,$$

and

$$h \sum_{j=2}^{n} \theta(j) \to 0 \quad \text{as} \quad n \to \infty.$$

(C.5) The function $r^{(1)}(\varphi_1\varphi_2, z)$ is continuous in z at the point t and

$$\sup_{j>1} \sup_{(z_1,z_j)\in\mathbb{R}^2} |\mathbb{E}(\breve{\varphi}_1(Y_1)\breve{\varphi}_2(Y_j) \mid X_1 = z_1, X_j = z_j| < \infty \quad \text{for} \quad \varphi_1, \varphi_2 \in \mathscr{F}.$$

(C.6) The sequence $\{Z_i = (X_i, Y_i)\}_{i\in\mathbb{N}^*}$ satisfies for $p > 0$: $\left(nh \max_{1\leq k\leq n} \beta_k^{1/p}\right) \to 0$ as $n \to \infty$. Moreover, $r^{(1)}(\|\breve{\varphi}\|_q, z_1) < \infty$, $r^{(1)}(\|\breve{\varphi}\|_s, z_j) < \infty$ for all $j \geqslant 2$, where $q > 1$, $s > 1$ and

$$\frac{1}{q} + \frac{1}{s} = 1 - \frac{1}{p},$$

where we define, for $\{X_j\}_{1\leq j\leq n}$ and almost all $\{z_j\}_{1\leq j\leq n} \in \mathbb{R}$, for $r = q, s$,

$$r^{(1)}(\|\breve{\varphi}\|_r, z_j) = \mathbb{E}^{1/r}\left(|\breve{\varphi}^r(Y_j)| \,\middle|\, X_j = z_j\right).$$

We are now equipped to state our theorem.

Theorem 2.1.1. *Let us consider the class of functions $\mathscr{FK}$ such that $\mathbb{E}\varphi^2(Y_1) < \infty$ and suppose that conditions* **(C.1)–(C.6)** *hold. Then, for $\mathfrak{M} \geq 1$ and $\psi_1, \ldots, \psi_{\mathfrak{M}} \in \mathscr{FK}$ we have*

$$\{\nu_n(\psi_i \mid t) : i = 1, \ldots, \mathfrak{M}\} \xrightarrow{d} \mathcal{N}(0, \Sigma) \quad \textit{as} \quad n \to \infty,$$

where $\Sigma := (\sigma_{i,j})_{i,j=1,\ldots,\mathfrak{M}}$ and

$$\sigma_{i,j} := \left(r^{(1)}(\psi_i\psi_j, t) - r^{(1)}(\psi_i, t)r^{(1)}(\psi_j, t)\right)\left\{\frac{\int_{\mathbb{R}} K^2(u)\, du}{\rho(t)}\right\}.$$

2.2. Asymptotic Uniform Equicontinuity (Tightness)

In this part, we give sufficient conditions for (2.2) to hold for the process

$$\left\{\nu_n(\psi \mid t) = \sqrt{nh}(\widehat{r}_n^{(1)}(\psi, t, h_n) - r^{(1)}(\psi, t), \; \psi \in \mathscr{FK}, \; t \in \mathbb{I}\right\}.$$

To do this we need to measure the size of the class $\mathscr{FK}$ that could be achieved in a simple way by metric entropy. The metric entropy of a class of functions $\mathscr{E}$ with respect to the pseudo-metric $d_{\text{p.m}}$ is defined by $\log \mathcal{N}(\epsilon, \mathscr{E}, d_{\text{p.m}})$, where

$$\mathcal{N}(\epsilon, \mathscr{E}, d_{\text{p.m}}) = \min\left\{m\colon \text{there are } f_1, \ldots, f_m \in \mathscr{E} \text{ such that } \sup_{f\in\mathscr{E}} \min_{1\leq j\leq m} d_{\text{p.m}}(f, f_j) \leq \epsilon\right\}$$

is the covering number. Recall that the envelope function given in (2.1), for $m = 1$, is

$$F(y) \geq \sup_{\varphi\in\mathscr{F}_1} |\varphi(y)|, \text{ for } y \in \mathbb{R}, \tag{2.3}$$

and from equation (1.2), we have $\sup_{x\in\mathbb{R}}|K(x)| =: \kappa < \infty$. Let us denote, for $m = 1$,

$$L_r(\mathbb{Q}) \;=\; ||\cdot||_{L_r(\mathbb{Q})} = \left(\int |\cdot|^r \, d\mathbb{Q}\right)^{\frac{1}{r}},$$

$$\gamma_n(t) \;:=\; \frac{1}{nh_n}\sum_{i=1}^{n}\varphi(Y_i)K\left(\frac{X_i - t}{h_n}\right).$$

(C.7) For some $2 < p < \infty$, the β-mixing coefficient satisfies

$$k^{\frac{p}{(p-2)}}(\log k)^{\frac{2(p-1)}{(p-2)}}\beta_k \longrightarrow 0 \quad \text{as} \quad k \to \infty.$$

(C.8) The conditional moment, for some $2 < p < \infty$,

$$\mu_p := \sup_{t\in\mathbb{I}} \mathbb{E}(F^p(Y) \mid X = t) < \infty. \tag{2.4}$$

(C.9) The metric entropy of the class $\mathscr{F}\mathscr{K}$ satisfies, for some $2 < p < \infty$,

$$\int_0^\infty (\log N(u, \mathscr{F}\mathscr{K}, \|\cdot\|_1))^{\frac{1}{2}} du < \infty,$$
$$\int_0^\infty (\log N(u, \mathscr{F}\mathscr{K}, \|\cdot\|_2))^{\frac{1}{2}} du < \infty,$$
$$\int_0^\infty (\log N(u, \mathscr{F}\mathscr{K}, \|\cdot\|_p))^{\frac{1}{2}} du < \infty.$$

(C.10) There are two positive constants b and ν such that:

$$N(\epsilon, \mathscr{F}\mathscr{K}, \|\cdot\|_{L_2(\mathbb{Q})}) \leq \left(\frac{b\|F\kappa\|_{L_2(\mathbb{Q})}}{\epsilon}\right)^{\nu}$$

for any $\epsilon > 0$ and each probability measure such that

$$\mathbb{Q}(F\kappa)^2 := \int (F\kappa)^2 d\mathbb{Q} < \infty,$$

where $\mathbb{Q}$ is any probability measure on $\big(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)\big)$, and $\mathcal{B}\,(\mathbb{R}^m)$ represents the σ-field of Borel sets of $\mathbb{R}^m$ (in the present section we consider $m = 1$).

Remark 2.2.1. Imposing such conditions is mainly for technical reasons:

- Condition **(C.1)** and $nh^{5/2} \to 0$ allow us to apply Proposition 3.4 of [73] in order to prove that $nh^{5/2}\mu_n' \to 0$ (defined in the proof). Imposing $nh^{5/2} \to 0$ could be replaced by the condition $\sqrt{nh^5} \searrow c = \text{const} < \infty$.
- Conditions **(C.2)**, **(C.3)** and **(C.8)** are classical conditions in kernel-type estimation.
- Conditions **(C.4)**–**(C.6)** and **(C.7)** are technical conditions used in order to calculate $\mathrm{Var}\,(\nu_{n1})$ and prove that ν_{ni} are asymptotically uncorrelated and also to get $(\upsilon_n - 1)\beta_{a_n} \to 0$ in the Eberlein's inequality.
- Imposing conditions **(C.9)** and **(C.10)** means that we are dealing with a class of functions neither too small nor too large with the property of being totally bounded. The latter allows us to study the asymptotic equicontinuity of the process. We impose such entropy properties in order to get that $(\mathscr{F}\mathscr{K}, \|\cdot\|_r)$ is totally bounded and also for technical reasons, it allows us to use some inequalities and propositions in order to get the asymptotic tightness.

We were inspired by [51] to impose conditions **(C.1)–(C.5)** and by [6] to impose **(C.7)**, **(C.9)** and **(C.10)**.

Our main result of this section is the following.

Theorem 2.2.2. *Suppose that conditions* **(C.2)**, **(C.7)–(C.10)** *hold and* $\mathbb{E}(\varphi^2(Y_1)) < \infty$ *for each* $\varphi \in \mathscr{F}$. *Then*

$$\lim_{b\to 0}\limsup_{n\to\infty}\mathbb{P}\Big\{\sup_{\substack{\|\psi_1-\psi_2)\|_p\le b\\ \psi_1,\psi_2\in\mathscr{F}\mathscr{K}}}\big|\nu_n\left((\psi_1-\psi_2)\mid t\right)\big|>\varepsilon\Big\}=0.$$

The proof of this theorem is based on the blocking approach and essentially on technics used in [6]: the main idea is to divide the strictly stationary sequence $(Z_1,\ldots,Z_n)$ into a number of blocks equal to $2\upsilon_n$ such that each one is of length equal to a_n and the remaining block is of length $n-2\upsilon_n a_n$, that is (for $1\le j\le \upsilon_n$)

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

The values of υ_n, a_n are given in the proof. Then introduce the sequence of independent blocks $(\xi_1,\ldots,\xi_n)$ such as:

$$\mathcal{L}(\xi_1,\ldots,\xi_n)=\mathcal{L}(Z_1,\ldots,Z_{a_n})\times\mathcal{L}(Z_{a_n+1},\ldots,Z_{2a_n})\times\cdots.$$

An application of the result of [21] implies that, for any measurable set A,

$$\begin{aligned} &\Big|\mathbb{P}\left\{(\xi_1,\ldots,\xi_{a_n},\xi_{2a_n+1},\ldots,\xi_{3a_n},\ldots,\xi_{2(\upsilon_n-1)a_n+1},\ldots,\xi_{2\upsilon_n a_n})\in A\right\}\\ &\qquad -\mathbb{P}\left\{(Z_1,\ldots,Z_{a_n},Z_{2a_n+1},\ldots,Z_{3a_n},\ldots,Z_{2(\upsilon_n-1)a_n+1},\ldots,Z_{2\upsilon_n a_n})\in A\right\}\Big|\\ &\le 2(\upsilon_n-1)\beta_{a_n}. \end{aligned} \tag{2.5}$$

Based on the work of Ibragimov [35] and that of Doukhan et al. [17], Arcones and Yu [6] showed that, for each stationary absolute regular sequence $(Z, Z_1,\ldots,Z_n)$ satisfying **(C.7)**, each function $g(\cdot)$ contained in some measurable class of functions $\mathscr{G}$, and for all $2<p<\infty$, there is a positive constant $c_{p,\beta}$ depending on p and the mixing coefficients $\{\beta_k\}_{k=1}^{\infty}$ such that

$$\mathbb{E}(\alpha_n(g))^2\le c_{p,\beta}\|g(Z)\|_p^2,$$

where $\{\alpha_n(g)\}_{\mathscr{G}}$ is an empirical process indexed by the class $\mathscr{G}$, i.e., for $g\in\mathscr{G}$,

$$\alpha_n(g)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(Z_i)-\mathbb{E}g(Z_i)).$$

They add that for each $g, g_{(1)}\in\mathscr{G}$,

$$\|\mathbb{G}(g)-\mathbb{G}(g_{(1)})\|_2\le c_{p,\beta}\|g(Z)-g_{(1)}(Z)\|_p, \tag{2.6}$$

where $\{\mathbb{G}(g)\}_{\mathscr{G}}$ is a Gaussian process indexed by $\mathscr{G}$. Arcones and Yu [6] assert also that if the class of functions $\mathscr{G}$ satisfies **(C.9)** and if **(C.7)** and (2.6) hold and with respect to Theorem 3.1 in [18], then the Gaussian process has a version with uniformly bounded and uniformly continuous paths with respect to the $\|\cdot\|_2$-norm. In the rest of this section, we assume that $\mathscr{F}$ is of VC-type, with characteristics A and v ("VC" for Vapnik and Červonenkis) meaning that

$$\text{(F.iii)}\qquad N(\epsilon,\mathscr{F},\|\cdot\|_{L_r(\mathbb{Q})})\le A\left(\frac{(\mathbb{Q}F^r)^{1/r}}{\epsilon}\right)^v,\qquad \epsilon>0,$$

for a given $1 \leq r < \infty$, such that $\mathbb{Q}F^r < \infty$, $\mathbb{Q}$ is any probability measure on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, and $\mathcal{B}(\mathbb{R}^m)$ represents the σ-field of Borel sets of $\mathbb{R}^m$. For instance, Examples 26 and 38 in [49], Lemma 22 in [47] § 4.7 in [19], Theorem 2.6.7 in [67], § 9.1 in [40] provide a number of sufficient conditions under which (F.iii) holds, we may also refer to § 3.2 in [13] for further discussions. For instance, it is satisfied, for general $d \geq 1$, whenever $g(x) = \phi(p(x))$, with $p(x)$ being a polynomial in d variables and $\phi(\cdot)$ being a real-valued function of bounded variation, we refer the reader to p. 1381 in [23]. If the class $\mathscr{F}$ is a VC-type class of functions then $\mathscr{F}$ satisfies (F.iii) with characteristics A and ν. Since it is generally recognized that the choice of the kernel is of less importance for the performance of a kernel estimator than the choice of bandwidth, we feel free to impose the following conditions on the kernel. Recall that $K(\cdot)$ is a kernel function with support contained in $[-B, B]$, $0 < B < \infty$, and satisfying

(K.ii) The class $\mathscr{K}$ is a VC-type class of functions.

Remark that condition (K.ii) is satisfied whenever $K(\cdot)$ is of bounded variation on $\mathbb{R}$. Notice that conditions (F.iii) and (K.ii) imply that the class of functions $\mathscr{F}\mathscr{K}$ is of VC-type, i.e., **(C.10)** is satisfied. Since we are dealing with a stationary absolute regular sequence $(Z_1, \ldots, Z_n)$, where $Z_i = (X_i, Y_i)$, the sequence of the independent blocks that we use in the following is given by $\{\xi_i = (\varsigma_i, \zeta_i)\}_{i\in\mathbb{N}^*}$.

Theorem 2.2.3. *Assume that conditions* **(C.1)–(C.8)** *hold. Suppose that the class $\mathscr{F}\mathscr{K}$ is of VC-type and for each $\varphi \in \mathscr{F}$, $\mathbb{E}(\varphi^2(Y_1)) < \infty$, then the process*

$$\left\{\nu_n(\psi \mid t) = \sqrt{nh}(\widehat{r}_n^{(1)}(\psi, t, h_n) - r^{(1)}(\psi, t)),\ \psi \in \mathscr{F}\mathscr{K},\ t \in \mathbb{I}\right\}$$

converges in law to a Gaussian process $\{G_n(\psi) : \psi \in \mathscr{F}\mathscr{K}\}$ that admits a version with uniformly bounded and uniformly continuous paths with respect to $\|\cdot\|_2$-norm.

Remark 2.2.4. If the VC type class of functions $\mathscr{F}\mathscr{K}$ is uniformly bounded, then to obtain the weak convergence of the process $\{\nu_n(\psi \mid t) = \sqrt{nh}(\widehat{r}_n^{(1)}(\psi, t, h_n) - r^{(1)}(\psi, t)), \psi \in \mathscr{F}\mathscr{K}, t \in \mathbb{I}\}$ we just need to assume that the mixing coefficient satisfies

$$\beta_k k^r \to 0 \quad \text{as} \quad k \to \infty$$

for some $r > 1$.

Remark 2.2.5. We denote by $\{\mathcal{M}(t)\colon t \geq 0\}$ a nonnegative continuous function increasing on $[0, \infty)$ and such that, for some $s > 2$, ultimately as $t \uparrow \infty$

$$(i)\ \ t^{-s}\mathcal{M}(t) \downarrow; \qquad (ii)\ \ t^{-1}\mathcal{M}(t) \uparrow .$$

For each $x \geq \mathcal{M}(0)$, we define $\mathcal{M}^{inv}(x) \geq 0$ by $\mathcal{M}(\mathcal{M}^{inv}(x)) = x$. We assume further that

$$\text{(C.8')} \qquad \sup_{t\in\mathbb{I}} \mathbb{E}(\mathcal{M}(F(Y)) \mid X = t) < \infty.$$

Assumption **(C.8)** on the class of functions $\mathscr{F}$ can be replaced by the general assumption (C.8') but this will add much extra complexity to the proofs. We will need also that the sequence $\{h_n\}_{n\geq 1}$ satisfies some appropriate conditions. For more details, we may refer to [22, 14 and 13].

3. CONDITIONAL U-PROCESSES

In this section, we shall establish weak convergence for conditional U-processes of β-mixing sequences. For a given strictly stationary β-mixing sequence $\{Z_i = (X_i, Y_i), i \in \mathbb{N}^*\}$ of random variables with $(X_i, Y_i) \in \mathbb{R}^2$, the conditional U-statistic based on $\{Z_i = (X_i, Y_i)\}_{\mathbb{N}^*}$ and the kernel $\varphi\widetilde{K}$ ($\widetilde{K}(\cdot)$ will be defined later) is given by

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \widehat{r}_n^{(m)}(\varphi, \mathbf{t}) = \frac{\sum_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) K\left(\frac{X_{i_1}-t_1}{h_n}\right) \cdots K\left(\frac{X_{i_m}-t_m}{h_n}\right)}{\sum_{(i_1,\ldots,i_m)\in I(m,n)} K\left(\frac{X_{i_1}-t_1}{h_n}\right) \cdots K\left(\frac{X_{i_m}-t_m}{h_n}\right)}, \quad (3.1)$$

where

$$I(m, n) = \{\mathbf{i} = (i_1, \ldots, i_m)\colon 0 \leq i_j \leq n \quad \text{and} \quad i_j \neq i_r \quad \text{if} \quad j \neq r\},$$

is the set of all m-tuples of different integers between 1 and n and $\{h = h_n\}_{n\geq 1}$ is a sequence of positive constants converging to zero at the rate $nh_n^m \to \infty$. The class of functions that we consider is $\mathscr{F}_m\mathscr{K}^m$ given in Section 2. We define the conditional U-process indexed by $\mathscr{F}_m\mathscr{K}^m$ by

$$\left\{U_n^{(m)}(\varphi, \mathbf{t}) := \sqrt{nh^m}(\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) - r^{(m)}(\varphi, \mathbf{t}))\right\}_{\mathscr{F}_m\mathscr{K}^m}, \tag{3.2}$$

for notational brevity, we suppose that the kernel $\varphi(\cdot)\widetilde{K}(\cdot)$ is symmetric ($\widetilde{K}(\cdot)$ will be defined later). To study (3.2) we introduce some slightly different definitions and notation, which however should not lead to a misunderstanding (some are borrowed from [6] and [16]). First, let $\pi_{k,m}f$ be a $\mathbb{P}$-canonical function (completely degenerate, or completely centered) defined for a (symmetric) measurable function $f\colon S^m \times S^m \to \mathbb{R}$ and $\mathbf{x}_k = (x_1, \ldots, x_k), \mathbf{y}_k = (y_1, \ldots, y_k) \in S^k$ by

$$\pi_{k,m}f(\mathbf{x}_k, \mathbf{y}_k) := (\delta_{(x_1,y_1)} - \mathbb{P}) \cdots (\delta_{(x_k,y_k)} - \mathbb{P})\mathbb{P}^{m-k}(f),$$

where for a measurable space $(S, \mathcal{S})$, and some measurable function f, the notation $\mathbb{P}f$ means $\mathbb{P}f = \int f\, d\mathbb{P}$, $\mathbb{P}^{m-k}(f)$ means that we are dealing with the integral related to the product measure composed of the probability measure $\mathbb{P}$, $(m-k)$ times. For $1 \leq i \leq k$, $\delta_{(x_i,y_i)}$ denotes the Dirac measure on (x_i, y_i). From now on,

$$\begin{aligned}
\mathbf{X} &:= (X_1, \ldots, X_m) \in \mathbb{R}^m, \quad \mathbf{Y} := (Y_1, \ldots, Y_m) \in \mathbb{R}^m,\\
\mathbf{X}_i &:= (X_{i_1}, \ldots, X_{i_m}), \quad \mathbf{Y}_i := (Y_{i_1}, \ldots, Y_{i_m}),\\
\widetilde{K}(\mathbf{t}) &:= \prod_{i=1}^m K(t_i), \quad \mathbf{t} = (t_1, \ldots, t_m),\\
K_h(x) &:= \frac{1}{h}K\left(\frac{x}{h}\right),\\
G_{\varphi,\mathbf{t}}(\mathbf{x}, \mathbf{y}) &:= \varphi(\mathbf{y})\widetilde{K}_h(\mathbf{x} - \mathbf{t}) = \frac{1}{h^m}\varphi(y_1, \ldots, y_m)\prod_{i=1}^m K\left(\frac{x_i - t_i}{h}\right) \quad \text{for } \mathbf{x} \in \mathbb{R}^m,\ \mathbf{y} \in \mathbb{R}^m,\ \mathbf{t} \in \mathbb{I}^m,\\
\mathcal{G} &:= \left\{G_{\varphi,\mathbf{t}}(\cdot,\cdot) = h^{-m}\varphi(\cdot)\widetilde{K}\left(\frac{\cdot - \mathbf{t}}{h}\right),\ \varphi \in \mathscr{F}_m,\ \widetilde{K} \in \mathscr{K}^m,\ \mathbf{t} = (t_1, \ldots, t_m)\right\},\\
\mathcal{G}^{(k)} &:= \{\pi_{k,m}G_{\varphi,\mathbf{t}}(\cdot,\cdot),\ \varphi \in \mathscr{F}_m,\ \mathbf{t} = (t_1, \ldots, t_m)\},\\
u_n(\varphi, \mathbf{t}) &= u_n^{(m)}(G_{\varphi,\mathbf{t}}) := \frac{(n-m)!}{n!}\sum_{i \in I(m,n)} G_{\varphi,\mathbf{t}}(\mathbf{X}_i, \mathbf{Y}_i),
\end{aligned}$$

and the U-statistic process

$$\mu_n(\varphi, \mathbf{t}) := \sqrt{nh^m}\left\{u_n(\varphi, \mathbf{t}) - \mathbb{E}(u_n(\varphi, \mathbf{t}))\right\}.$$

The main result of this section is summarized in the following theorem.

Theorem 3.0.1. *Let $\mathscr{F}_m\mathscr{K}^m$ be a measurable VC subgraph class of functions from $(\mathbb{R}^m, \mathbb{R}^m) \longrightarrow \mathbb{R}$ such that, for some $2 < p < \infty$,*

$$\mu_p^{(m)} := \sup_{\boldsymbol{t} \in \mathbb{R}^m} \mathbb{E}(F^p(\boldsymbol{Y}) \mid \boldsymbol{X} = \boldsymbol{t}) < \infty. \tag{3.3}$$

If the β coefficients of the mixing stationary sequence $\{Z_i = (X_i, Y_i)\}_{i\in\mathbb{N}^*}$ fulfill

$$\beta_k k^r \longrightarrow 0 \quad as \quad k \to \infty, \tag{3.4}$$

for some $r > 1$, then $\{U_n^{(m)}(\varphi, \boldsymbol{t})\}_{\mathscr{F}_m\mathscr{K}^m}$ converges in law to a Gaussian process $\{\mathbb{G}(\varphi)\}_{\mathscr{F}_m\mathscr{K}^m}$, which has a version with uniformly bounded and uniformly continuous paths with respect to $\|\cdot\|_2$-norm.

Remark 3.0.2. The index r in equation (3.4) denotes a constant $r > 1$ different from the r used in (F.iii).

Remark 3.0.3. If $G_{\varphi,\mathbf{t}}$ is not symmetric, we will need to symmetrize it. To do this we have:

$$\overline{G}_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y}) := \frac{1}{m!}\sum_{\sigma\in I_m^m} G_{\varphi,\mathbf{t}}(\mathbf{x}_\sigma,\mathbf{y}_\sigma) = \frac{1}{m!}\sum_{\sigma\in I_m^m} \varphi(\mathbf{y}_\sigma)\widetilde{K}(\mathbf{x}_\sigma - \mathbf{t}),$$

where $\mathbf{x}_\sigma = (x_{\sigma_1},\ldots,x_{\sigma_m})$ and $\mathbf{y}_\sigma = (y_{\sigma_1},\ldots,y_{\sigma_m})$. After symmetrization the expectation

$$\mathbb{E}\left(\overline{G}_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y})\right) = \mathbb{E}\left(G_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y})\right),$$

and the U-statistic $u_n^{(m)}(G_{\varphi,\mathbf{t}}) = u_n^{(m)}(\overline{G}_{\varphi,\mathbf{t}}) = u_n(\varphi,\mathbf{t})$ do not change, so the U-process could be redefined using the symmetric kernels $\overline{\varphi\widetilde{K}}$ as

$$\mu_n(\varphi,\mathbf{t}) = \sqrt{nh^m}\left\{u_n^{(m)}(\overline{G}_{\varphi,\mathbf{t}}) - \mathbb{E}(u_n^{(m)}(\overline{G}_{\varphi,\mathbf{t}}))\right\},$$

so Hoeffding's decomposition is:

$$\mu_n(\varphi,\mathbf{t}) = \sqrt{nh^m}\sum_{k=1}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}\overline{G}_{\varphi,\mathbf{t}}).$$

Note that if the class of functions $\mathscr{F}_m\mathscr{K}^m$ satisfies the entropy condition (7.22), the class $\overline{\mathscr{F}_m\mathscr{K}^m}$ of symmetrized functions satisfies it too with some characteristics $\overline{a}$, $\overline{b}$, and its envelope function is

$$\overline{F}(\mathbf{y}) \equiv \overline{F}(\mathbf{x},\mathbf{y}) = k^m\sum_{\sigma\in I_m^m} F(\mathbf{y}_\sigma).$$

Further, if the class of functions $\mathscr{F}_m\mathscr{K}^m$ is of VC-type, then so is also the class $\overline{\mathscr{F}_m\mathscr{K}^m}$.

4. STATISTICAL APPLICATION TO THE INDEPENDENCE TEST

In this section we present a statistical application of the theoretical results given in Section 3. To test the independence of one-dimensional random variables X and Y, Kendall [38] proposed a method based on the U-statistic K_n with the kernel function

$$\varphi\left((s_1,t_1),(s_2,t_2)\right) = \mathbf{1}_{\{(s_2-s_1)(t_2-t_1)>0\}} - \mathbf{1}_{\{(s_2-s_1)(t_2-t_1)\le 0\}}. \tag{4.1}$$

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

$$\widehat{r}_n^{(2)}(\varphi,\mathbf{t}) = \frac{\sum_{i\neq j}^n \varphi(Y_i,Y_j)K\left(\frac{t_1-X_i}{h_n}\right)K\left(\frac{t_2-X_j}{h_n}\right)}{\sum_{i\neq j}^n K\left(\frac{t_1-X_i}{h_n}\right)K\left(\frac{t_2-X_j}{h_n}\right)},$$

where $\mathbf{t} = (t_1,t_2) \in \mathbf{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (4.1). 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 $Y_1,\ldots,Y_n$ are observations of $(\boldsymbol{\xi},\boldsymbol{\eta})$. We are interested in testing

$$H_0\colon \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X \quad \text{vs} \quad H_a\colon H_0 \text{ is not true.} \tag{4.2}$$

Let $\mathbf{a} = (\mathbf{a}_1,\mathbf{a}_2) \in \mathbb{R}^d$ be such that $\|\mathbf{a}\| = 1$ and $\mathbf{a}_1 \in \mathbb{R}^{d_1}$, $\mathbf{a}_2 \in \mathbb{R}^{d_2}$, and let $F(\cdot)$, $G(\cdot)$ be the distribution functions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ respectively. Suppose that $F^{\mathbf{a}_1}(\cdot)$ and $G^{\mathbf{a}_2}(\cdot)$ are continuous for any unit vector $\mathbf{a} = (\mathbf{a}_1,\mathbf{a}_2)$, where $F^{\mathbf{a}_1}(t) = \mathbb{P}(\mathbf{a}_1^\tau\boldsymbol{\xi} < t)$ and $G^{\mathbf{a}_2}(t) = \mathbb{P}(\mathbf{a}_2^\tau\boldsymbol{\eta} < t)$ and $\mathbf{a}_i^\tau$ means the transpose of the vector $\mathbf{a}_i$, $1 \le i \le 2$. 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

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

As in [72], for $m = 2$ and the class of functions

$$\mathscr{F}_{\mathbf{a}}\mathscr{K}^2 = \left\{\varphi^{\mathbf{a}}(\cdot,\cdot)K\left(\frac{\cdot - t_1}{h}\right)K\left(\frac{\cdot - t_2}{h}\right), \mathbf{a} \in \mathbb{R}^d,\ \|\mathbf{a}\| = 1\right\},$$

it is easy to see that $\sup_{\|\mathbf{a}\|=1}\sqrt{nh^2}\widehat{r}_2^{(2)}(\varphi^{\mathbf{a}},\mathbf{t})$ could be used as a test statistic for (4.2). If the null hypothesis is true, then

$$D_n = \sup_{\|\mathbf{a}\|=1}\sqrt{nh^2}\,|\widehat{r}_2^{(2)}(\varphi^{\mathbf{a}},\mathbf{t})| = \sup_{\varphi^{\mathbf{a}}\times K\times K\in\mathscr{F}_{\mathbf{a}}\mathscr{K}^2}\sqrt{nh^2}\,|\widehat{r}_2^{(2)}(\varphi^{\mathbf{a}},\mathbf{t})|.$$

An application of Theorem 3.0.1 gives

$$D_n \to \sup_{\varphi^{\mathbf{a}}\times K\times K\in\mathscr{F}_{\mathbf{a}}\mathscr{K}^2}\sqrt{nh^2}\,|\widehat{r}_2^{(2)}(\varphi^{\mathbf{a}},\mathbf{t})|.$$

It will be of interest to give more details how to perform this statistical tests. We will not investigate this question in the present paper.

5. EXAMPLES

5.1. Examples of Classes of Functions

Example 5.1.1. The set $\mathscr{F}$ of all indicator functions $1\!\!1_{\{(-\infty,t]\}}$ of cells in $\mathbb{R}$ satisfies

$$N\big(\epsilon,\mathscr{F},d_{\mathbb{P}}^{(2)}\big) \le \frac{2}{\epsilon^2},$$

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

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

For more details and discussion on this example refer to Example 2.5.4 in [67] and p. 157 in [40]. 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 in [40].

Example 5.1.2. (*Classes of functions that are Lipschitz in a parameter*, Section 2.7.4 in [67]). Let $\mathscr{F}$ be the class of functions $x \mapsto \varphi(t,x)$ that are Lipschitz in the index parameter $t \in T$. Suppose that

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

for some metric d on the index set T and the function $\kappa(\cdot)$ defined on the sample space $\mathcal{X}$, and all x. According to Theorem 2.7.11 in [67] and Lemma 9.18 in [40], it follows, for any norm $\|\cdot\|_{\mathscr{F}}$ on $\mathscr{F}$, that

$$N\big(\epsilon\|F\|_{\mathscr{F}},\mathscr{F},\|\cdot\|_{\mathscr{F}}\big) \le N(\epsilon/2,T,d).$$

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 $\mathscr{F}$.

Example 5.1.3. Let us consider as example the classes of functions that are smooth up to order α defined as follows, see Section 2 in [66]. For $0 < \alpha < \infty$ let $\lfloor\alpha\rfloor$ be the greatest integer strictly smaller than α. For any vector $k = (k_1,\ldots,k_d)$ of d integers define the differential operator:

$$D^{k.} := \frac{\partial^{k.}}{\partial^{k_1}\ldots\partial^{k_d}},$$

where

$$k_. := \sum_{i=1}^d k_i.$$

Then, for a function $\varphi : \mathcal{X} \to \mathbb{R}$, let

$$\|\varphi\|_\alpha := \max_{k.\le\lfloor\alpha\rfloor}\sup_x |D^k\varphi(x)| + \max_{k.=\lfloor\alpha\rfloor}\sup_{x,y}\frac{D^k\varphi(x) - D^k\varphi(y)}{\|x-y\|^{\alpha-\lfloor\alpha\rfloor}},$$

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 $\varphi\colon \mathcal{X} \to \mathbb{R}$ with

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

Note that for $\alpha \leq 1$ this class consists of bounded functions φ that satisfy a Lipschitz condition. Kolmogorov and Tihomirov [38] computed the entropy of the classes of $C_M^\alpha(\mathcal{X})$ for the uniform norm. As a consequence of their results, van der Vaart [66] 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 $\epsilon > 0$

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

$N_{[\,]}$ is the bracketing number, refer to Definition 2.1.6 in [67] and we refer to Theorem 2.7.1 in [67] for a variant of the last inequality. By Lemma 9.18 in [40], we have

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

5.2. Examples of U-Statistics

Generally speaking, we may take for $\varphi(\cdot)$ any function which has been found interesting in the unconditional setup; cf. [57]. As was mentioned before, the case $m = 1$ leads to the Nadaraya–Watson estimator if we set $\varphi = Id$, the identity function; $\varphi = \mathbf{1}\{\cdot \leq t\}$ yields the conditional d.f. evaluated at t; [62]. We now discuss several examples for $m = 2$. We suppose that (X_1, Y_1) and (X_2, Y_2) are independent.

Example 5.2.1. Let $\varphi(Y_1, Y_2) = Y_1Y_2$, then

$$\begin{aligned} r^{(2)}(\varphi(Y_1, Y_2) \mid t_1, t_2) &= \mathbb{E}(Y_1Y_2 \mid X_1 = t_1, X_2 = t_2) \\ &= \mathbb{E}(Y_1 \mid X_1 = t_1)\mathbb{E}(Y_2 \mid X_2 = t_2) \\ &= \overline{r}^{(2)}(t_1)\overline{r}^{(2)}(t_2), \end{aligned}$$

with $\overline{r}^{(2)}$ denoting the regression of $\mathbf{Y}$ on $\mathbf{X} = \mathbf{t}$. The above $\varphi(\cdot)$ is a simple example of kernel for a conditional U-statistic, where one is interested in functions of $\overline{r}^{(2)}$.

Example 5.2.2. For

$$\varphi(Y_1, Y_2) = \frac{1}{2}(Y_1 - Y_2)^2$$

we obtain

$$r^{(2)}(\varphi(Y_1, Y_2) \mid t_1, t_2) = \mathrm{Var}(Y_1 \mid X_1 = t_1).$$

Example 5.2.3. For $\varphi(Y_1, Y_2) = \mathbf{1}\{Y_1 + Y_2 > 0\}$, we obtain a conditional U-statistic which may be viewed as a conditional version of the Wilcoxon one-sample statistic. It may be used for testing the hypothesis that the conditional distribution at X_1 is symmetric at zero. Obviously:

$$r^{(2)}(\varphi(Y_1, Y_2) \mid t_1 = t_2) = \mathbb{P}(Y_1 + Y_2 > 0 \mid X_1 = t_1 = X_2).$$

Example 5.2.4. For $\varphi(Y_1, Y_2) = \mathbf{1}\{Y_1 \leq Y_2\}$,

$$r^{(2)}(\varphi(Y_1, Y_2) \mid t_1, t_2) = \mathbb{P}(Y_1 \leq Y_2 \mid X_1 = t_1, X_2 = t_2) \quad \text{for} \quad t_1 \neq t_2$$

equals the probability that the output pertaining to t_1 is less than or equal to the one pertaining to $\mathbf{t}_2$.

Example 5.2.5. Assume $\{Y_i = (Y_{i,1}, Y_{i,2})^\top\}_{i=1,2}$ and define φ by

$$\varphi(y_1, y_2) := \frac{1}{2}(y_{1,1}y_{1,2} + y_{2,1}y_{2,2} - y_{1,1}y_{2,2} - y_{1,2}y_{2,1}),$$

and

$$\begin{aligned} r^{(2)}(\varphi(Y_1, Y_2) \mid t_1, t_2) \;=\; & \frac{1}{2}\{\mathbb{E}(Y_{1,1}Y_{1,2} \mid X_1 = t_1) + \mathbb{E}(Y_{2,1}Y_{2,2} \mid X_2 = t_2) \\ & - \mathbb{E}(Y_{1,1}Y_{2,2} \mid X_1 = t_1, X_2 = t_2) - \mathbb{E}(Y_{1,2}Y_{2,1} \mid X_1 = t_1, X_2 = t_2)\}. \end{aligned}$$

In particular,

$$r^{(2)}(\varphi(Y_1, Y_2) \mid t_1) = \mathbb{E}(Y_{1,1}Y_{1,2} \mid X_1 = t_1) - \mathbb{E}(Y_{1,1} \mid X_1 = t_1)\mathbb{E}(Y_{1,2} \mid X_1 = t_1)$$

is the conditional covariance of Y_1 given $X_1 = t_1$.

Example 5.2.6. For m=3, let

$$\varphi(Y_1, Y_2, Y_3) = \mathbf{1}\{Y_1 - Y_2 - Y_3 > 0\},$$

We have

$$r^{(3)}(\varphi(Y_1, Y_2, Y_3) \mid t_1 = t_2 = t_3 = t) = \mathbb{P}(Y_1 > Y_2 + Y_3 \mid X_1 = X_2 = X_3 = t)$$

and the corresponding conditional unbiased statistic can be looked upon as a conditional analogue of the Hollander–Proschan test-statistic [34]. 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 5.2.7. Let

$$\psi(Y_1, Y_2, Y_3) = \mathbf{1}\{Y_2 \leq Y_1\} - \mathbf{1}\{Y_3 \leq Y_1\}$$

and for $m = 5$ define

$$\varphi(Y_1, \ldots, Y_5) = \frac{1}{4}\psi(Y_1, Y_2, Y_3)\psi(Y_1, Y_4, Y_5) \times \psi(Y_1, Y_2, Y_3)\psi(Y_1, Y_4, Y_5).$$

We have

$$r^{(5)}\big(\varphi(Y_1, \ldots, Y_5) \mid t_1 = t_2 = t_3 = t_4 = t_5 = t\big) = \mathbb{E}\big(\varphi(Y_1, \ldots, Y_5) \mid X_1 = X_2 = X_3 = X_4 = X_5 = t\big).$$

The corresponding U-statistics may be used to test the conditional independence.

Example 5.2.8. Let $\widehat{Y_1Y_2}$ 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

$$\varphi_t(Y_1, Y_2) = \mathbf{1}\{\widehat{Y_1Y_2} \leq t\} - t/\pi, \quad \text{for} \quad t \in [0, \pi).$$

Silverman [60] has used this kernel in order to propose a U-process to test uniformity on the circle. Let

$$r^{(2)}(\varphi_t(Y_1, Y_2) \mid t_1 = t_2 = t) = \mathbb{E}(\varphi_t(Y_1, Y_2) \mid X_1 = X_2 = t).$$

In this setting, one can propose conditional U-process to test conditional uniformity on the circle.

Example 5.2.9. Let $\{\mathbf{Z}_1 = (X_1, Y_1), \mathbf{Z}_2 = (X_2, Y_2)\}$ be two random variables. If we want to test the symmetry about zero of the conditional distribution at $X_1 = X_2 = t_1$, we often use the so-called Wilcoxon one-sample statistic. In this case the U-statistic

$$\widehat{r}_2^{(2)}(\varphi(Y_1, Y_2) \mid t_1, t_2) = \frac{1}{2}\frac{\mathbf{1}_{\{Y_1+Y_2>0\}}K\left(\frac{X_1-t_1}{h}\right)K\left(\frac{X_2-t_1}{h}\right)}{K\left(\frac{X_1-t_1}{h}\right)K\left(\frac{X_2-t_1}{h}\right)}$$

may be viewed as a conditional version of the Wilcoxon one-sample statistic with kernel

$$\mathbf{1}_{\{Y_1+Y_2>0\}}K\left(\frac{X_1 - t_1}{h}\right)K\left(\frac{X_2 - t_1}{h}\right)$$

for

$$\varphi(y_1, y_2) = \mathbf{1}_{\{y_1+y_2>0\}}.$$

6. CONCLUDING REMARKS

In the present work we have considered the weak convergence to Gaussian processes of the conditional empirical processes and the conditional U-processes from stationary β-mixing sequences indexed by classes of functions satisfying some entropy conditions. In particular, we have extended the results of [6] to the conditional U-processes. We have treated the weak convergence in both cases when the class of functions is bounded or unbounded satisfying some moment conditions. The unbounded case remains still open until present. The fact that the limits in our theorems depend on the unknown parameters makes it important that good approximations of these limiting distributions be found and that is where the bootstrap proved to be a very effective tool. It would be of interest to provide a complete investigation of the approximation of the bootstrapped conditional empirical and conditional U-processes based upon stationary mixing sequences which requires nontrivial mathematics, that goes well beyond the scope of the present paper, we leave this problem open for future. A natural question is how to relax the dependence assumption on the sequence of random variables to cope with more general framework by considering weak dependence or by assuming only ergodicity.

7. MATHEMATICAL DEVELOPMENTS

This section is devoted to the proofs of our results. The aforementioned notation is also used in what follows. First, we introduce the following easy lemma that will turn out to be useful later on.

Lemma 7.0.1. *Let $\{\omega_{nk}, 1 \le k \le k_n\}$ be a triangular array such that following conditions are fulfilled*

$$\mathbb{E}(\omega_{nk}) = 0, \quad 1 \le k \le k_n, \qquad \lim_{n\to\infty} \sum_{k=1}^{k_n} \mathbb{E}(\omega_{nk}^2) = 1,$$

and

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \mathbb{E}[\omega_{nk}^2 \mathbf{1}_{\{|\omega_{nk}|>\epsilon\}}] = 0, \quad \forall \epsilon > 0. \tag{7.1}$$

Then

$$\lim_{n\to\infty} k_n \max_{1\le j\le k_n} \beta_j = 0 \tag{7.2}$$

implies that

$$S_n = \sum_{k=1}^{k_n} \omega_{nk} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof of Lemma 7.0.1

This lemma is a particular case of Theorem 5.3 in [20]. It can be proved with the help of Corollary 2.3 and Lemma 5.3 in [20]. Lemma 5.3 in [20] asserts that for some random variable (r.v) $\boldsymbol{\xi}$ (complex) satisfying $|\boldsymbol{\xi}| < 1$, we have

$$\mathbb{E}\left[\mathbb{E}(\boldsymbol{\xi} \mid \mathcal{C}) - \mathbb{E}(\boldsymbol{\xi})\right] \le 2\pi\alpha(\mathcal{B}(\boldsymbol{\xi}), \mathcal{C}),$$

where $\mathcal{B}(\boldsymbol{\xi})$ is the σ-algebra generated by $\boldsymbol{\xi}$ and $\mathcal{C}$ is some σ-algebra in the same probability space. It is known that $2\alpha(\mathcal{A},\mathcal{B}) \le \beta(\mathcal{A},\mathcal{B})$ for any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, hence

$$\mathbb{E}\big[\mathbb{E}(\boldsymbol{\xi} \mid \mathcal{C}) - \mathbb{E}(\boldsymbol{\xi})\big] \le \pi\beta(\mathcal{B}(\boldsymbol{\xi}), \mathcal{C}).$$

Making use of this result, we obtain for the triangular array $\{\omega_{ni}, 1 \le i \le k_n\}$ that

$$\mathbb{E}\left[\mathbb{E}(\exp\{it\omega_{nk}\} \mid \mathscr{B}_{n,k-1}) - \mathbb{E}(\exp\{it\omega_{nk}\})\right] \le \pi\beta(\mathscr{B}_{n,k}, \mathscr{B}_{n,k-1}),$$

where $\mathscr{B}_{n,k} = \sigma(\omega_{nk})$. Consequently, we obtain

$$\sum_{k=2}^{k_n} \mathbb{E}\left[\mathbb{E}(\exp\{it\omega_{nk}\} \mid \mathscr{B}_{n,k-1}) - \mathbb{E}(\exp\{it\omega_{nk}\})\right] \leq \pi k_n \max_{1\leq j\leq k_n} \beta_j.$$

Combining (7.1) and (7.2) and applying Corollary 2.3 in [20], we then get

$$\sum_{k=2}^{k_n} \mathbb{E}\left[\mathbb{E}(\exp\{it\omega_{nk}\} \mid \mathscr{B}_{n,k-1}) - \mathbb{E}(\exp\{it\omega_{nk}\})\right] \leq \pi k_n \max_{1\leq j\leq k_n} \beta_j \overset{n\to\infty}{\longrightarrow} 0,$$

and it follows that

$$S_n \xrightarrow{d} N(0,1).$$

Hence the proof is complete. □

Proof of Theorem 2.1.1

The Cramér–Wold device states, indirectly, that the convergence of finite-dimensional distributions can be obtained from the convergence of one-dimensional distributions. Therefore we will just prove the one-dimensional convergence. Notice that we have

$$\begin{aligned}
\nu_n(\varphi \mid t) &= \sqrt{nh}(\widehat{r}_n^{(1)}(\varphi,t) - r^{(1)}(\varphi,t)) \\
&= \sqrt{nh}\left(\frac{\frac{1}{n}\sum_{i=1}^n \varphi(Y_i)K_h(X_i-t)}{\frac{1}{n}\sum_{i=1}^n K_h(X_i-t)} - r^{(1)}(\varphi,t)\right) \\
&= \frac{\sum_{i=1}^n \sqrt{\frac{h}{n}}(f(Y_i) - r^{(1)}(\varphi,t))K_h(X_i-t)}{\frac{1}{n}\sum_{i=1}^n K_h(X_i-t)} \\
&= \frac{\sum_{i=1}^n(\nu_{ni} - \mathbb{E}\nu_{ni} + \mathbb{E}\nu_{ni})}{\rho_n(t)} \\
&= \frac{S_n}{\rho_n(t)} + \frac{\sum_{i=1}^n \mathbb{E}\nu_{ni}}{\rho_n(t)} \\
&= \frac{S_n}{\rho_n(t)} + nh^{5/2}\mu_n',
\end{aligned}$$

where the following is used

$$S_n = \sum_{i=1}^n(\nu_{ni} - \mathbb{E}\nu_{ni}), \quad \rho_n(t) = \frac{1}{n}\sum_{i=1}^n K_h(X_i-t) \quad \text{and} \quad nh^{5/2}\mu_n' = \frac{\sum_{i=1}^n \mathbb{E}\nu_{ni}}{\rho_n(t)}.$$

It is known that the Akaike–Parzen–Rosenblatt [2, 48, 54] kernel density estimators $\rho_n(t)$ are consistent estimators of $\rho(t)$, that is

$$\rho_n(t) \xrightarrow{\mathbb{P}} \rho(t).$$

Making use of the Cramér–Slutsky lemma, it suffices to show that

$$S_n \xrightarrow{d} N(0,\sigma^2(\varphi)), \quad \text{where} \quad \sigma^2(f) = \left(r^{(1)}(\varphi^2,t) - (r^{(1)}(\varphi,t))^2\right)\rho(t)\int_{\mathbb{R}} K^2(u)\,du,$$

and

$$nh^{5/2}\mu_n' \to 0.$$

Let us begin by showing that $nh^{5/2}\mu_n' \to 0$. We have

$$\frac{1}{nh^{5/2}}\sum_{i=1}^n \mathbb{E}\nu_{ni} = \frac{1}{nh^{5/2}}\sum_{i=1}^n \mathbb{E}\left\{\sqrt{\frac{h}{n}}\breve{\varphi}(Y_i)K_h(X_i-t)\right\},$$

$$= \frac{1}{\sqrt{nh^6}}\mathbb{E}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)K\Big(\frac{X_1-t}{h}\Big)\Big\}.$$

An application of Lemma 7.0.5 in combination with Proposition 3.4 in [73] implies that, as $n \to \infty$,

$$\begin{aligned}&\frac{1}{\sqrt{nh^6}}\mathbb{E}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)K\Big(\frac{X_1-t}{h}\Big)\Big\}\\ &\quad= \frac{1}{\sqrt{nh^6}}\int_{\mathbb{R}} \mathbb{E}\big(\varphi(Y_1) - r^{(1)}(\varphi,t) \mid X = z\big)K\Big(\frac{z-t}{h}\Big)\rho(z)\,dz,\\ &\quad= \frac{1}{\sqrt{nh^6}}\int_{\mathbb{R}} \big(r^{(1)}(\varphi,z) - r^{(1)}(\varphi,t)\big)K\Big(\frac{z-t}{h}\Big)\rho(z)\,dz \to 0.\end{aligned}$$

Hence we have $\mu_n' \to 0$ as $n \to \infty$. This when combined with condition **(C.3)** implies that

$$nh^{5/2}\mu_n' \to 0.$$

We shall now prove that

$$S_n \xrightarrow{d} N(0,\sigma^2(\varphi)).$$

We start by calculating $n\mathrm{Var}(\nu_{n1})$. We get

$$\begin{aligned}n\mathrm{Var}(\nu_{n1}) &= \frac{1}{h}\mathrm{Var}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)K\Big(\frac{X_1-t}{h}\Big)\Big\}\\ &= \frac{1}{h}\mathbb{E}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)^2K^2\Big(\frac{X_1-t}{h}\Big)\Big\}\\ &\quad -h\Big\{\frac{1}{h}\mathbb{E}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)K\Big(\frac{X_1-t}{h}\Big)\Big\}\Big\}^2.\end{aligned}$$

Notice that we have, as $n \to \infty$,

$$\begin{aligned}&\frac{1}{h}\mathbb{E}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)K\Big(\frac{X_1-t}{h}\Big)\Big\}\\ &\quad= \frac{1}{h}\int_{\mathbb{R}} \big(r^{(1)}(\varphi,z) - r^{(1)}(\varphi,t)\big)K\Big(\frac{z-t}{h}\Big)\rho(z)\,dz \to 0. \qquad (7.3)\end{aligned}$$

We apply again Lemma 7.0.5 to infer that, as $n \to \infty$,

$$\begin{aligned}&\frac{1}{h}\mathbb{E}\Big\{\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)^2K^2\Big(\frac{X_1-t}{h}\Big)\Big\}\\ &\quad= \frac{1}{h}\int \mathbb{E}\Big(\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)^2 \mid X = z\Big)K^2\Big(\frac{z-t}{h}\Big)\rho(z)\,dz\\ &\quad= \frac{1}{h}\int r^{(1)}\Big(\big(\varphi(Y_1) - r^{(1)}(\varphi,t)\big)^2, z\Big)K^2\Big(\frac{z-t}{h}\Big)\rho(z)\,dz\\ &\quad\to r^{(1)}\Big(\big(\varphi - r^{(1)}(\varphi,t)\big)^2, t\Big)\rho(t)\int K^2(z)\,dz. \qquad (7.4)\end{aligned}$$

By combining equations (7.3), (7.4), we readily obtain

$$n\mathrm{Var}(\nu_{n1}) \overset{n\to\infty}{\longrightarrow} r^{(1)}\Big(\big(f - r^{(1)}(\varphi,t)\big)^2, t\Big)\rho(t)\int K^2(z)\,dz.$$

Recall that our goal is to calculate $\mathrm{Var}(S_n)$. By using the stationarity of $\{\nu_{ni}, 1 \le i \le n\}$, we have

$$\mathrm{Var}(S_n) = \mathbb{E}S_n^2 = n\mathrm{Var}(\nu_{n1}) + 2\sum_{j=2}^{n}(n-j+1)\mathrm{Cov}(\nu_{n1},\nu_{nj}).$$

It is straightforward to see that

$$\begin{aligned} n\mathrm{Cov}(\nu_{n1},\nu_{nj}) &= n\mathrm{Cov}\Big(\sqrt{\frac{h}{n}}\breve{\varphi}(Y_1)K_h(X_1-t),\sqrt{\frac{h}{n}}\breve{\varphi}(Y_j)K_h(X_j-t)\Big)\\ &= h\mathrm{Cov}\big(\breve{\varphi}(Y_1)K_h(X_1-t),\breve{\varphi}(Y_j)K_h(_j-t)\big)\\ &= h\Big(\mathbb{E}\big(\breve{\varphi}(Y_1)K_h(X_1-t)\breve{\varphi}(Y_j)K_h(X_j-t)\big)\\ &\qquad -\mathbb{E}\big(\breve{\varphi}(Y_1)K_h(X_1-t)\big)\mathbb{E}\big(\breve{\varphi}(Y_j)K_h(X_j-t)\big)\Big). \end{aligned} \tag{7.5}$$

We have first to calculate

$$\begin{aligned} &\mathbb{E}(\breve{\varphi}(Y_1)K_h(X_1-t)\breve{\varphi}(Y_j)K_h(X_j-t))\\ &\quad=\int_{\mathbb{R}^2}\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid X_1=z_1,X_j=z_j\big)K_h(z_1-t)K_h(z_j-t)\rho_j(z_1,z_j)\,dz_1\,dz_j. \end{aligned} \tag{7.6}$$

A similar calculus yields:

$$\begin{aligned} \mathbb{E}\big(\breve{\varphi}(Y_1)K_h(X_1-t)\big) &= \int_{\mathbb{R}}\mathbb{E}\big(\breve{\varphi}(Y_1)\mid X_1=z_1\big)K_h(z_1-t)\rho(z_1)dz_1\\ &= \int_{\mathbb{R}} r^{(1)}\big(\breve{\varphi},z_1\big)K_h(z_1-t)\rho(z_1)dz_1. \end{aligned} \tag{7.7}$$

In the same way we calculate $\mathbb{E}\big(\breve{\varphi}(Y_j)K_h(X_j-t)\big)$. Thus, by combining equations (7.5), (7.6), and (7.7), we infer that

$$\begin{aligned} n\mathrm{Cov}(\nu_{n1},\nu_{nj}) &= h\int_{\mathbb{R}^2}\Big\{\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid X_1=z_1,X_j=z_j\big)\rho_j(z_1,z_j)\\ &\qquad -r^{(1)}(\breve{\varphi},z_1)r^{(1)}(\breve{\varphi},z_j)\rho(z_1)\rho(z_j)\Big\}K_h(z_1-t)K_h(z_j-t)\,dz_1\,dz_j\\ &= h\int_{\mathbb{R}^2}\breve{m}(z_1,z_j)K_h(z_1-t)K_h(z_j-t)\,dz_1\,dz_j, \end{aligned}$$

where we use the notation

$$\begin{aligned} &\breve{m}(z_1,z_j)=\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j\big)\rho_j(z_1,z_j)-r^{(1)}(\breve{\varphi},z_1)r^{(1)}(\breve{\varphi},z_j)\rho(z_1)\rho(z_j),\\ &\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j\big)=\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid X_1=z_1,X_j=z_j\big). \end{aligned}$$

Notice that by adding and subtracting the quantity $\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j\big)\rho(z_1)\rho(z_j)$ we get

$$\begin{aligned} \breve{m}(z_1,z_j) &\le \big|\rho_j(z_1,z_j)-\rho(z_1)\rho(z_j)\big|\big|\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j)\big|\\ &\qquad+\rho(z_1)\rho(z_j)\big|\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j)-r^{(1)}(\breve{\varphi},z_1)r^{(1)}(\breve{\varphi},z_j)\big|. \end{aligned}$$

Let us recall that

$$\mathbb{E}(Y\mid X=t)=\lim_{\epsilon\to 0}\frac{\mathbb{E}\big(Y\mathbf{1}_{\{X\in V_\epsilon(t)\}}\big)}{\mathbb{P}\big(X\in V_\epsilon(t)\big)},$$

where $V_\epsilon(t)$ is the ϵ-neighborhood of t. We have

$$\begin{aligned} &\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j)-r^{(1)}(\breve{\varphi},z_1)r^{(1)}(\breve{\varphi},z_j)\\ &=\lim_{\epsilon\to 0}\frac{\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mathbf{1}_{\{X_1\in V_\epsilon(z_1),X_j\in V_\epsilon(z_j)\}}\big)}{\mathbb{P}\big(X_1\in V_\epsilon(z_1),X_j\in V_\epsilon(z_j)\big)}-\mathbb{E}(\breve{\varphi}(Y_1)\mid X_1=z_1)\mathbb{E}(\breve{\varphi}(Y_j)\mid X_j=z_j)\\ &=\lim_{\epsilon\to 0}\frac{\mathbb{E}\big(\breve{\varphi}(Y_1)\mathbf{1}_{\{X_1\in V_\epsilon(z_1)\}}\breve{\varphi}(Y_j)\mathbf{1}_{\{X_j\in V_\epsilon(z_j)\}}\big)}{\mathbb{P}(X_1\in V_\epsilon(z_1))\mathbb{P}(X_j\in V_\epsilon(z_j))}-\lim_{\epsilon\to 0}\frac{\mathbb{E}\big(\breve{\varphi}(Y_1)\mathbf{1}_{\{X_1\in V_\epsilon(z_1)\}}\big)\mathbb{E}\big(\breve{\varphi}(Y_j)\mathbf{1}_{\{X_j\in V_\epsilon(z_j)\}}\big)}{\mathbb{P}(X_1\in V_\epsilon(z_1))\mathbb{P}(X_j\in V_\epsilon(z_j))}\\ &\quad-\lim_{\epsilon\to 0}\frac{\mathbb{E}\big(\breve{\varphi}(Y_1)\mathbf{1}_{\{X_1\in V_\epsilon(z_1)\}}\breve{\varphi}(Y_j)\mathbf{1}_{\{X_j\in V_\epsilon(z_j)\}}\big)}{\mathbb{P}(X_1\in V_\epsilon(z_1))\mathbb{P}(X_j\in V_\epsilon(z_j))}+\lim_{\epsilon\to 0}\frac{\mathbb{E}\big(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mathbf{1}_{\{X_1\in V_\epsilon(z_1),X_j\in V_\epsilon(z_j)\}}\big)}{\mathbb{P}(X_1\in V_\epsilon(z_1),X_j\in V_\epsilon(z_j))} \end{aligned}$$

$$= \lim_{\epsilon\to 0} \mathrm{Cov}\left(\frac{\breve{\varphi}(Y_1)\mathbf{1}_{\{X_1\in V_\epsilon(z_1)\}}}{\mathbb{P}(X_1\in V_\epsilon(z_1))}\frac{\breve{\varphi}(Y_j)\mathbf{1}_{\{X_j\in V_\epsilon(z_j)\}}}{\mathbb{P}(X_j\in V_\epsilon(z_j))}\right)$$
$$+\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1 z_j)\left(1-\frac{\mathbb{P}(X_1\in V_\epsilon(z_1), X_j\in V_\epsilon(z_j))}{\mathbb{P}(X_1\in V_\epsilon(z_1))\mathbb{P}(X_j\in V_\epsilon(z_j))}\right).$$

The integrability of $\rho(\cdot)$ and $\rho_j(\cdot)$ implies, for almost all $(z_1, z_j)\in\mathbb{R}^2$, that

$$\lim_{\epsilon\to 0}\frac{\mathbb{P}\left(X_1\in V_\epsilon(z_1), X_j\in V_\epsilon(z_j)\right)}{\mathbb{P}(X_1\in V_\epsilon(z_1))\mathbb{P}(X_j\in V_\epsilon(z_j))}=\frac{\rho_j(z_1,z_j)}{\rho(z_1)\rho(z_j)}.$$

This when combined with Davydov's inequality, implies readily

$$\lim_{\epsilon\to 0}\left|\mathrm{Cov}\left(\frac{\breve{\varphi}(Y_1)\mathbf{1}_{\{X_1\in V_\epsilon(z_1)\}}}{\mathbb{P}(X_1\in V_\epsilon(z_1))},\ \frac{\breve{\varphi}(Y_j)\mathbf{1}_{\{X_j\in V_\epsilon(z_j)\}}}{\mathbb{P}(X_j\in V_\epsilon(z_j))}\right)\right|\le 2p\beta_j^{1/p}r^{(1)}(\|\breve{\varphi}\|_q, z_1)r^{(1)}(\|\breve{\varphi}\|_s, z_j).$$

We could choose

$$\frac{1}{q}=\frac{1}{s}=\frac{1}{p}=\frac{1}{3}.$$

By using Lemma 3.1 in [51] and conditions **(C.1)–(C.6)**, it follows, as $n\to\infty$,

$$\begin{aligned}\mathrm{Cov}(S_{n1,nj})\ \le\ &\sup_{j>1}\left(\int_{\mathbb{R}^2}|\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j)|\rho(z_1)\rho(z_j)K_h(z_1-t)K_h(z_j-t)\,dz_1\,dz_j\right)h\sum_{j=2}^{n}\theta(j)\\ &+h\sum_{j=2}^{n}2p\beta_j^{1/p}\int_{\mathbb{R}^2}r^{(1)}(\|\breve{\varphi}\|_q,z_1)r^{(1)}(\|\breve{\varphi}\|_r,z_j)\rho(z_1)\rho(z_j)K_h(z_1-t)K_h(z_j-t)\,dz_1\,dz_j\\ &+h\sum_{j=2}^{n}\int_{\mathbb{R}^2}|\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j)|\rho_j(z_1,z_j)K_h(z_1-t)K_h(z_j-t)\,dz_1\,dz_j\\ &+h\sup_{j>1}\int_{\mathbb{R}^2}|\mathbb{E}(\breve{\varphi}(Y_1)\breve{\varphi}(Y_j)\mid z_1,z_j)|\rho(z_1)\rho(z_j)K_h(z_1-t)K_h(z_j-t)\,dz_1\,dz_j\to 0,\end{aligned}$$

where

$$\mathrm{Cov}(S_{n1,nj})=\sum_{j=2}^{n}(n-j+1)\mathrm{Cov}(\nu_{n1},\nu_{nj}).$$

Thus

$$\mathrm{Var}(S_n)\overset{n\to\infty}{\longrightarrow} r^{(1)}\big((\varphi-r^{(1)}(\varphi,t))^2,t\big)\rho(t)\int K^2(\mathbf{z})\,d\mathbf{z}.$$

Finally, all what remains is to check condition (7.1), i.e., for all $\epsilon>0$,

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\big\{(\nu_{ni}-\mathbb{E}(\nu_{ni}))^2\mathbf{1}_{\{|\nu_{ni}-\mathbb{E}(\nu_{ni})|>\epsilon\}}\big\}=0. \tag{7.8}$$

But the event $(\nu_{ni}-\mathbb{E}(\nu_{ni}))^2\mathbf{1}_{\{|\nu_{ni}-\mathbb{E}(\nu_{ni})|>\epsilon\}}$ only makes sense when $|\nu_{ni}-\mathbb{E}(\nu_{ni})|>\epsilon$, so proving (7.8) is to prove that

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{P}\{|\nu_{ni}-\mathbb{E}(\nu_{ni})|>\epsilon\}=0.$$

We have, by applying Markov's inequality:

$$\sum_{i=1}^{n}\mathbb{P}\{|\nu_{ni}-\mathbb{E}(\nu_{ni})|>\epsilon\}\le\frac{\sum_{i=1}^{n}\mathbb{E}\{|\nu_{ni}-\mathbb{E}(\nu_{ni})|\}}{\epsilon^2}$$

$$\leq \frac{2n\mathbb{E}|\nu_{n1}|}{\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0\,.$$

Therefore the proof is complete. □

Proof of Theorem 2.2.2

To study the asymptotic equicontinuity of the conditional empirical process

$$\big\{\nu_n(\psi \mid t) = \sqrt{nh}(\widehat{r}_n^{(1)}(\psi, t, h_n) - r^{(1)}(\psi, t),\ \psi \in \mathscr{F}\mathscr{K},\ t \in \mathbb{I}\big\},$$

we decompose it so that to get a sum of simple empirical processes. Towards this end, we introduce the following process: for any $\varphi K \in \mathscr{F}\mathscr{K}$ and $t \in \mathbb{I}$,

$$\mathcal{W}_n(t,\varphi) = \sum_{i=1}^{n} \varphi(Y_i) K\bigg(\frac{X_i - t}{h_n}\bigg) - n\mathbb{E}\bigg\{\varphi(Y_1)K\bigg(\frac{X_1 - t}{h_n}\bigg)\bigg\}. \tag{7.9}$$

We first decompose

$$\nu_n(\varphi \mid t) = \sqrt{nh}(\widehat{r}_n^{(1)}(\varphi, t) - r^{(1)}(\varphi, t)) := \sqrt{nh}(\widehat{r}_n^{(1)}(\psi, t, h_n) - r^{(1)}(\psi, t)) = \nu_n(\psi \mid t)$$

to

$$\begin{aligned}
\nu_n(\varphi \mid t) &= \sqrt{nh}\big(\widehat{r}_n^{(1)}(\varphi, t) - r^{(1)}(\varphi, t)\big) \\
&= \sqrt{nh}\bigg(\frac{\sum_{i=1}^n \varphi(Y_i)K(\frac{X_i - t}{h_n})}{\sum_{i=1}^n K(\frac{X_i - t}{h_n})} - \mathbb{E}(\varphi(Y) \mid X = t)\bigg) \\
&= \sqrt{nh}\bigg(\frac{\gamma_n(t)}{\rho_n(t)} - \mathbb{E}(\varphi(Y) \mid X = t)\bigg) \\
&= \sqrt{nh}\bigg(\frac{\gamma_n(t)}{\rho_n(t)} - \frac{\mathbb{E}(\gamma_n(t))}{\mathbb{E}(\rho_n(t))}\bigg) - \sqrt{nh}\bigg(\mathbb{E}(\varphi(Y) \mid X = t) - \frac{\mathbb{E}(\gamma_n(t))}{\mathbb{E}(\rho_n(t))}\bigg) \\
&= \frac{1}{\rho_n(t)}\sqrt{nh}\big(\gamma_n(t) - \mathbb{E}(\gamma_n(t))\big) - \frac{\mathbb{E}(\gamma_n(t))}{\rho_n(t)\mathbb{E}(\rho_n(t))}\sqrt{nh}\big(\rho_n(t) - \mathbb{E}(\rho_n(t))\big) \\
&\quad -\sqrt{nh}\bigg(\mathbb{E}(\varphi(Y) \mid X = t) - \frac{\mathbb{E}(\gamma_n(t))}{\mathbb{E}(\rho_n(t))}\bigg) \\
&= \frac{1}{\rho_n(t)}\frac{1}{\sqrt{nh}}\mathcal{W}_n(t,\varphi) - \frac{\mathbb{E}(\gamma_n(t))}{\rho_n(t)\mathbb{E}(\rho_n(t))}\frac{1}{\sqrt{nh}}\mathcal{W}_n(t,1) - \sqrt{nh}R_n(t).
\end{aligned} \tag{7.10}$$

Then we study the equicontinuity of each of the terms of (7.10) in order to establish the equicontinuity of the process. Let $\alpha_n(\cdot)$ denote the bivariate empirical process based upon $(X_1, Y_1), \ldots, (X_n, Y_n)$ and indexed by a class of functions $\mathscr{G}$. Namely, $\alpha_n(\cdot)$ is defined for $g \in \mathscr{G}$ by

$$\alpha_n(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(g(X_i, Y_i) - \mathbb{E}g(X_i, Y_i)\big).$$

For any class of functions $\mathscr{G}$, set

$$\|\alpha_n(g)\|_{\mathscr{G}} = \sup_{g\in\mathscr{G}}|\alpha_n(g)|\,,$$

and for any measurable function $\varphi(\cdot)$ and $t \in \mathbb{I}$, set

$$\eta_{n,t,\varphi,K}(u,v) = \varphi(v)K\bigg(\frac{u - t}{h}\bigg) \quad \text{for} \quad u, v \in \mathbb{R}.$$

Recalling (7.9), notice that

$$\frac{1}{\sqrt{nh}}\mathcal{W}_n(t,\varphi) = \frac{1}{\sqrt{h}}\alpha_n(\eta_{n,t,\varphi,K}),$$

so we shall first obtain the equicontinuity of the following empirical process

$$\left\{\frac{1}{\sqrt{h}}\alpha_n(\eta_{n,t,\varphi,K})\colon \eta_{n,t,\varphi,K} \in \mathscr{F}\mathscr{K}\right\},$$

that is

$$\lim_{b\to 0}\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{\|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq b}\frac{1}{\sqrt{h}}\left|\alpha_n(\eta_{n,t,\varphi_1,K_1})-\alpha_n(\eta_{n,t,\varphi_2,K_2})\right|>\varepsilon\right\}=0$$

for every $\varepsilon > 0$, or equivalently,

$$\lim_{b\to 0}\limsup_{n\to\infty}\mathbb{P}\left\{\frac{1}{\sqrt{h}}\left\|\alpha_n(\eta_{n,t,\varphi,K})\right\|_{\mathscr{F}\mathscr{K}(b,\|\cdot\|_p)}>\varepsilon\right\}=0 \tag{7.11}$$

where

$$\mathscr{F}\mathscr{K}_{(b,\|\cdot\|_p)}=\left\{\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\colon \left\|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\right\|_p<b,\ \eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in\mathscr{F}\mathscr{K}\right\}.$$

Now we translate our problem to that of the independent block sequence $\{\xi_j=(\zeta_j,\varsigma_j)\}_{j=1}^{\infty}$. We can symmetrize the independent block sequence and work with random entropies. By Eq. (2.5), we have

$$\begin{aligned}
&\mathbb{P}\left\{\left\|(nh)^{-1/2}\sum_{j=1}^{n}\left(\varphi(Y_j)K\left(\frac{X_j-t}{h}\right)-\mathbb{P}(\eta_{n,t,\varphi,K})\right)\right\|_{\mathscr{F}\mathscr{K}(b,\|\cdot\|_p)}>\delta\right\}\\
&\leq 2\mathbb{P}\left\{\left\|(nh)^{-1/2}\sum_{j=1}^{\upsilon_n}\sum_{i\in H_j}\left(\varphi(\zeta_i)K\left(\frac{\varsigma_i-t}{h}\right)-\mathbb{P}(\eta_{n,t,\varphi,K})\right)\right\|_{\mathscr{F}\mathscr{K}(b,\|\cdot\|_p)}>\delta'\right\}\\
&\quad+2(\upsilon_n-1)\beta_{a_n}.
\end{aligned}\tag{7.12}$$

We adapt the choice of [6] for

$$a_n=\left[(\log n)^{-1}(n^{p-2}h^p)^{1/2(p-1)}\right]\quad\text{and}\quad \upsilon_n=\left[\frac{n}{2a_n}\right]-1.$$

So condition **(C.7)** implies $(\upsilon_n-1)\beta_{a_n}\longrightarrow 0$ as $n\to 0$, thus it suffices to treat the first term in the right-hand side of (7.12). To do this and because of independence of the blocks, we symmetrize using a sequence $\{\epsilon_j\}_{j\in\mathbb{N}^*}$ of i.i.d. Rademacher variables, i.e., r.v's with

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

Note that the sequence $\{\epsilon_j\}_{j\in\mathbb{N}^*}$ is independent of the sequence $\{\boldsymbol{\xi}_i=(\varsigma_i,\zeta_i)\}_{i\in\mathbb{N}^*}$, therefore it suffices to prove, for all $\delta>0$,

$$\lim_{b\to 0}\limsup_{n\to\infty}\mathbb{P}\left\{\left\|(nh)^{-1/2}\sum_{j=1}^{\upsilon_n}\epsilon_j\sum_{i\in H_j}\left(\varphi(\zeta_i)K\left(\frac{\varsigma_i-t}{h}\right)\right)\right\|_{\mathscr{F}\mathscr{K}(b,\|\cdot\|_p)}>\delta\right\}=0.$$

Making use of condition **(C.8)**, we can truncate and we get, for each $\lambda>0$, as $n\to\infty$,

$$\begin{aligned}
&(nh)^{-1/2}\sum_{j=1}^{n}\mathbb{E}\left(\kappa F(\zeta_i)\mathbf{1}_{\{F(\zeta_i)\geq\lambda(n/h)^{1/2(p-1)}\}}\right)\\
&=\sqrt{nh^{-1}}\int_0^{\infty}\mathbb{P}\left(\kappa F\mathbf{1}_{\{F\geq\lambda(n/h)^{1/2(p-1)}\}}\geq t\right)dt\\
&=\sqrt{nh^{-1}}\int_0^{\lambda(n/h)^{1/2(p-1)}}\mathbb{P}(F\geq\lambda(n/h)^{1/2(p-1)})\,dt\\
&\quad+\sqrt{nh^{-1}}\int_{\lambda(n/h)^{1/2(p-1)}}^{\infty}\mathbb{P}(F\geq t)\,dt\underset{n\to\infty}{\longrightarrow}0.
\end{aligned}\tag{7.13}$$

Thus from (7.13) there exists a sequence $(\lambda_n) \rightarrow 0$, as $n \to \infty$, for which

$$\sqrt{nh^{-1}}\mathbb{E}(\kappa F \mathbf{1}_{\{F \geq \lambda_n (n/h)^{1/2(p-1)}\}}) \underset{n\to\infty}{\longrightarrow} 0.$$

Therefore we have only to show

$$\lim_{b\to 0} \limsup_{n\to\infty} \mathbb{P}\left\{ \left\| (nh)^{-1/2} \sum_{j=1}^{\upsilon_n} \epsilon_j \sum_{i\in H_j} \left(\varphi(\zeta_i) K\left(\frac{\varsigma_i - t}{h}\right)\right) \mathbf{1}_{\{\kappa F(\zeta_i) \leq \lambda_n (n/h)^{1/2(p-1)}\}} \right\|_{\mathscr{F}\mathscr{K}_{(b, \|\cdot\|_p)}} > \delta \right\}.$$

We apply the chaining argument to the process

$$\nu_n^{(2)}(\eta_{n,t,\varphi,K}) = (nh)^{-1/2} \sum_{j=1}^{\upsilon_n} \epsilon_j \sum_{i\in H_j} \left(\varphi(\zeta_i) K\left(\frac{\varsigma_i - t}{h}\right)\right) \mathbf{1}_{\{\kappa F(\zeta_i) \leq \lambda_n (n/h)^{1/2(p-1)}\}}.$$

And as done in [6], we define $b_k = b2^{-k}$, $k = 0, \ldots, k_n$, where k_n is such that

$$2^{-1}\lambda_n (\log (n))^{-1} \leq b_{k_n}^2 \leq \lambda_n (\log (n))^{-1} \tag{7.14}$$

and $\mathscr{F}\mathscr{K}_k$ is the class of measurable functions of $\mathscr{F}\mathscr{K}$ satisfying

$$\#\mathscr{F}\mathscr{K}_k = N_k := N(b_k, \mathscr{F}\mathscr{K}, \|\cdot\|_p) \quad \sup_{\eta_{n,t,\varphi_1,K_1}\in\mathscr{F}\mathscr{K}} \ \min_{\eta_{n,t,\varphi_2,K_2}\mathscr{F}\mathscr{K}_k} \|\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2}\|_p \leq b_k,$$

so there is a map $\pi_k : \mathscr{F}\mathscr{K} \longrightarrow \mathscr{F}\mathscr{K}_k$ that takes each $\eta_{n,t,\varphi,K} \in \mathscr{F}\mathscr{K}$ to its closest function in $\mathscr{F}\mathscr{K}_k$ such that

$$\|\eta_{n,t,\varphi,K} - \pi_k(\eta_{n,t,\varphi,K})\|_p \leq b_k.$$

By the chaining method,

$$\begin{aligned}
&\sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in\mathscr{F}\mathscr{K}\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq b}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2})\\
&\leq \sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in\mathscr{F}\mathscr{K}\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq b_{k_n}}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2})\\
&\quad + 2\sum_{k=1}^{k_n} \sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in(\mathscr{F}\mathscr{K})_{k-1}\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq 3b_k}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2})\\
&\quad + \sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in(\mathscr{F}\mathscr{K})_0\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq 2b}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2}).
\end{aligned} \tag{7.15}$$

For computational reasons, keep

$$\delta_k = (b_k)^{1/2} \vee (3b_k(8 + c_{p,\beta}^2)^{1/2}(\log N_k)^{1/2}). \tag{7.16}$$

Choosing r so small that $2\sum_{k=1}^{\infty} \delta_k \leq \delta$, we get from (7.15)

$$\begin{aligned}
&\mathbb{P}\left\{ \sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in\mathscr{F}\mathscr{K}\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq b}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2}) \geq 3\delta \right\}\\
&\leq \mathbb{P}\left\{ \sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in\mathscr{F}\mathscr{K}\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq b_{k_n}}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2}) \geq \delta \right\}\\
&\quad + 2\sum_{k=1}^{k_n} \mathbb{P}\left\{ \sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in(\mathscr{F}\mathscr{K})_{k-1}\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq 3b_k}} \nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1} - \eta_{n,t,\varphi_2,K_2}) \geq \delta_k \right\}
\end{aligned}$$

$$+\mathbb{P}\Big\{\sup_{\substack{\eta_{n,t,\varphi_1,K_1},\eta_{n,t,\varphi_2,K_2}\in(\mathscr{F}\mathscr{K})_0\\ \|\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2}\|_p\leq 2b}}\nu_n^{(2)}(\eta_{n,t,\varphi_1,K_1}-\eta_{n,t,\varphi_2,K_2})\geq\delta\Big\}$$
$$=: \mathbb{A}+\mathbb{B}+\mathbb{C}.$$

By the fact that the terms composing $\nu_n^{(2)}(\eta_{n,t,\varphi,K})$ are bounded by $a_n\lambda_n(n/h)^{1/2(p-1)}$, and by applying Bernstein's inequality, we infer that

$$\mathbb{B}\leq 2\sum_{k=1}^{k_n}\exp\left(2\log N_k-\frac{\delta_k^2(nh)}{nb_k^2c_{p,\beta}^2+(4/3)\delta_k a_n\lambda_n n^{p/2(p-1)}h^{(p-2)/2(p-1)}}\right).$$

By the boundedness imposed on b_k in (7.14) we obtain

$$\delta_k a_n\lambda_n n^{p/2(p-1)}h^{(p-2)/2(p-1)}=(4/3)\delta_k\lambda_n(nh)(\log(n))^{-1}\leq(8/3)nb_k^2\delta_k\leq 8nb_k^2,$$

which readily implies that

$$\begin{aligned}\mathbb{B} &\leq 2\sum_{k=1}^{k_n}\exp\left(2\log N_k-\frac{\delta_k^2}{(8+c_{p,\beta}^2)b_k^2}\right)\leq 2\sum_{k=1}^{k_n}\exp\left(-\frac{\delta_k^2}{2(8+c_{p,\beta}^2)b_k^2}\right)\\ &\leq 2\sum_{k=1}^{\infty}\exp\left(-\frac{2^k}{2(8+c_{p,\beta}^2)b}\right)\longrightarrow 0\quad\text{as}\quad b\to 0.\end{aligned}\tag{7.17}$$

In view of (7.16), we assume that $\delta<3$. In a similar way, we have

$$\mathbb{C}\leq 2\exp\left(2\log N_0-\frac{\delta^2}{(8+c_{p,\beta}^2)b^2}\right)\longrightarrow 0\quad\text{as}\quad b\to 0.$$

Finally, by (7.14) it suffices to prove, for each $\delta>0$,

$$\lim_{n\to\infty}\mathbb{P}\Big\{\big\|\nu_n^{(2)}(\eta_{n,t,\varphi,K})\big\|_{\mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}}\geq\delta\Big\}=0.$$

As in [6], we apply to the last expression the square root trick (Lemma 5.2 in [28] and see also [41]). Notice that

$$\begin{aligned}&\mathbb{P}\Big\{\Big\|(nh)^{-1/2}\sum_{j=1}^{\upsilon_n}\epsilon_j\sum_{i\in H_j}\Big(\varphi(\boldsymbol{\zeta}_i)K\Big(\frac{\varsigma_i-t}{h}\Big)\Big)\mathbf{1}_{\{\kappa F(\zeta_i)\leq\lambda_n(n/h)^{1/2(p-1)}\}}\Big\|_{\mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}}\geq 2\delta\Big\}\\
&\leq\mathbb{P}\Big\{\Big\|(nh)^{-1/2}\sum_{j=1}^{\upsilon_n}\epsilon_j\sum_{i\in H_j}\varphi(\boldsymbol{\zeta}_i)K\Big(\frac{\varsigma_i-t}{h}\Big)\mathbf{1}_{\{\kappa F(\boldsymbol{\zeta}_i)\leq\lambda_n(n/h)^{1/2(p-1)}\}}\Big\|_{\mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}}\geq 2\delta,\\
&\quad\Big\|(nh)^{-1}\sum_{j=1}^{\upsilon_n}\Big(\sum_{i\in H_j}\varphi(\boldsymbol{\zeta}_i)K\Big(\frac{\varsigma_i-t}{h}\Big)\mathbf{1}_{\{\kappa F(\zeta_i)\leq\lambda_n(n/h)^{1/2(p-1)}\}}\Big)^2\Big\|_{\mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}}\\
&\qquad\leq 64\lambda_n c_{p,\beta}^2(\log(n))^{-1}\Big\}\\
&\quad+\mathbb{P}\Big\{\Big\|(nh)^{-1}\sum_{j=1}^{\upsilon_n}\Big(\sum_{i\in H_j}\varphi(\boldsymbol{\zeta}_i)K\Big(\frac{\varsigma_i-\mathbf{t}}{h}\Big)\mathbf{1}_{\{\kappa F(\boldsymbol{\zeta}_i)\leq\lambda_n(n/h)^{1/2(p-1)}\}}\Big)^2\Big\|_{\mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}}\\
&\qquad\leq 64\lambda_n c_{p,\beta}^2(\log(n))^{-1}\Big\}\\
&=: \mathbb{P}(\mathbb{A}_1)+\mathbb{P}(\mathbb{A}_2).\end{aligned}$$

So using the semi-norm

$$\widetilde{d}_{nh,2}:=\Big((nh)^{-1}\sum_{j=1}^{\upsilon_n}\sum_{i\in H_j}\big|\eta_{n,t,\varphi_1,K_1}(\varsigma_i,\boldsymbol{\zeta}_i)-\eta_{n,t,\varphi_2,K_2}(\varsigma_i,\boldsymbol{\zeta}_i)\big|^2\Big)^{1/2}$$

and the covering number defined for any class of functions $\mathscr{E}$ by

$$\widetilde{N}_{nh,2}(u,\mathscr{E}) := N_{nh,2}(u,\mathscr{E},\widetilde{d}_{nh,2}),$$

The probability $\mathbb{P}(\mathbb{A}_1)$ can be bounded, as in [6], in the following way:

$$\begin{aligned}\mathbb{P}(\mathbb{A}_1) \;\leq\; & \mathbb{P}\Big\{\mathbb{A}_1, \log \widetilde{N}_{nh,2}\big(\delta a_n^{-1/2}(nh)^{-1/2}, \mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}\big) \\ & \qquad \geq 2^{-4}\min\left(n, 2^{-6}c_{p,\beta}^{-2}\delta^2\lambda_n^{-1}\log(n)\right)\Big\} \\ & + \mathbb{P}\Big\{\mathbb{A}_1, \log \widetilde{N}_{nh,2}\big(\delta a_n^{-1/2}(nh)^{-1/2}, \mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(n))^{-1/2},\|\cdot\|_p)}\big) \\ & \qquad < 2^{-4}\min\left(n, 2^{-6}c_{p,\beta}^{-2}\delta^2\lambda_n^{-1}\log(n)\right)\Big\} =: \mathrm{I}_{\mathbb{A}_1} + \mathrm{II}_{\mathbb{A}_1}.\end{aligned}$$

By using condition **(C.10)** we establish $\mathrm{I}_{\mathbb{A}_1} \to 0$. Now for treating $\mathrm{II}_{\mathbb{A}_1} \to 0$, let us consider the dense net $\mathscr{E}_{(\delta a_n^{-1/2}(nh)^{-1/2},\widetilde{d}_{nh,2})}$ of $\mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(nh))^{-1/2},\|\cdot\|_p)}$ of cardinality

$$\#\mathscr{E}_{(\delta a_n^{-1/2}(nh)^{-1/2},\widetilde{d}_{nh,2})} := \widetilde{N}_{nh,2}\big(\delta a_n^{-1/2}(nh)^{-1/2}, \mathscr{F}\mathscr{K}_{(\lambda_n^{1/2}(\log(nh))^{-1/2},\|\cdot\|_p)}\big).$$

An application of Hoeffding's inequality to

$$\left\{\left\|(nh)^{-1/2}\sum_{j=1}^{\upsilon_n}\epsilon_j\sum_{i\in H_j}\varphi(\boldsymbol{\zeta}_i)K\Big(\frac{\varsigma_i - t}{h}\Big)\mathbf{1}_{\{\kappa F(\boldsymbol{\zeta}_i)\leq\lambda_n(n/h)^{1/2(p-1)}\}}\right\|_{\mathscr{E}_{(\delta a_n^{-1/2}(nh)^{-1/2},\widetilde{d}_{nh,2})}} \geq \lambda\right\}$$

gives that $\mathrm{II}_{\mathbb{A}_1} \to 0$. Consider now the probability of $\mathbb{A}_2$. Following [6], since the blocks are i.i.d., we can apply again Lemma 5.2 in [28] to bound $\mathbb{P}(\mathbb{A}_2)$ from above. This is achieved by using the condition **(C.10)** to get $\mathbb{P}(\mathbb{A}_2) \to 0$, therefore the process

$$\left\{\frac{1}{\rho_n(t)}\frac{1}{\sqrt{nh}}\mathcal{W}_n(t,\varphi)\colon \varphi K \in \mathscr{F}\mathscr{K}\right\}$$

satisfies (7.11). In a similar way we treat

$$\left\{\frac{1}{\sqrt{h}}\alpha_{nh}(\eta_{n,t,1,K})\colon \eta_{n,t,1,K} \in \mathscr{K}\right\}.$$

Further, the class $\mathscr{F}\mathscr{K}$ meets all of the conditions **(C.7)**, **(C.9)** and **(C.10)** in addition to $\kappa < \infty$ so the process

$$\left\{\frac{\mathbb{E}(\gamma_n(t))}{\rho_n(t)\mathbb{E}(\rho_n(t))}\frac{1}{\sqrt{h}}\alpha_{nh}(\eta_{n,t,1,K}\colon \eta_{n,t,1,K} \in \mathscr{K}\right\}$$

satisfies (2.2.2). Hence the proof of the theorem is complete. □

Proof of Theorem 2.2.3

Under conditions **(C.1)–(C.6)** convergence of the finite-dimensional distributions follows directly from Theorem 2.2.1 and from **(C.7)–(C.8)**. Since the VC-type class, whose envelope is in L_p, satisfies conditions **(C.9)–(C.10)**, the asymptotic uniform equicontinuity condition follows directly from Theorem 2.2.2. □

Proof of Theorem 3.0.1

Our work-plan to establish the convergence of our previously defined conditional U-process indexed by a class of functions not necessarily uniformly bounded is to break the U-process into two parts:

- One that is called the truncated part, where we assume that the class $\mathscr{F}_m\mathscr{K}^m$ is uniformly bounded, and which will in turn be divided into two parts (linear and non-linear).
- The second is called the remainder part, which we will prove later to be asymptotically negligible.

Notice that, in order to simplify notation and calculations, we assume that the zero function belongs to the class $\mathscr{F}_m\mathscr{K}^m$.

Preliminaries of the Proof of Theorem 3.0.1

Here we develop some details that will be used in the proof of Theorem 3.0.1. Making use of (3.3) we have, for each $\lambda > 0$,

$$\begin{aligned}G_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y}) &= G_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y})\mathbf{1}_{\{\kappa^m F(\mathbf{y})\leq\lambda(n/h^m)^{1/2(p-1)}\}} + G_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y})\mathbf{1}_{\{\kappa^m F(\mathbf{y})>\lambda(n/h^m)^{1/2(p-1)}\}}\\ &=: G^{(T)}_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y}) + G^{(R)}_{\varphi,\mathbf{t}}(\mathbf{x},\mathbf{y}).\end{aligned}$$

We can write the U-statistic as follows:

$$\begin{aligned}\mu_n(\varphi,\mathbf{t}) &= \sqrt{nh^m}\big\{u_n^{(m)}(G^{(T)}_{\varphi,\mathbf{t}}) - \mathbb{E}(u_n^{(m)}\big(G^{(T)}_{\varphi,\mathbf{t}}))\big\} + \sqrt{nh^m}\big\{u_n^{(m)}(G^{(R)}_{\varphi,\mathbf{t}}) - \mathbb{E}(u_n^{(m)}(G^{(R)}_{\varphi,\mathbf{t}}))\big\}\\ &=: \sqrt{nh^m}\big\{u_n^{(T)}(\varphi,\mathbf{t}) - \mathbb{E}(u_n^{(T)}(\varphi,\mathbf{t}))\big\} + \sqrt{nh^m}\big\{u_n^{(R)}(\varphi,\mathbf{t}) - \mathbb{E}(u_n^{(R)}(\varphi,\mathbf{t}))\big\}\\ &=: \mu_n^{(T)}(\varphi,\mathbf{t}) + \mu_n^{(R)}(\varphi,\mathbf{t}). \qquad (7.18)\end{aligned}$$

We call the first term of the right-hand side of (7.18) $\mu_n^{(T)}(\varphi,\mathbf{t})$ truncated part and the second $\mu_n^{(R)}(\varphi,\mathbf{t})$ remainder part. First we are interested in $\mu_n^{(T)}(\varphi,\mathbf{t})$. An application of Hoeffding's decomposition tells us that

$$\begin{aligned}u_n^{(T)}(\varphi,\mathbf{t}) &= \sum_{k=0}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}})\\ &= \mathbb{E}G^{(T)}_{\varphi,\mathbf{t}}(\mathbf{X}',\mathbf{Y}') + \sum_{k=1}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}}), \qquad (7.19)\end{aligned}$$

where $\{\mathbf{Z}'_i = (\mathbf{X}'_i,\mathbf{Y}'_i)\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. r.v.'s with $\mathcal{L}(\mathbf{Z}'_i) = \mathcal{L}(\mathbf{Z}_i)$ for each i, and $\mathbf{X}'$ and $\mathbf{Y}'$ are respectively defined as **X** and **Y**. In view of (7.19), we have

$$\mu_n^{(T)}(\varphi,\mathbf{t}) = \sqrt{nh^m}\Big\{\mathbb{E}G^{(T)}_{\varphi,t}(\mathbf{X}',\mathbf{Y}') + \sum_{k=1}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}}) - \mathbb{E}(u_n^{(T)}(\varphi,\mathbf{t}))\Big\},$$

the stationarity assumption and routine calculus of $\mathbb{E}(u_n^{(T)}(\varphi,\mathbf{t}))$ show that

$$\mathbb{E}(u_n^{(T)}(\varphi,\mathbf{t})) = \mathbb{E}G^{(T)}_{\varphi,\mathbf{t}}(\mathbf{X}',\mathbf{Y}').$$

From this, we infer that

$$\begin{aligned}\mu_n^{(T)}(\varphi,\mathbf{t}) &= \sqrt{nh^m}\Big\{\sum_{k=1}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}})\Big\}\\ &= \sqrt{nh^m}\Big\{mu_n^{(1)}(\pi_{1,m}G^{(T)}_{\varphi,\mathbf{t}}) + \sum_{k=2}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}})\Big\}. \qquad (7.20)\end{aligned}$$

Yoshihara [70] proved that if $\varphi\widetilde{K}$ is $\mathbb{P}$-canonical and $k \geq 2$, then

$$\mathbb{E}\big((nh^m)^{\frac{1}{2}}u_n^{(k)}(\varphi\widetilde{K},\mathbf{t})\big)^2 = O\big(n^{1-p(p-1)r/(p+1)(p-2)}h^{m/2}\big) = O\big(n^{1-p(p-1)r/(p+1)(p-2)}\big). \qquad (7.21)$$

By the fact that $\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}}$ is $\mathbb{P}$-canonical and making use of (7.21) we obtain that

$$\Big(\sqrt{nh^m}\sum_{k=2}^{m}u_n^{(k)}(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}})\Big) \xrightarrow{\mathbb{P}} 0.$$

So that to establish the weak convergence of the U-process $\{\mu_n^{(T)}(\varphi,\mathbf{t})\}_{\mathscr{F}_m\mathscr{K}^m}$, by Hoeffding's decomposition it is enough to show

$$m\sqrt{nh^m}u_n^{(1)}(\pi_{1,m}G^{(T)}_{\varphi,\mathbf{t}}) \xrightarrow{w} \mathbb{G}(\varphi) \qquad \text{in} \quad \ell_\infty(m\mathcal{G}^{(1)}),$$

where $\{\mathbb{G}(\varphi)\}_{m\mathcal{G}^{(1)}}$ is a Gaussian process indexed by $m\mathcal{G}^{(1)}$, and

$$\left\|\sqrt{nh^m}u_n^{(k)}(\pi_{k,m}G_{\varphi,\mathbf{t}}^{(T)})\right\|_{\mathscr{F}_m\mathscr{K}^m} \xrightarrow{\mathbb{P}} 0,$$

for $2 \le k \le m$. Then we have to prove that the remainder part is negligible, in the sense that

$$\left\|\sqrt{nh^m}\{u_n^{(R)}(\varphi,\mathbf{t}) - \mathbb{E}(u_n^{(R)}(\varphi,\mathbf{t}))\}\right\|_{\mathscr{F}_m\mathscr{K}^m} \xrightarrow{\mathbb{P}} 0.$$

The following technical lemma will be instrumental in the proof of our theorem.

Lemma 7.0.2. *Let $\mathscr{F}_m\mathscr{K}^m$ be a uniformaly bounded class of measurable canonical functions from $\mathbb{R}^m \times \mathcal{X}^m \longrightarrow \mathbb{R}$, $m \ge 2$. Suppose that there are finite constants $\boldsymbol{a}$ and $\boldsymbol{b}$ such that the $\mathscr{F}_m\mathscr{K}^m$ covering number satisfies*

$$N(\epsilon, \mathscr{F}_m\mathscr{K}^m, \|\cdot\|_{L_2(Q)}) \le \boldsymbol{a}\epsilon^{-\boldsymbol{b}} \tag{7.22}$$

for every $\epsilon > 0$ and every probability measure Q. If the mixing coefficients β of the stationary sequence $\{Z_i = (X_i, Y_i)\}_{i\in\mathbb{N}^*}$ fulfill

$$\beta_k k^r \longrightarrow, \text{ as } k \to \infty \tag{7.23}$$

for some $r > 1$, then

$$\left\|h^{m/2}n^{-m+\frac{1}{2}}\sum_{\mathbf{i}\in I_m^n} G_{\varphi,\mathbf{t}}(\mathbf{X}_i, \mathbf{Y}_i)\right\|_{\mathscr{F}_m\mathscr{K}^m} \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 7.0.2

For clarity of exposition we present the proof for $m = 2$, this case already contains the main idea. As in the proof of Theorem 2.2.2, we divide the sequence $\{(\mathbf{X}_i, \mathbf{Y}_i)\}$ into υ_n alternate blocks, of sizes a_n, b_n, which are different and satisfy

$$b_n \ll a_n, \qquad (\upsilon_n - 1)(a_n + b_n) < n \le \upsilon_n(a_n + b_n). \tag{7.24}$$

Set, for $1 \le j \le \upsilon_n - 1$,

$$\begin{aligned}
H_j^{(U)} &= \{i\colon (j-1)(a_n+b_n)+1 \le i \le (j-1)(a_n+b_n)+a_n\},\\
T_j^{(U)} &= \{i\colon (j-1)(a_n+b_n)+a_n+1 \le i \le (j-1)(a_n+b_n)+a_n+b_n\},\\
H_{\upsilon_n}^{(U)} &= \{i\colon (\upsilon_n-1)(a_n+b_n)+1 \le i \le n \wedge (\upsilon_n-1)(a_n+b_n)+a_n\},\\
T_{\upsilon_n}^{(U)} &= \{i\colon (\upsilon_n-1)(a_n+b_n)+a_n+1 \le i \le n\}.
\end{aligned}$$

Note that b_n used here and in the proof of Theorem 3.0.1 denotes the size of the alternative blocks. However in the proof of Theorem 2.2.2 it denotes the radius of the nets of the class of functions.

We decompose the process according to the distribution of the blocks:

$$\begin{aligned}
&\sum_{i\neq j}^{n} \frac{1}{h^2}\varphi(Y_i, Y_j)K\Big(\frac{X_i - t_1}{h}\Big)K\Big(\frac{X_j - t_2}{h}\Big)\\
&\quad= \sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}} \frac{1}{h^2}\varphi(Y_i, Y_j)K\Big(\frac{X_i - t_1}{h}\Big)K\Big(\frac{X_j - t_2}{h}\Big)\\
&\qquad+ \sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in H_p^{(U)}} \frac{1}{h^2}\varphi(Y_i, Y_j)K\Big(\frac{X_i - t_1}{h}\Big)K\Big(\frac{X_j - t_2}{h}\Big)
\end{aligned}$$

$$+2\sum_{p=1}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\frac{1}{h^2}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)$$

$$+2\sum_{p=1}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\leq 1}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\frac{1}{h^2}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)$$

$$+\sum_{p\neq q}^{\upsilon_n}\sum_{i\in T_p^{(U)}}\sum_{j\in T_q^{(U)}}\frac{1}{h^2}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)$$

$$+\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in T_p^{(U)}}\frac{1}{h^2}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)$$

$$=:\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V}+\mathrm{VI}. \tag{7.25}$$

We have to treat each of the terms I − VI. The treatment of V and VI is readily achieved through the similar techniques used to investigate V and VI, which we omit.

(I). The same type of block but not the same block. Suppose that the sequence of independent blocks $\{\xi_i=(\varsigma_i,\zeta_i)\}_{i\in\mathbb{N}^*}$ is of size a_n. An application of (2.5) shows that

$$\mathbb{P}\Big\{\Big\|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}>\delta\Big\}$$

$$\leq 2\upsilon_n\beta_{b_n}+\mathbb{P}\Big\{\Big\|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}>\delta\Big\}.$$

We keep the choice of b_n and υ_n such that:

$$\upsilon_n b_n^r\leq 1, \tag{7.26}$$

which implies that $2\upsilon_n\beta_{b_n}\to 0$ as $n\to\infty$, so the term to consider is the second summand. By combining Lemma A.1 in [11] with Proposition 7 in The Appendix, we obtain:

$$\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$

$$\leq c_2\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\epsilon_p\epsilon_q\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$

$$\leq c_2\mathbb{E}\int_0^{D_{nh}^{(U_1)}}N(u,\mathscr{F}_2\mathscr{K}^2,\widetilde{d}_{nh,2}^{(1)})\,du, \tag{7.27}$$

where $D_{nh}^{(U_1)}$ is the diameter of $\mathscr{F}_2\mathscr{K}^2$ according to the distance $\widetilde{d}_{nh,2}^{(1)}$, which are defined respectively by

$$D_{nh}^{(U_1)}:=\Big\|\mathbb{E}_\epsilon\Big|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\epsilon_p\epsilon_q\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big|\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$

and

$$\widetilde{d}_{nh,2}^{(1)}(\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2)\ :=\ \mathbb{E}_\epsilon\Big|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\epsilon_p\epsilon_q\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\Big[\varphi_1(\zeta_i,\zeta_j)K_1\Big(\frac{\varsigma_i-t_1}{h}\Big)K_1\Big(\frac{\varsigma_j-t_2}{h}\Big)$$

$$-\varphi_2(\zeta_i,\zeta_j)K_2\Big(\frac{\varsigma_i-t_1}{h}\Big)K_2\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big]\Big|.$$

Let us consider another semi-norm $\widetilde{d}^{(2)}_{nh,2}$,

$$\widetilde{d}^{(2)}_{nh,2}(\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2) = \frac{1}{nh^2}\Big[\sum_{i\neq j}^{\upsilon_n}\Big(\varphi_1(\zeta_i,\zeta_j)K_1\Big(\frac{\varsigma_i-t_1}{h}\Big)K_1\Big(\frac{\varsigma_j-t_2}{h}\Big) -\varphi_2(\zeta_i,\zeta_j)K_2\Big(\frac{\varsigma_i-t_1}{h}\Big)K_2\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)^2\Big]^{1/2}.$$

One can see that

$$\widetilde{d}^{(1)}_{nh,2}(\varphi_1K_1,\varphi_2K_2) \le a_n n^{-1/2}h\widetilde{d}^{(2)}_{nh,2}(\varphi_1K_1,\varphi_2K_2) \le a_n n^{-1/2}\widetilde{d}^{(2)}_{nh,2}(\varphi_1K_1,\varphi_2K_2).$$

We readily infer that

$$\begin{aligned}
&\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\\
&\le c_2\mathbb{E}\int_0^{D_{nh}^{(U_1)}} N\big(ua_n^{-1}n^{1/2},\mathscr{F}_2\mathscr{K}^2,\widetilde{d}^{(2)}_{nh,2}\big)\,du\\
&\le c_2a_nn^{-1/2}\mathbb{P}\big\{D_{nh}^{(U_1)}a_n^{-1}n^{1/2}\geqslant\lambda_n\big\}+c_ma_nn^{-1/2}\int_0^{\lambda_n}\log u^{-1}\,du,
\end{aligned}$$

where $\lambda_n\to 0$. Notice that as $\lambda\to 0$, we have

$$\Big(\int_0^{\lambda}\log u^{-1}du\Big)\Big/(\lambda\log\lambda^{-1})\to 0,$$

where a_n and λ_n are chosen so that

$$a_n\lambda_n n^{-1/2}\log\lambda_n^{-1}\to 0. \tag{7.28}$$

Making use of the triangle inequality, in combination with Hoeffding's trick, we obtain readily that

$$\begin{aligned}
&a_nn^{-1/2}\mathbb{P}\big\{D_{nh}^{(U_1)}\geqslant\lambda_na_nn^{-1/2}\big\}\\
&\le\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\mathbb{E}\Big\|\sum_{p\neq q}^{\upsilon_n}\Big[\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big]^2\Big\|_{\mathscr{F}_2\mathscr{K}^2}\\
&\le c_2\upsilon_n\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\mathbb{E}\Big\|\sum_{p=1}^{\upsilon_n}\Big[\sum_{i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta'_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma'_j-t_2}{h}\Big)\Big]^2\Big\|_{\mathscr{F}_2\mathscr{K}^2},
\end{aligned}\tag{7.29}$$

where $\{\xi'_i=(\varsigma'_i,\zeta'_i)\}_{i\in\mathbb{N}^*}$ are independent copies of $\{\xi_i=(\varsigma_i,\zeta_i)\}_{i\in\mathbb{N}^*}$. By imposing

$$\lambda_n^{-2}a_n^{1-r}n^{-1/2}\to 0, \tag{7.30}$$

we readily infer from (7.21) that

$$\Big\|\upsilon_n\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\mathbb{E}\sum_{p=1}^{\upsilon_n}\Big[\sum_{i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta'_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma'_j-t_2}{h}\Big)\Big]^2\Big\|_{\mathscr{F}_2\mathscr{K}^2}=O\big(\lambda_n^{-2}a_n^{1-r}n^{-1/2}\big).$$

By symmetrizing the expression in (7.29) and applying again Proposition 7 in the Appendix, we get

$$\begin{aligned}
&\upsilon_n\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\mathbb{E}\Big\|\sum_{p=1}^{\upsilon_n}\Big[\sum_{i,j\in H_p^{(U)}}\epsilon_p\varphi(\zeta_i,\zeta'_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma'_j-t_2}{h}\Big)\Big]^2\Big\|_{\mathscr{F}_2\mathscr{K}^2}\\
&\le c_2\mathbb{E}\Big(\int_0^{D_{nh}^{(U_2)}}\big(\log N(u,\mathscr{F}_2\mathscr{K}^2,\widetilde{d}'_{nh,2})\big)^{1/2}\Big),
\end{aligned}\tag{7.31}$$

where

$$D_{nh}^{(U_2)} = \Big\|\mathbb{E}_\epsilon\Big|\upsilon_n\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\sum_{p=1}^{\upsilon_n}\epsilon_p\Big[\sum_{i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j')K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j'-t_2}{h}\Big)\Big]^2\Big|\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$

and for $\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2\in\mathscr{F}_2\mathscr{K}^2$,

$$\widetilde{d}_{nh,2}'(\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2) = \mathbb{E}_\epsilon\Big|\upsilon_n\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\sum_{p=1}^{\upsilon_n}\epsilon_p\Big[\Big(\sum_{i,j\in H_p^{(U)}}\varphi_1(\zeta_i,\zeta_j')K_1\Big(\frac{\varsigma_i-t_1}{h}\Big)K_1\Big(\frac{\varsigma_j'-t_2}{h}\Big)\Big)^2$$
$$-\Big(\sum_{i,j\in H_p^{(U)}}\varphi_2(\zeta_i,\zeta_j')K_2\Big(\frac{\varsigma_i-t_1}{h}\Big)K_2\Big(\frac{\varsigma_j'-t_2}{h}\Big)\Big)^2\Big]\Big|.$$

By the fact that

$$\mathbb{E}_\epsilon\Big|\upsilon_n\lambda_n^{-2}a_n^{-1}n^{-5/2}h^{-2}\sum_{p=1}^{\upsilon_n}\epsilon_p\Big(\sum_{i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j')K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j'-t_2}{h}\Big)\Big)^2\Big|$$
$$\leq a_n^{3/2}\lambda_n^{-2}n^{-1}\Big[\upsilon_n^{-1}a_n^{-2}h^{-4}\sum_{p=1}^{\upsilon_n}\sum_{i,j\in H_p^{(U)}}\Big(\varphi(\zeta_i,\zeta_j')K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j'-t_2}{h}\Big)\Big)^4\Big]^{1/2},$$

so

$$a_n^{3/2}\lambda_n^{-2}n^{-1}h^{-2}\to 0, \tag{7.32}$$

we have the convergence of (7.31) to zero. For the choice of a_n, b_n and υ_n, it should be noted that all the values satisfying (7.24), (7.26), (7.28), (7.30) and (7.32) are accepted.

(II). The same blocks. We have

$$\mathbb{P}\Big\{\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in H_p^{(U)}}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)\Big\|_{\mathscr{F}\mathscr{K}^2}>\lambda\Big\}$$
$$\leq 2\upsilon_n\beta_{b_n}+\mathbb{P}\Big\{\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}>\lambda\Big\}.$$

In a similar way as in the preceding proof, it suffices to prove that

$$\mathbb{E}\Big(\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\Big)\to 0.$$

Notice that we treat uniformly bounded classes of functions. We obtain uniformly in $\mathscr{F}_2\mathscr{K}^2$

$$\mathbb{E}\Big(\sum_{i\neq j\, i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)=O(a_n).$$

This implies that we have to prove that

$$\mathbb{E}\Big(\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in H_p^{(U)}}\Big[\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)$$
$$-\mathbb{E}\Big(\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)\Big]\Big\|_{\mathscr{F}_2\mathscr{K}^2}\Big)\to 0. \tag{7.33}$$

Like for empirical processes, to prove (7.33), it suffices to symmetrize and show that

$$\mathbb{E}\Big(\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j}\sum_{i,j\in H_p^{(U)}}\epsilon_p\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\Big)\to 0.$$

Similarly to (7.27), we infer that

$$\begin{aligned}&\mathbb{E}\Big(\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j}\sum_{i,j\in H_p^{(U)}}\epsilon_p\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\Big)\\&\quad\leq\mathbb{E}\Big(\int_0^{D_{nh}^{(U_3)}}\big(\log N(u,\mathscr{F}_2\mathscr{K}^2,\widetilde{d}_{nh,2}^{(3)})\big)^{1/2}du\Big),\end{aligned}$$

where

$$D_{nh}^{(U_3)}=\Big\|\mathbb{E}_\epsilon\Big|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\epsilon_p\sum_{i\neq j}\sum_{i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big|\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$

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

$$\begin{aligned}\widetilde{d}_{nh,2}^{(3)}(\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2)\;=\;&\mathbb{E}_\epsilon\Big|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\epsilon_p\sum_{i\neq j}\sum_{i,j\in H_p^{(U)}}\Big(\varphi_1(\zeta_i,\zeta_j)K_1\Big(\frac{\varsigma_i-t_1}{h}\Big)K_1\Big(\frac{\varsigma_j-t_2}{h}\Big)\\&-\varphi_2(\zeta_i,\zeta_j)K_2\Big(\frac{\varsigma_i-t_1}{h}\Big)K_2\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)\Big|.\end{aligned}$$

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

$$\begin{aligned}&\mathbb{E}_\epsilon\Big|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\epsilon_p\sum_{i\neq j}\sum_{i,j\in H_p^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big|\\&\leq a_n^{3/2}(nh)^{-1}\Big[\frac{1}{\upsilon_n a_n^2}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j}\sum_{i,j\in H_p^{(U)}}\Big(\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)^2\Big]^{1/2}\\&=O(a_n^{3/2}(nh)^{-1}).\end{aligned}$$

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

(III) Different types of blocks. An application of (2.5), shows that

$$\begin{aligned}&\sum_{p=1}^{\upsilon_n}\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\\&\quad\leq\sum_{p=1}^{\upsilon_n}\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\\&\qquad+n^{-3/2}h^{-1}\upsilon_n^2a_nb_n\beta_{a_n}.\end{aligned}$$

By the last choice of the parameters a_n, b_n, υ_n and condition (7.23) imposed on the β-coefficients, we have

$$n^{-3/2}h^{-1}\upsilon_n^2a_nb_n\beta_{a_n}\to 0.$$

For $p = 1$ and $p = \upsilon_n$, since we have independent exchangeable blocks, we infer that

$$\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$=\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_{\upsilon_n}^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$=\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}.$$

For $2\leq p\leq \upsilon_n-1$, we obtain

$$\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$=\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=4}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$\leq\mathbb{E}\Big\|n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2},$$

therefore it suffices to treat the convergence

$$\mathbb{E}\Big\|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\longrightarrow 0.$$

By similar arguments to those in [6], the usual symmetrization applies and

$$\mathbb{E}\Big\|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$\leq 2\mathbb{E}\Big\|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\epsilon_q\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$=2\mathbb{E}\Big\{\Big\|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\epsilon_q\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\mathbf{1}_{\{D_{nh}^{(U_4)}\leq\gamma_n\}}\Big\}$$
$$+2\mathbb{E}\Big\{\Big\|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\epsilon_q\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\mathbf{1}_{\{D_{nh}^{(U_4)}>\gamma_n\}}\Big\}$$
$$=2\mathrm{III}_1+2\mathrm{III}_2, \tag{7.34}$$

where:

$$D_{nh}^{(U_4)}=\Big\|\upsilon_n n^{-3/2}h^{-1}\Big[\sum_{q=3}^{\upsilon_n}\Big(\sum_{j\in T_q^{(U)}}\sum_{i\in H_1^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)^2\Big]^{1/2}\Big\|_{\mathscr{F}_2\mathscr{K}^2}.$$

Similarly to (7.27), we infer that

$$\mathrm{III}_1\leq c_2\int_0^{\gamma_n}\big(\log N(u,\mathscr{F}_2\mathscr{K}^2,\widetilde{d}_{nh,2}^{(4)})\big)^{1/2}\,du, \tag{7.35}$$

where

$$\widetilde{d}_{nh,2}^{(4)}(\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2) \;=\; \mathbb{E}_\epsilon\Big|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\epsilon_q\Big[\varphi_1(\zeta_i,\zeta_j)K_1\Big(\frac{\varsigma_i-t_1}{h}\Big)K_1\Big(\frac{\varsigma_j-t_2}{h}\Big)$$
$$-\varphi_2(\zeta_i,\zeta_j)K_2\Big(\frac{\varsigma_i-t_1}{h}\Big)K_2\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big]\Big|.$$

Since we have

$$\mathbb{E}_\epsilon\Big|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\epsilon_q\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big|$$
$$\leq a_n^{-1/2}b_nh\Big[\Big(\frac{1}{a_nb_n\upsilon_nh^4}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\Big[\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big]^2\Big)^{1/2}\Big],$$

and by considering the semi-metric

$$\widetilde{d}_{nh,2}^{(5)}(\varphi_1\tilde{K}_1,\varphi_2\tilde{K}_2) \;=\; \Big(\frac{1}{a_nb_n\upsilon_nh^4}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\Big[\varphi_1(\zeta_i,\zeta_j)K_1\Big(\frac{\varsigma_i-t_1}{h}\Big)K_1\Big(\frac{\varsigma_j-t_2}{h}\Big)$$
$$-\varphi_2(\zeta_i,\zeta_j)K_2\Big(\frac{\varsigma_i-t_1}{h}\Big)K_2\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big]^2\Big)^{1/2},$$

we show that the expression in (7.35) is bounded by

$$\upsilon_n^{1/2}b_nn^{-1/2}h\int_0^{\upsilon_n^{-1/2}b_n^{-1}n^{1/2}h^{-1}\gamma_n}\big(\log N(u,\mathscr{F}_2\mathscr{K}^2,\widetilde{d}_{nh,2}^{(5)})\big)^{1/2}\,du.$$

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 in the right-hand side of (7.34), we remark that

$$\begin{aligned}\mathrm{III}_2 \;&=\; \mathbb{E}\Big\{\Big\|\upsilon_n n^{-3/2}h^{-1}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\epsilon_q\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}\mathbf{1}_{\{D_{nh}^{(U_4)}>\gamma_n\}}\Big\}\\ &\leq\; a_n^{-1}b_nn^{1/2}h^{-1}\\ &\times\;\mathbb{P}\Big\{\Big\|\upsilon_n^2n^{-3}h^{-2}\sum_{q=3}^{\upsilon_n}\Big(\sum_{j\in T_q^{(U)}}\sum_{i\in H_1^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)^2\Big\|_{\mathscr{F}_2\mathscr{K}^2}\geqslant\gamma_n^2\Big\},.\end{aligned}\tag{7.36}$$

Now we apply the square root trick to the last expression conditionally on H_1^U. We denote by $\mathbb{E}_T$ the expectation with respect to $\sigma\{(\varsigma_j,\zeta_j)\colon j\in T_q, q\geqslant 3\}$ and we get by equation (2.4) for $2r/(r-1)<s<\infty$ (in the notation in Lemma 5.2 in [28])

$$M_n=\upsilon_n^{1/2}\mathbb{E}_T\Big(\sum_{j\in T_q^{(U)}}\sum_{i\in H_1^{(U)}}\varphi(\zeta_i,\zeta_j)K\Big(\frac{\varsigma_i-t_1}{h}\Big)K\Big(\frac{\varsigma_j-t_2}{h}\Big)\Big)^2$$
$$t=\gamma_n^2a_n^{5/2}n^{1/2}h^{-1},\quad \rho=\lambda=2^{-4}\gamma_na_n^{5/4}n^{1/4}h^{-1/2},\quad m=\exp\big(\gamma_n^2nh^{-2}b_n^{-2}\big)$$

Since we need $t>8M_n$, and $m\to\infty$, by similar arguments as in [6], p. 69, we get the convergence of (7.35) and (7.36) to zero.

(IV). Different types of blocks. We have

$$\Big\|n^{-3/2}h^{-1}\sum_{p=1}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\leq 1}^{\upsilon_n}\sum_{j\in T_q^{(U)}}\varphi(Y_i,Y_j)K\Big(\frac{X_i-t_1}{h}\Big)K\Big(\frac{X_j-t_2}{h}\Big)\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$

$$\le c_2 v_n a_n b_n n^{-3/2} h^{-1} \to 0.$$

Hence the proof of the lemma is complete. □

Armed with Lemma 7.0.2, we are now ready to complete the proof of Theorem 3.0.1. This will be achieved in the forthcoming section.

Proof of Theorem 3.0.1

In order to prove the convergence of the process $\{U_n^{(m)}(\varphi, \mathbf{t})\}_{\mathscr{F}_m\mathscr{K}^m}$, we shall proceed as follows:

$$\begin{aligned} U_n^{(m)}(\varphi, \mathbf{t}) &= \sqrt{nh^m}\Big(\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) - r^{(m)}(\varphi, \mathbf{t})\Big) \\ &= \sqrt{nh^m}\Big\{\frac{u_n(\varphi, \mathbf{t})}{u_n(1, \mathbf{t})} - \frac{\mathbb{E}(u_n(\varphi, \mathbf{t}))}{\mathbb{E}(u_n(1, \mathbf{t}))} - r^{(m)}(\varphi, \mathbf{t}) + \frac{\mathbb{E}(u_n(\varphi, \mathbf{t}))}{\mathbb{E}(u_n(1, \mathbf{t}))}\Big\} \\ &= \frac{1}{u_n(1, \mathbf{t})}\sqrt{nh^m}\big(u_n(\varphi, \mathbf{t}) - \mathbb{E}(u_n(\varphi, \mathbf{t}))\big) - \frac{\mathbb{E}(u_n(\varphi, \mathbf{t}))}{u_n(1, \mathbf{t})\mathbb{E}(u_n(1, \mathbf{t}))} \\ &\times \sqrt{nh^m}\big(u_n(1, \mathbf{t}) - \mathbb{E}(u_n(1, \mathbf{t}))\big) - \sqrt{nh^m}\Big(r^{(m)}(\varphi, \mathbf{t}) - \frac{\mathbb{E}(u_n(\varphi, \mathbf{t}))}{\mathbb{E}(u_n(1, \mathbf{t}))}\Big). \end{aligned}$$

Our conditional U-process is a sum of U-processes that will be treated. Let us begin with the finite-dimensional convergence of $\{\mu_n(\varphi, \mathbf{t})\}_{\mathscr{F}_m\mathscr{K}^m}$ that can be obtained from Theorem 1 in [70], which asserts that

$$\mu_n(\varphi, \mathbf{t}) = \sqrt{nh^m}\big\{u_n(\varphi, \mathbf{t}) - \mathbb{E}(u_n(\varphi, \mathbf{t}))\big\}$$

converges in distribution to a Gaussian r.v. if the functions, for fixed φ, K and the β-mixing coefficients, satisfy the conditions:

(H.1) there are constants M' and $p > 2$ such that

$$\mathbb{E}\Big|\varphi(Y_{i_1}, \ldots, Y_{i_m})\prod_{j=1}^{m} K\Big(\frac{X_{i_j} - t_j}{h}\Big)\Big| \le M';$$

(H.2) the β-mixing coefficients satisfy, for some $r > 1$,

$$\beta_n = O\big((n)^{-rp/(p-2)}\big).$$

These conditions are fulfilled thanks to (1.2), (3.3) and (3.4) respectively, so we have the finite-dimensional convergence. Let us consider now the tightness of the process $\big\{\mu_n(\varphi, \mathbf{t})\big\}_{\mathscr{F}_m\mathscr{K}^m}$. As was mentioned earlier, we decompose the U-process $\mu_n(\varphi, \mathbf{t})$ into two parts, the truncated and remainder parts,

$$\mu_n(\varphi, \mathbf{t}) = \mu_n^{(T)}(\varphi, \mathbf{t}) + \mu_n^{(R)}(\varphi, \mathbf{t}).$$

The truncated part $\mu_n^{(T)}(\varphi, \mathbf{t})$ is decomposed according to the Hoeffding's decomposition as it is displayed in the preliminaries:

$$\mu_n^{(T)}(\varphi, \mathbf{t}) = \sqrt{nh^m}\Big\{mu_n^{(1)}(\pi_{1,m}G_{\varphi,\mathbf{t}}^{(T)}) + \sum_{k=2}^{m}\frac{m!}{(m-k)!}u_n^{(k)}(\pi_{k,m}G_{\varphi,\mathbf{t}}^{(T)})\Big\}.$$

We shall first investigate the linear term $m\sqrt{nh^m}u_n^{(1)}(\pi_{1,m}G_{\varphi,\mathbf{t}}^{(T)})$. Notice that

$$m\sqrt{nh^m}u_n^{(1)}\big(\pi_{1,m}G_{\varphi,\mathbf{t}}^{(T)}\big) = \frac{m\sqrt{h^m}}{\sqrt{n}}\sum_{i=1}^{n}\pi_{1,m}G_{\varphi,\mathbf{t}}^{(T)}(\mathbf{X}_i, \mathbf{Y}_i).$$

We can write

$$\pi_{1,m}G^{(T)}_{\varphi,\mathbf{t}}(x,y) = \mathbb{E}\big[G^{(T)}_{\varphi,\mathbf{t}}(x,X_2,\cdots,X_m),(y,X_2,\cdots,X_m)\big] - \mathbb{E}\big[G^{(T)}_{\varphi,\mathbf{t}}(\mathbf{X},\mathbf{Y})\big]$$
$$= \mathbb{E}\big[G^{(T)}_{\varphi,\mathbf{t}}(\mathbf{X},\mathbf{Y}) \mid (X_1,Y_1)=(x,y)\big] - \mathbb{E}\big[G^{(T)}_{\varphi,\mathbf{t}}(\mathbf{X},\mathbf{Y})\big].$$

For notational brevity, we use a function introduced in [16]:

$$S_{\varphi,\mathbf{t}} : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$$
$$(x,y) \longrightarrow m\mathbb{E}\Big[\varphi(y)\widetilde{K}\Big(\frac{t-x}{h}\Big) \mid (X_1,Y_1)=(x,y)\Big].$$

Hence we get

$$m\pi_{1,m}G^{(T)}_{\varphi,\mathbf{t}}(x,y) = h^{-m}\big(S_{\varphi,\mathbf{t}}(x,y) - \mathbb{E}[S_{\varphi,\mathbf{t}}(x,y)]\big).$$

The linear term of the process is given by

$$m\sqrt{nh^m}u^{(1)}_n(\pi_{1,m}G^{(T)}_{\varphi,\mathbf{t}}) = \frac{1}{\sqrt{nh^m}}\sum_{i=1}^{n}\big\{S_{\varphi,\mathbf{t}}(X_i,Y_i) - \mathbb{E}[S_{\varphi,\mathbf{t}}(X_i,Y_i)]\big\} := \alpha_n(S_{\varphi,\mathbf{t}}).$$

Therefore the linear term of the U-process $\{\mu_n(\varphi,\mathbf{t})\}_{\mathscr{F}_m\mathscr{K}^m}$ is an empirical process indexed by the class of functions $\mathcal{S}$ defined by

$$\mathcal{S} = \big\{S_{\varphi,\mathbf{t}}(\cdot,\cdot)\ \varphi\in\mathscr{F}_m,\ \mathbf{t}=(t_1,\ldots,t_m)\in\mathbb{I}\big\},$$

therefore its weak convergence may be established in a similar way to the proof of Theorem 2.2.2. It is clear that $\mathcal{S}\subset m\mathcal{G}^{(1)}$. Consider now the nonlinear part. We have to show that

$$\big\|\sqrt{nh^m}u^{(k)}_n(\pi_{k,m}G^{(T)}_{\varphi,\mathbf{t}})\big\|_{\mathscr{F}_m\mathscr{K}^m} \xrightarrow{\mathbb{P}} 0, \quad \text{for} \quad 2\le k\le m.$$

This is a consequence of Lemma 7.0.2, noting that to get the convergence of the truncated term we will need to choose the parameters a_n, $b_n\upsilon_n$ in a way that the terms I–VI converge toward 0.

Let us investigate now the remainder part $\mu^{(R)}_n(\varphi,\mathbf{t})$. Our main goal is to prove that

$$\mathbb{P}\big\{\|\mu^{(R)}_n(\varphi,\mathbf{t})\|_{\mathscr{F}_m\mathscr{K}^m} > \lambda\big\} \to 0 \quad \text{as} \quad n\to\infty.$$

For clarity we restrict ourselves to $m=2$. We have

$$\begin{aligned}
\mu^{(R)}_n(\varphi,\mathbf{t}) &= \sqrt{nh^2}\big\{u^{(R)}_n(\varphi,\mathbf{t}) - \mathbb{E}(u^{(R)}_n(\varphi,\mathbf{t}))\big\}\\
&= \frac{\sqrt{nh^2}}{n(n-1)}\sum_{i\neq j}^{n}\big\{G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j)) - \mathbb{E}\big[G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j))\big]\big\}\\
&\le \frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H^{(U)}_p}\sum_{j\in H^{(U)}_q} h^2\big\{G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j)) - \mathbb{E}\big[G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j))\big]\big\}\\
&+\frac{1}{\sqrt{nh^2}}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in H^{(U)}_p} h^2\big\{G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j)) - \mathbb{E}\big[G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j))\big]\big\}\\
&+2\frac{1}{\sqrt{nh^2}}\sum_{p=1}^{\upsilon_n}\sum_{i\in H^{(U)}_p}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T^{(U)}_q} h^2\big\{G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j)) - \mathbb{E}\big[G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j))\big]\big\}\\
&+2\frac{1}{\sqrt{nh^2}}\sum_{p=1}^{\upsilon_n}\sum_{i\in H^{(U)}_p}\sum_{q:|q-p|\le 1}^{\upsilon_n}\sum_{j\in T^{(U)}_q} h^2\big\{G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j)) - \mathbb{E}\big[G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j))\big]\big\}\\
&+\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in T^{(U)}_p}\sum_{j\in T^{(U)}_q} h^2\big\{G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j)) - \mathbb{E}\big[G^{(R)}_{\varphi,\mathbf{t}}((X_i,X_j),(Y_i,Y_j))\big]\big\}
\end{aligned}$$

$$+\frac{1}{\sqrt{nh^2}}\sum_{p=1}^{\upsilon_n}\sum_{i\neq j\, i,j\in T_p^{(U)}} h^2\big\{G_{\varphi,\mathbf{t}}^{(R)}((X_i,X_j),(Y_i,Y_j))-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((X_i,X_j),(Y_i,Y_j))\big]\big\}$$
$$=: \mathrm{I}'+\mathrm{II}'+\mathrm{III}'+\mathrm{IV}'+\mathrm{V}'+\mathrm{VI}'.$$

We will use blocking arguments and treat the resulting terms. We start by considering the first I'. We have

$$\mathbb{P}\Big\{\Big\|\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}} h^2\big\{G_{\varphi,\mathbf{t}}^{(R)}((X_i,X_j),(Y_i,Y_j))-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((X_i,X_j),(Y_i,Y_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2}>\delta\Big\}$$
$$\leq \mathbb{P}\Big\{\Big\|\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}} h^2\big\{G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2}>\delta\Big\}$$
$$+2\upsilon_n\beta_{b_n}.$$

Notice that (3.4) readily implies that $\upsilon_n\beta_{b_n}\to 0$ and recall that for all $\varphi\in\mathscr{F}_m$ and $\mathbf{x},\mathbf{y},\mathbf{t}\in\mathbb{R}^2$,

$$\kappa^2F(\mathbf{y})\geqslant\varphi(\mathbf{y})\widetilde{K}\Big(\frac{\mathbf{x}-\mathbf{t}}{h}\Big).$$

By symmetry of the function $F(\cdot)$, it holds that:

$$\Big\|\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}} h^2\big\{G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2}$$
$$\leq\Big|\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\big\{\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}$$
$$-\mathbb{E}\big[\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]\big\}\Big|, \tag{7.37}$$

hence we have to investigate the following probability:

$$\mathbb{P}\Big\{\Big|\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\big\{\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}$$
$$-\mathbb{E}\big[\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]\big\}\Big|>\delta\Big\}.$$

We apply respectively Chebyshev's inequality, Hoeffding's trick, and Hoeffding's inequality to get

$$\mathbb{P}\Big\{\Big|\frac{1}{\sqrt{nh^2}}\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\Big\{\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}$$
$$-\mathbb{E}\big[\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]\Big\}\Big|>\delta\Big\}$$
$$\leq\delta^{-2}n^{-1}h^{-2}\mathrm{Var}\Big(\sum_{p\neq q}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{j\in H_q^{(U)}}\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\Big)$$
$$\leq c_2\upsilon_n\delta^{-2}n^{-1}h^{-2}\mathrm{Var}\Big(\sum_{p=1}^{\upsilon_n}\sum_{i,j\in H_p^{(U)}}\kappa^2F(\zeta_i,\zeta_j')\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\Big)$$
$$\leq 2c_2\upsilon_n\delta^{-2}n^{-2}h^{-2}\mathbb{E}\big[(\kappa^2F(\zeta_1,\zeta_2))^2\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]. \tag{7.38}$$

Since the moment condition (3.3) is fulfilled, we have for each $\lambda>0$,

$$c_2\upsilon_n\delta^{-2}n^{-2}h^{-2}\mathbb{E}\big[\big(\kappa^2F(\zeta_1,\zeta_2)\big)^2\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]$$

$$= c_2\upsilon_n\delta^{-2}n^{-2}h^{-2}\int_0^\infty \mathbb{P}\big\{\big(\kappa^2F(\zeta_1,\zeta_2)\big)^2\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}} \geqslant t\big\}\,dt$$
$$= c_2\upsilon_n\delta^{-2}n^{-2}h^{-2}\int_0^{\lambda(n/h^m)^{1/2(p-1)}} \mathbb{P}\big\{\kappa^2F > \lambda(n/h^m)^{1/2(p-1)}\big\}\,dt$$
$$+c_2\upsilon_n\delta^{-2}n^{-2}h^{-2}\int_{\lambda(n/h^m)^{1/2(p-1)}}^\infty \mathbb{P}\big\{\big(\kappa^2F\big)^2 > t\big\}\,dt,$$

which tends to 0 as $n\to\infty$. The terms II′, V′ and VI′ are treated in the same way as the first, except that for II′, VI′ we do not need to apply Hoeffding's trick because our variables $\{\zeta_i,\zeta_j\}_{i,j\in H_p^{(U)}}$ (or $\{\zeta_i,\zeta_j\}_{i,j\in T_p^{(U)}}$ for VI′) are in the same blocks, and for the term IV′ we deduce its study from those of I′ and III′. Let us consider the term III′. As for the truncated part, we have

$$\mathbb{P}\Big\{\Big\|\frac{1}{\sqrt{nh^2}}\sum_{p=1}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}} h^2\big\{G_{\varphi,\mathbf{t}}^{(R)}((X_i,X_j),(Y_i,Y_j))$$
$$-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((X_i,X_j),(Y_i,Y_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2} > \delta\Big\}$$
$$\le \mathbb{P}\Big\{\Big\|\frac{1}{\sqrt{nh^2}}\sum_{p=1}^{\upsilon_n}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}} \big\{G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))$$
$$-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2} > \delta\Big\} + \frac{\upsilon_n^2a_nb_n\beta_{a_n}}{\sqrt{nh^2}}.$$

We also have

$$\mathbb{P}\Big\{\Big\|\frac{1}{\sqrt{nh^2}}\sum_{i\in H_p^{(U)}}\sum_{q:|q-p|\geqslant 2}^{\upsilon_n}\sum_{j\in T_q^{(U)}} \big\{G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))$$
$$-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2} > \delta\Big\}$$
$$\le \mathbb{P}\Big\{\Big\|\frac{1}{\sqrt{nh^2}}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}} \big\{G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))$$
$$-\mathbb{E}\big[G_{\varphi,\mathbf{t}}^{(R)}((\varsigma_i,\varsigma_j),(\zeta_i,\zeta_j))\big]\big\}\Big\|_{\mathscr{F}_2\mathscr{K}^2} > \delta\Big\}.$$

Since the equation (7.37) is still satisfied, the problem is reduced to

$$\mathbb{P}\Big\{\Big|\frac{1}{\sqrt{nh^2}}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}} \big\{\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}$$
$$-\mathbb{E}\big[\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]\big\}\Big| > \delta\Big\}.$$

We have the following bound:

$$\mathbb{P}\Big\{\Big|\frac{1}{\sqrt{nh^2}}\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in T_q^{(U)}} \big\{\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}$$
$$-\mathbb{E}\big[\kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\big]\big\}\Big| > \delta\Big\}$$
$$\le \delta^{-2}n^{-1}h^{-2}\mathrm{Var}\Big(\sum_{i\in H_1^{(U)}}\sum_{q=3}^{\upsilon_n}\sum_{j\in H_q^{(U)}} \kappa^2F(\zeta_i,\zeta_j)\mathbf{1}_{\{\kappa^2F>\lambda(n/h^m)^{1/2(p-1)}\}}\Big),$$

we follow the same procedure as in (7.38). The rest has just been shown to be asymptotically negligible, so the process $\{\mu_n(\varphi,\mathbf{t})\}_{\mathscr{F}_m\mathscr{K}^m}$ converges in law to a Gaussian process which has a version with uniformly bounded and uniformly continuous paths with respect to $\|\cdot\|_2$-norm. We treat $\{\mu_n(1,\mathbf{t})\}_{\mathscr{K}^m}$ in a similar way, and the treatment of $\mathbb{E}(u_n(\varphi,\mathbf{t}))$, $u_n(1,\mathbf{t})$ and $\mathbb{E}(u_n(1,\mathbf{t}))$ is done as in the proof of Theorem 2.2.2. □

Remark.

- In the treatment of the remainder part, we think that we could choose alternative blocks of the same size.
- For any measurability question regarding the calculation of the probability tails when dealing either with the conditional empirical process or the conditional U-process, we invite the reader to check [71], p. 110.

APPENDIX

To estimate the dependence between two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ defined on a probability space $(\Omega,\mathcal{D},P)$, we are going to use some classical measures of dependence (see, for example, [10]):

$$\alpha(\mathcal{A},\mathcal{B}) := \sup\left|P(A\cap B)-P(A)P(B)\right|,\quad A\in\mathcal{A},\quad B\in\mathcal{B},$$

$$\beta(\mathcal{A},\mathcal{B}) := \sup \tfrac{1}{2}\textstyle\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P(A_i\cap B_j)-P(A_i)P(B_j)\right|,$$

$$\phi(\mathcal{A},\mathcal{B}) := \sup|P(B|A)-P(A)P(B)|,\quad A\in\mathcal{A},\quad B\in\mathcal{B},\quad P(A)>0,$$

$$\psi(\mathcal{A},\mathcal{B}) := \sup\left|P(A\cap B)-P(A)P(B)\right|/P(A)P(B),\quad A\in\mathcal{A},\quad B\in\mathcal{B},$$

where for $\beta(\mathcal{A},\mathcal{B})$, the sup is taken over all pairs of finite partitions $\{A_1,\ldots,A_I\}$ and $\{B_1,\ldots,B_J\}$ of Ω such that $A_i\in\mathcal{A}$ for all $1\le i\le I$ and $B_j\in\mathcal{B}$ for all $1\le j\le J$. Let

$$\sigma_J^L := \sigma(\mathbf{Z}_i, J\le i\le L),$$

a $*$-mixing sequence is defined by requiring the $*$-mixing coefficient $*_k$ to satisfy

$$*_k := \sup_{J\in Z} *(\sigma_{-\infty}^{J},\sigma_{J+k}^{\infty}) \underset{k\to\infty}{\longrightarrow} 0.$$

Definition 7.0.3. A class of subsets $\mathcal{C}$ 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 $\mathcal{C}$ picks out at most $P(N)$ distinct subsets.

Definition 7.0.4. A class of functions $\mathscr{F}$ is called a VC subgraph class if the graphs of the functions in $\mathscr{F}$ form a VC class of sets, that is, if we define the subgraph of a real-valued function f on some space S as the following subset $\mathcal{G}_f$ on $S\times\mathbb{R}$:

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

the class $\{\mathcal{G}_f\colon f\in\mathscr{F}\}$ is a VC class of sets on $S\times\mathbb{R}$.

Lemma 7.0.5 (Bochner) *Let $G\colon(\mathbb{R},\mathcal{B}_\mathbb{R})\to(\mathbb{R},\mathcal{B}_\mathbb{R})$ be a bounded integrable function such that*

$$|z|G(z) \overset{|z|\to\infty}{\longrightarrow} 0,$$

and $g\colon(\mathbb{R},\mathcal{B}_\mathbb{R})\to(\mathbb{R},\mathcal{B}_\mathbb{R})$ an integrable function. Set

$$g_n(x) = \frac{1}{h_n}\int_{\mathbb{R}} G\Big(\frac{z}{h_n}\Big)g(x-z)\,dz,$$

where $0<h_n\to 0$ as $n\to\infty$. If $g(\cdot)$ is continuous at the point x, then

$$\lim_{n\to\infty} g_n(x) = g(x)\int_{-\infty}^{+\infty} G(z)\,dz.$$

Further, if $g(\cdot)$ is uniformly continuous the convergence is uniform.

The Nadaraya–Watson estimator of regression is defined by

$$\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n) = \frac{\sum_{i=1}^n \mathbf{Y}_i K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h_n}\right)},$$

where Id denote the identity function. In our analysis, we shall consider another, but more appropriate and more computationally convenient, centering factor than the expectation $\mathbb{E}\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n)$, which is delicate to handle. This is given by

$$\widetilde{\mathbb{E}}[\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n)] = \frac{\mathbb{E}\left(\sum_{i=1}^n \mathbf{Y}_i K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h_n}\right)\right)}{\mathbb{E}\left(\sum_{i=1}^n K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h_n}\right)\right)}.$$

Proposition 7.0.6. *If $\mathbf{Y}$ is a bounded random variable and $nh \to \infty$, $h \to 0$, then*

$$\mathbb{E}[\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n)] = \widetilde{\mathbb{E}}[\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n)] + O((nh)^{-1}).$$

If $\mathbb{E}(\mathbf{Y}^2) < \infty$ and $nh^2 \to \infty$, then we have

$$\mathbb{E}[\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n)] = \widetilde{\mathbb{E}}[\widehat{r}_n^{(1)}(\mathrm{Id}, \mathbf{t}, h_n)] + O((n^{\frac{1}{2}}h)^{-1}).$$

Proposition 7.0.7. (see [3], Proposition 3.6). *Let $\{\mathbf{X}_t : t \in \boldsymbol{T}\}$ be a process satisfying, for $m \geqslant 1$,*

$$\left(\mathbb{E}\|\mathbf{X}_t - \mathbf{X}_s\|^p\right)^{1/p} \le \left(\frac{p-1}{q-1}\right)^{m/2} \left(\mathbb{E}\|\mathbf{X}_t - \mathbf{X}_s\|^q\right)^{1/q}, \quad 1 < q < p < \infty,$$

and the semi-metric

$$\rho(s,t) = \left(\mathbb{E}\|\mathbf{X}_t - \mathbf{X}_s\|^2\right)^{1/2}.$$

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

$$\mathbb{E} \sup_{s,t \in \boldsymbol{T}} \|\mathbf{X}_t - \mathbf{X}_s\| \le K \int_0^D [\log N(\epsilon, \boldsymbol{T}, \rho)]^{m/2} \, d\epsilon,$$

D being the ρ-diameter of $\boldsymbol{T}$.

ACKNOWLEDGMENTS

We would like to acknowledge the reviewer for detailed and useful comments which led to a more sharply focused presentation.

REFERENCES

1. J. Abrevaya and W. Jiang, "A Nonparametric Approach to Measuring and Testing Curvature", J. Bus. Econom. Statist. **23** (1), 1–19 (2005).
2. H. Akaike, "An Approximation to the Density Function", Ann. Inst. Statist. Math., Tokyo **6**, 127–132 (1954).
3. M. A. Arcones and E. Giné, "Limit Theorems for U-Processes", Ann. Probab. **21** (3), 1494–1542 (1993).
4. M. A. Arcones and E. Giné, "On the Law of the Iterated Logarithm for Canonical U-Statistics and Processes", Stochastic Process. Appl. **58** (2), 217–245 (1995).
5. M. A. Arcones and Y. Wang, "Some New Tests for Normality Based on U-Processes", Statist. Probab. Lett. **76** (1), 69–82 (2006).
6. M. A. Arcones and B. Yu, "Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences", J. Theoret. Probab. **7** (1), 47–71 (1994).
7. M. A. Arcones, Z. Chen, and E. Giné, "Estimators Related to U-Processes with Applications to Multivariate Medians: Asymptotic Normality", Ann. Statist. **22** (3), 1460–1477 (1994).
8. S. Borovkova, R. Burton, and H. Dehling, "Limit Theorems for Functionals of Mixing Processes with Applications to U-Statistics and Dimension Estimation", Trans. Amer. Math. Soc. **353** (11), 4261–4318 (2001).

9. Y. V. Borovskikh, *U-Statistics in Banach Spaces* (VSP, Utrecht, 1996).
10. R. C. Bradley, “Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions”, Probab. Surv. **2**, 107–144 (2005). Update of and a supplement to the 1986 original.
11. V. H. de la Peña, “Decoupling and Khintchin’s Inequalities for U-Statistics”, Ann. Probab. **20** (4), 1877–1892 (1992).
12. V. H. de la Peña and E. Giné, *Decoupling*, in *Probability and its Applications (New York)* (Springer-Verlag, New York, 1999). From dependence to independence. Randomly stopped processes. U-statistics and processes. Martingales and beyond.
13. P. Deheuvels, “One Bootstrap Suffices to Generate Sharp Uniform Bounds in Functional Estimation”, Kybernetika (Prague) **47** (6), 855–865 (2011).
14. P. Deheuvels and D. M. Mason, “General Asymptotic Confidence Bands Based on Kernel-Type Function Estimators”, Statist. Inference Stoch. Process. **7** (3), 225–277 (2004).
15. M. Denker and G. Keller, “On U-Statistics and von Mises’ Statistics for Weakly Dependent Processes”, Z. Wahrsch. Verw. Gebiete **64** (4), 505–522 (1983).
16. J. Dony and D. M. Mason, “Uniform in Bandwidth Consistency of Conditional U-Statistics”, Bernoulli **14** (4), 1108–1133 (2008).
17. P. Doukhan, P. Massart, and E. Rio, “The Functional Central Limit Theorem for Strongly Mixing Processes”, Ann. Inst. H. Poincaré Probab. Statist. **30** (1), 63–82 (1994).
18. R. M. Dudley, “The Sizes of Compact Subsets of Hilbert Space and Continuity of Gaussian Processes”, J. Functional Analysis **1**, 290–330 (1967).
19. R. M. Dudley, *Uniform Central Limit Theorems*, in *Cambridge Studies in Advanced Mathematics* (Cambridge Univ. Press, Cambridge, 1999), Vol. 63.
20. A. Dvoretzky, “Asymptotic Normality for Sums of Dependent Random Variables”, pp. 513–535 (1972).
21. E. Eberlein, “Weak Convergence of Partial Sums of Absolutely Regular Sequences”, Statist. Probab. Lett. **2** (5), 291–293 (1984).
22. U. Einmahl and D. M. Mason, “An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators”, J. Theoret. Probab. **13** (1), 1–37 (2000).
23. U. Einmahl and D. M. Mason, “Uniform in Bandwidth Consistency of Kernel-Type Function Estimators”, Ann. Statist. **33** (3), 1380–1403 (2005).
24. E. W. Frees, “Infinite Order U-Statistics”, Scand. J. Statist. **16** (1), 29–45 (1989).
25. S. Ghosal, A. Sen, and A. W. van der Vaart, “Testing Monotonicity of Regression”, Ann. Statist. **28** (4), 1054–1082 (2000).
26. E. Giné and D. M. Mason, “Laws of the Iterated Logarithm for the Local U-Statistic Process”, J. Theoret. Probab. **20** (3), 457–485 (2007a).
27. E. Giné and D. M. Mason, “On Local U-Statistic Processes and the Estimation of Densities of Functions of Several Sample Variables”, Ann. Statist. **35** (3), 1105–1145 (2007b).
28. E. Giné and J. Zinn, “Some Limit Theorems for Empirical Processes”, Ann. Probab. **12** (4), 929–998 (1984). With discussion.
29. P. R. Halmos, “The Theory of Unbiased Estimation”, Ann. Math. Statist. **17**, 34–43 (1946).
30. M. Harel and M. L. Puri, “Conditional U-Statistics for Dependent Random Variables”, J. Multivariate Anal. **57** (1), 84–100 (1996).
31. C. Heilig and D. Nolan, “Limit Theorems for the Infinite-Degree U-Process”, Statist. Sinica **11** (1), 289–302 (2001).
32. W. Hoeffding, “A Class of Statistics with Asymptotically Normal Distribution”, Ann. Math. Statistics **19**, 293–325 (1948).
33. J. Hoffmann-Jørgensen, “Convergence of Stochastic Processes on Polish Spaces”, (1984). Unpublished.
34. M. Hollander and F. Proschan, “Testing Whether New is Better Than Used”, Ann. Math. Statist. **43**, 1136–1146 (1972).
35. I. A. Ibragimov, “Some Limit Theorems for Stationary Processes”, Teor. Verojatnost. i Primenen. **7**, 361–392 (1962).
36. E. Joly and G. Lugosi, “Robust Estimation of U-Statistics”, Stochastic Process. Appl. **126** (12), 3760–3773 (2016).
37. M. G. Kendall, “A New Measure of Rank Correlation”, Biometrika **30** (1/2), 81–93 (1938).
38. A. N. Kolmogorov and V. M. Tihomirov, “ε-Entropy and ε-Capacity of Sets in Functional Space”, Amer. Math. Soc. Transl. (2) **17**, 277–364 (1961).
39. V. S. Koroljuk and Y. V. Borovskich, *Theory of U-Statistics*, in *Mathematics and its Applications* (Kluwer Academic Publishers Group, Dordrecht, 1994), Vol. 273. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.
40. M. R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference*, in *Springer Series in Statistics* (Springer, New York, 2008).

41. L. Le Cam, "A Remark on Empirical Measures", In a Festschrift for Erich L. Lehmann, Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA (1983), pp. 305–327.
42. A. J. Lee, *U-Statistics*, in *Statistics: Textbooks and Monographs* (Marcel Dekker Inc., New York, 1990), Vol. 110. Theory and practice.
43. S. Lee, O. Linton, and Y.-J. Whang, "Testing for Stochastic Monotonicity", Econometrica **77** (2), 585–602 (2009).
44. A. Leucht, "Degenerate U- and V-Statistics under Weak Dependence: Asymptotic Theory and Bootstrap Consistency", Bernoulli **18** (2), 552–585 (2012).
45. A. Leucht and M. H. Neumann, "Degenerate U- and V-Statistics under Ergodicity: Asymptotics, Bootstrap and Applications in Statistics", Ann. Inst. Statist. Math. **65** (2), 349–386 (2013).
46. E. A. Nadaraja, "On a Regression Estimate", Teor. Verojatnost. i Primenen. **9**, 157–159 (1964).
47. D. Nolan and D. Pollard, "U-Processes: Rates of Convergence", Ann. Statist. **15** (2), 780–799 (1987).
48. E. Parzen, "On Estimation of a Probability Density Function and Mode", Ann. Math. Statist. **33**, 1065–1076 (1962).
49. D. Pollard, *Convergence of Stochastic Processes*, in *Springer Series in Statistics* (Springer-Verlag, New York, 1984).
50. W. Polonik and Q. Yao, "Set-Indexed Conditional Empirical and Quantile Processes Based on Dependent Data", J. Multivariate Anal. **80** (2), 234–255 (2002).
51. D. V. Poryvaĭ, "An Invariance Principle for Conditional Empirical Processes Formed by Dependent Random Variables", Izv. Ross. Akad. Nauk Ser. Mat. **69** (4), 129–148 (2005).
52. B. L. S Prakasa Rao and A. Sen, "Limit Distributions of Conditional U-Statistics", J. Theoret. Probab. **8** (2), 261–301 (1995).
53. G. Rempala and A. Gupta, "Weak Limits of U-Statistics of Infinite Order", Random Oper. Stochastic Equations **7** (1), 39–52 (1999).
54. M. Rosenblatt, "A Central Limit Theorem and a Strong Mixing Condition", Proc. Nat. Acad. Sci. U.S.A. **42**, 43–47 (1956).
55. A. Schick, Y. Wang, and W. Wefelmeyer, "Tests for Normality Based on Density Estimators of Convolutions", Statist. Probab. Lett. **81** (2), 337–343 (2011).
56. A. Sen, "Uniform Strong Consistency Rates for Conditional U-Statistics", Sankhyā Ser. A **56** (2), 179–194 (1994).
57. R. J. Serfling, *Approximation Theorems of Mathematical Statistics* (John Wiley & Sons, Inc., New York, 1980). Wiley Series in Probability and Mathematical Statistics.
58. R. P. Sherman, "The Limiting Distribution of the Maximum Rank Correlation Estimator", Econometrica **61** (1), 123–137 (1993).
59. R. P. Sherman, "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators", Ann. Statist. **22** (1), 439–459 (1994).
60. B. W. Silverman, "Distances on Circles, Toruses and Spheres", J. Appl. Probability **15** (1), 136–143 (1978).
61. Y. Song, X. Chen, and K. Kato, *Approximating High-Dimensional Infinite-Order U-Statistics: Statistical and Computational Guarantees* (2019), arXiv e-prints, page arXiv:1901.01163.
62. W. Stute, "Conditional Empirical Processes", Ann. Statist. **14** (2), 638–647 (1986).
63. W. Stute, "Conditional U-Statistics", Ann. Probab. **19** (2), 812–825 (1991).
64. W. Stute, "Almost Sure Representations of the Product-Limit Estimator for Truncated Data", Ann. Statist. **21** (1), 146–156 (1993).
65. W. Stute, "Symmetrized NN-Conditional U-Statistics", in *Research Developments in Probability and Statistics*, (VSP, Utrecht, 1996), pp. 231–237.
66. A. van der Vaart, "New Donsker Classes", Ann. Probab. **24** (4), 2128–2140 (1996).
67. A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, in *Springer Series in Statistics* (Springer-Verlag, New York, 1996).
68. R. von Mises, On the Asymptotic Distribution of Differentiable Statistical Functions. Ann. Math. Statist. **18**, 309–348 (1947).
69. G. S. Watson, "Smooth Regression Analysis", Sankhyā Ser. A **26**, 359–372 (1964).
70. K.-i. Yoshihara, "Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes", Z. Wahrsch. und Verw. Gebiete **35** (3), 237–252 (1976).
71. B. Yu, "Rates of Convergence for Empirical Processes of Stationary Mixing Sequences", Ann. Probab. **22** (1), 94–116 (1994).
72. D. Zhang, "Bayesian Bootstraps for U-Processes, Hypothesis Tests and Convergence of Dirichlet U-Processes", Statist. Sinica **11** (2), 463–478 (2001).
73. K. Ziegler, "On the Asymptotic Normality of Kernel Regression Estimators of the Mode in the Nonparametric Random Design Model", J. Statist. Plann. Inference **115** (1), 123–144 (2003).