Bernoulli **18**(2), 2012, 552–585
DOI: 10.3150/11-BEJ354

Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency

ANNE LEUCHT

Friedrich-Schiller-Universität Jena, Institut für Stochastik, Ernst-Abbe-Platz 2, D-07743 Jena, Germany. E-mail: anne.leucht@uni-jena.de

We devise a general result on the consistency of model-based bootstrap methods for U- and V-statistics under easily verifiable conditions. For that purpose, we derive the limit distributions of degree-2 degenerate U- and V-statistics for weakly dependent $\mathbb{R}^d$-valued random variables first. To this end, only some moment conditions and smoothness assumptions concerning the kernel are required. Based on this result, we verify that the bootstrap counterparts of these statistics have the same limit distributions. Finally, some applications to hypothesis testing are presented.

Keywords: bootstrap; consistency; U-statistics; V-statistics; weak dependence

1. Introduction

Numerous test statistics can be formulated or approximated in terms of degenerate U- or V-type statistics. Examples include the Cramér–von Mises statistic, the Anderson–Darling statistic or the χ^2-statistic. For i.i.d. random variables the limit distributions of U- and V-statistics can be derived via a spectral decomposition of their kernel if the latter is squared integrable. To use the same method for dependent data, often restrictive assumptions are required whose validity is quite complicated or even impossible to verify in many cases. The first of our two main results is the derivation of the asymptotic distributions of U- and V-statistics under assumptions that are fairly easy to check. This approach is based on a wavelet decomposition instead of a spectral decomposition of the kernel.

The limit distributions for both independent and dependent observations depend on certain parameters which in turn depend on the underlying situation in a complicated way. Therefore, problems arise as soon as critical values for test statistics of U- and V-type have to be determined. The bootstrap offers a convenient way to circumvent these problems; see Arcones and Giné [2], Dehling and Mikosch [10] or Leucht and Neumann [25] for the i.i.d. case. To our knowledge, there are no results concerning bootstrapping general degenerate U-statistics of non-independent observations. As a second main result of the paper, we establish consistency of model-based bootstrap methods for U- and V-type statistics of weakly dependent data.

In order to describe the dependence structure of the sample, we do not invoke the concept of mixing although a great variety of processes satisfy these constraints and various tools of probability theory and statistics such as central limit theorems, probability and moment inequalities can be carried over from the i.i.d. setting to mixing processes. However, these methods of

1350-7265

measuring dependencies are inappropriate in the present context since not only the asymptotic behaviour of U- and V-type statistics but also bootstrap consistency is focused. Model-based bootstrap methods can yield samples that are no longer mixing even though the original sample satisfies some mixing condition. A simple example is presented in Section 4.2. There we consider a model-specification test within the class of nonlinear AR(1) processes. Under $\mathcal{H}_0$, $X_k = g_0(X_{k-1}) + \varepsilon_k$, where g_0 is Lipschitz contracting and $(\varepsilon_k)_k$ is a sequence of i.i.d. centered innovations. It is most natural to draw the bootstrap innovations $(\varepsilon_k^*)_k$ via Efron's bootstrap from the recentered residuals first. Then the bootstrap counterpart of $(X_k)_k$ is generated iteratively by choosing an initial variable X_0^* independently of $(\varepsilon_k^*)_k$ and defining $X_k^* = g_0(X_{k-1}^*) + \varepsilon_k^*$. Due to the discreteness of the bootstrap innovations, commonly used coupling techniques to prove mixing properties for Markovian processes fail; see also Andrews [1]. It turns out that the characterization of dependence structures introduced by Dedecker and Prieur [9] is exceptionally suitable here. Based on their τ-dependence coefficient it is possible to construct an L_1-coupling in the following sense. Let $\mathcal{M}$ denote a σ-algebra generated by sample variables of the "past" and let X be a random variable of a certain "future" time point. Then, the minimal L_1-distance between X and a random variable that has the same distribution as X but that is independent of $\mathcal{M}$ is equivalent to the τ-dependence coefficient $\tau(\mathcal{M}, X)$.

We exploit this coupling property in order to derive the asymptotic distribution for the original as well as the bootstrap statistics of degenerate U-type. Basically, both proofs follow the same lines. First, the (almost) Lipschitz continuous kernels of the U-statistics are approximated by a finite wavelet series expansion. There are two crucial points that assure asymptotic negligibility of the approximation error. On the one hand, the smoothness of the kernel function carries over to its wavelet approximation uniformly in scale, cf. Lemma 5.2. On the other hand, Lipschitz continuity of the kernel and the L_1-coupling property of the underlying τ-dependent sample perfectly fit together. A next step contains the application of a central limit theorem and the continuous mapping theorem to determine the limits of the approximating statistics of U-type. Based on these investigations, the asymptotic distribution of the U-statistic and its bootstrap counterpart is then deduced via passage to the limit. It can be expressed as an infinite weighted sum of normal variables.

Our paper is organized as follows. We start with an overview of asymptotic results on degenerate U-type statistics of dependent random variables. In Section 2.2, we introduce the underlying concept of weak dependence and derive the asymptotic distributions of U- and V-statistics. On the basis of these results, we deduce consistency of general bootstrap methods in Section 3. Some applications of the theory to hypothesis testing are presented in Section 4. All proofs are deferred to a final Section 5.

2. Asymptotic distributions of U- and V-statistics

2.1. Survey of literature

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variables with common distribution P_X. In the case of i.i.d. random variables, the limit distributions of degenerate U- and V-type statistics, that is,

$$nU_n = \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j} h(X_j, X_k) \quad \text{and} \quad nV_n = \frac{1}{n}\sum_{j,k=1}^{n} h(X_j, X_k),$$

with h: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ symmetric and $\int_{\mathbb{R}^d} h(x, y)P_X(\mathrm{d}x) = 0, \forall y \in \mathbb{R}^d$, can be derived by using a spectral decomposition of the kernel, $h(x, y) = \sum_{k=1}^{\infty} \lambda_k \Phi_k(x)\Phi_k(y)$, which holds true in the L_2-sense. Here, $(\Phi_k)_k$ denote orthonormal eigenfunctions and $(\lambda_k)_k$ the corresponding eigenvalues of the integral equation

$$\int_{\mathbb{R}^d} h(x, y)g(y)P_X(\mathrm{d}y) = \lambda g(x). \tag{2.1}$$

Approximate nU_n by $nU_n^{(K)} = \sum_{k=1}^{K} \lambda_k\{(n^{-1/2}\sum_{i=1}^{n} \Phi_k(X_i))^2 - n^{-1}\sum_{i=1}^{n} \Phi_k^2(X_i)\}$. Then the sum under the round brackets is asymptotically standard normal while the latter sum converges in probability to 1. Finally, one obtains

$$nU_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k(Z_k^2 - 1), \tag{2.2}$$

where $(Z_k)_k$ is a sequence of i.i.d. standard normal random variables; cf. Serfling [27]. If additionally $\mathbb{E}|h(X_1, X_1)| < \infty$, the weak law of large numbers and Slutsky's theorem imply $V_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k(Z_k^2 - 1) + \mathbb{E}h(X_1, X_1)$. (Here, $\xrightarrow{d}$ denotes convergence in distribution.)

So far, most previous attempts to derive the limit distributions of degenerate U- and V-statistics of dependent random variables are based on the adoption of this method of proof. Eagleson [15] developed the asymptotic theory in the case of a strictly stationary sequence of ϕ-mixing, real-valued random variables under the assumption of absolutely summable eigenvalues. This condition is satisfied if the kernel function is of the form $h(x, y) = \int_{\mathbb{R}} h_1(x, z)h_1(z, y)P_X(\mathrm{d}z)$ and h_1 is squared integrable w.r.t. P_X. Using general heavy-tailed weight functions instead of P_X, the eigenvalues are not necessarily absolutely summable; see, for example, de Wet [7]. Carlstein [5] analysed U-statistics of α-mixing, real-valued random variables in the case of finitely many eigenfunctions. He derived a limit distribution of the form (2.2), where $(Z_k)_{k\in\mathbb{N}}$ is a sequence of centered normal random variables. Denker [11] considered stationary sequences $(X_n = f(Y_n, Y_{n+1}, \ldots))_n$ of functionals of β-mixing random variables $(Y_n)_n$. He assumed f and the cumulative distribution function of X_1 to be Hölder continuous. Imposing some smoothness condition on h, the limit distribution of nU_n was derived under the additional assumption $\|\Phi_k\|_\infty < \infty, \forall k \in \mathbb{N}$. The condition on $(\Phi_k)_k$ is difficult or even impossible to check in a multitude of cases since this requires to solve the associated integral equation (2.1). Similar difficulties occur if one wants to apply the results of Dewan and Prakasa Rao [12] or Huang and Zhang [21]. They studied U-statistics of associated, real-valued random variables. Besides the absolute summability of the eigenvalues, certain regularity conditions have to be satisfied uniformly by the eigenfunctions in order to obtain the asymptotic distribution of nU_n.

A different approach was used by Babbel [3] to determine the limit distribution of U-statistics of ϕ- and β-mixing random variables. She deduced the limit distribution via a Haar wavelet decomposition of the kernel and empirical process theory without imposing the critical conditions mentioned above. However, she presumed that $\iint h(x,y) P_{X_k,X_{k+n}}(\mathrm{d}x,\mathrm{d}y) = 0, \forall k \in \mathbb{Z}, n \in \mathbb{N}$. This assumption does in general not hold true within our applications in Section 3. Moreover, this approach is not suitable when dealing with U-statistics of τ-dependent random variables since Lipschitz continuity will be the crucial property of the (approximating) kernel in order to exploit the underlying dependence structure.

2.2. Main results

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$ with common distribution P_X. In this subsection, we derive the limit distributions of

$$nU_n = \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j} h(X_j, X_k) \quad \text{and} \quad nV_n = \frac{1}{n}\sum_{j,k=1}^{n} h(X_j, X_k),$$

where $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a symmetric function with $\int_{\mathbb{R}^d} h(x,y) P_X(\mathrm{d}x) = 0, \forall y \in \mathbb{R}^d$. In order to describe the dependence structure of $(X_n)_{n\in\mathbb{N}}$, we recall the definition of the τ-dependence coefficient for $\mathbb{R}^d$-valued random variables of Dedecker and Prieur [9].

Definition 2.1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{M}$ a sub-σ-algebra of $\mathcal{A}$ and X an $\mathbb{R}^d$-valued random variable. Assume that $\mathbb{E}\|X\|_{l_1} < \infty$, where $\|x\|_{l_1} = \sum_{i=1}^{d} |x_i|$, and define*

$$\tau(\mathcal{M}, X) = \mathbb{E}\left(\sup_{f\in\Lambda_1(\mathbb{R}^d)}\left|\int_{\mathbb{R}^d} f(x) P_{X|\mathcal{M}}(\mathrm{d}x) - \int_{\mathbb{R}^d} f(x) P_X(\mathrm{d}x)\right|\right).$$

Here, $P_{X|\mathcal{M}}$ denotes the conditional distribution of X given $\mathcal{M}$ and $\Lambda_1(\mathbb{R}^d)$ denotes the set of 1-Lipschitz functions from $\mathbb{R}^d$ to $\mathbb{R}$.

We assume

(A1) (i) $(X_n)_{n\in\mathbb{N}}$ is a (strictly) stationary sequence of $\mathbb{R}^d$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$ with common distribution P_X and $\mathbb{E}\|X_1\|_{l_1} < \infty$.

(ii) The sequence $(\tau_r)_{r\in\mathbb{N}}$, defined by

$$\tau_r = \sup\{\tau(\sigma(X_{s_1}, \dots, X_{s_u}), (X'_{t_1}, X'_{t_2}, X'_{t_3})') \mid$$
$$u \in \mathbb{N}, s_1 \le \cdots \le s_u < s_u + r \le t_1 \le t_2 \le t_3 \in \mathbb{N}\},$$

satisfies $\sum_{r=1}^{\infty} r\tau_r^{\delta} < \infty$ for some $\delta \in (0,1)$. (Here, prime denotes the transposition.)

Remark 1. If Ω is rich enough, due to Dedecker and Prieur [8] the validity of (A1) allows for the construction of a random vector $(\widetilde{X}'_{t_1}, \widetilde{X}'_{t_2}, \widetilde{X}'_{t_3})' \stackrel{d}{=} (X'_{t_1}, X'_{t_2}, X'_{t_3})'$ that is independent of $X_{s_1}, \dots, X_{s_u}$ and such that

$$\sum_{i=1}^{3} \mathbb{E}\|\widetilde{X}_{t_i} - X_{t_i}\|_{l_1} \le \tau_r. \tag{2.3}$$

The notion of τ-dependence is more general than mixing. If, for example, $(X_n)_n$ is β-mixing, we obtain an upper bound for the dependence coefficient $\tau_r \le 6\int_0^{\beta(r)} Q_{|X_1|}(u)\,\mathrm{d}u$, where $Q_{|X_1|}(u) = \inf\{t \in \mathbb{R} | P(\|X_1\|_{l_1} > t) \le u\}, u \in [0,1]$, and $\beta(r)$ denotes the ordinary β-mixing coefficient $\beta(r) := \mathbb{E}\sup_{B\in\sigma(X_s, s\ge t+r), t\in\mathbb{Z}} |P(B|\sigma(X_s, s\le t)) - P(B)|$. This is a consequence of Remark 2 of Dedecker and Prieur [8]. Moreover, inequality (2.3) immediately implies

$$|\operatorname{cov}(h(X_{s_1},\dots,X_{s_u}), k(X_{t_1},\dots,X_{t_v}))| \le 2\|h\|_\infty \operatorname{Lip}(k)\left\lceil \frac{v}{3}\right\rceil \tau_r \tag{2.4}$$

for $s_1 \le \cdots \le s_u < s_u + r \le t_1 \le \cdots \le t_v \in \mathbb{N}$ and for all functions $h : \mathbb{R}^u \to \mathbb{R}$ and $k : \mathbb{R}^v \to \mathbb{R}$ in $\mathcal{L} := \{f : \mathbb{R}^p \to \mathbb{R}$ for some $p \in \mathbb{N}|$Lipschitz continuous and bounded$\}$. Therefore, a sequence of random variables that satisfies (A1) is $((\tau_r)_r, \mathcal{L}, \psi)$-weakly dependent in the sense of Doukhan and Louhichi [14] with $\psi(h,k,u,v) = 2\|h\|_\infty \operatorname{Lip}(k)\lceil\frac{v}{3}\rceil$. (Here and in the sequel, $\operatorname{Lip}(g)$ denotes the Lipschitz constant of a generic function g.) A list of examples for τ-dependent processes including causal linear and functional autoregressive processes is provided by Dedecker and Prieur [9].

Besides the conditions on the dependence structure of $(X_n)_{n\in\mathbb{N}}$, we make the following assumptions concerning the kernel:

(A2) (i) The kernel $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a symmetric, measurable function and degenerate under P_X, that is, $\int_{\mathbb{R}^d} h(x,y) P_X(\mathrm{d}x) = 0, \forall y \in \mathbb{R}^d$.

(ii) For a δ satisfying (A1)(ii), the following moment constraints hold true with some $\nu > (2-\delta)/(1-\delta)$ and an independent copy $\widetilde{X}_1$ of X_1:

$$\sup_{k\in\mathbb{N}} \mathbb{E}|h(X_1, X_{1+k})|^\nu < \infty \quad \text{and} \quad \mathbb{E}|h(X_1, \widetilde{X}_1)|^\nu < \infty.$$

(A3) The kernel h is Lipschitz continuous.

Using an appropriate kernel truncation, it is possible to reduce the problem of deriving the asymptotic distribution of nU_n to statistics with bounded kernel functions.

Lemma 2.1. *Suppose that* (A1), (A2), *and* (A3) *are fulfilled. Then there exists a family of bounded functions* $(h_c)_{c\in\mathbb{R}^+}$ *satisfying* (A2) *and* (A3) *uniformly such that*

$$\lim_{c\to\infty} \sup_{n\in\mathbb{N}} n^2 \mathbb{E}(U_n - U_{n,c})^2 = 0, \tag{2.5}$$

where $U_{n,c} = n^{-2}\sum_{j=1}^n \sum_{k\ne j} h_c(X_j, X_k)$.

After this simplification of the problem, we intend to develop a decomposition of the kernel that allows for the application of a central limit theorem (CLT) for weakly dependent random variables. One could try to imitate the proof of the i.i.d. case. According to the discussion in the previous subsection, this leads to prerequisites that can hardly be checked in numerous cases. Therefore, we do not use a spectral decomposition of the kernel but a wavelet decomposition. It turns out that Lipschitz continuity is the central property the kernel function should satisfy in order to exploit (2.3). For this reason, the choice of Haar wavelets, as they were employed by

Babbel [3], is inappropriate in the present situation. Instead, the application of Lipschitz continuous scale and wavelet functions is more suitable.

In the sequel, let ϕ and ψ denote scale and wavelet functions associated with an one-dimensional multiresolution analysis. As illustrated by Daubechies [6], Section 8, these functions can be selected in such a manner that they possess the following properties:

(1) ϕ and ψ are Lipschitz continuous,
(2) ϕ and ψ have compact support,
(3) $\int_{-\infty}^{\infty} \phi(x)\,\mathrm{d}x = 1$ and $\int_{-\infty}^{\infty} \psi(x)\,\mathrm{d}x = 0$.

It is well known that an orthonormal basis in $L_2(\mathbb{R}^d)$ can be constructed from ϕ and ψ. For this purpose, define $E := \{0, 1\}^d \setminus \{0_d\}$, where 0_d denotes the d-dimensional null vector. In addition, set

$$\varphi^{(i)} := \begin{cases} \phi & \text{for } i = 0, \\ \psi & \text{for } i = 1 \end{cases}$$

and define functions $\Psi_{j,k}^{(e)} : \mathbb{R}^d \to \mathbb{R}$, $j \in \mathbb{Z}$, $k = (k_1, \ldots, k_d)' \in \mathbb{Z}^d$, by

$$\Psi_{j,k}^{(e)}(x) := 2^{jd/2} \prod_{i=1}^{d} \varphi^{(e_i)}(2^j x_i - k_i) \qquad \forall e = (e_1, \ldots, e_d)' \in E, x = (x_1, \ldots, x_d)' \in \mathbb{R}^d.$$

The system $(\Psi_{j,k}^{(e)})_{e \in E, j \in \mathbb{Z}, k \in \mathbb{Z}^d}$ is an orthonormal basis of $L_2(\mathbb{R}^d)$, see Wojtaszczyk [29], Section 5. The same holds true for $(\Phi_{0,k})_{k \in \mathbb{Z}^d} \cup (\Psi_{j,k}^{(e)})_{j \geq 0, e \in E, k \in \mathbb{Z}^d}$, where the functions $\Phi_{j,k} : \mathbb{R}^d \to \mathbb{R}$ are given by $\Phi_{j,k}(x) := 2^{jd/2} \prod_{i=1}^{d} \phi(2^j x_i - k_i)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}^d$.

Now, an L_2-approximation of $nU_{n,c}$ by a statistic based on a wavelet approximation of h_c can be established. To this end, we introduce $\widetilde{h}_c^{(K,L)}$ with

$$\begin{aligned} \widetilde{h}_c^{(K,L)}(x, y) := & \sum_{k_1, k_2 \in \{-L, \ldots, L\}^d} \alpha_{k_1,k_2}^{(c)} \Phi_{0,k_1}(x) \Phi_{0,k_2}(y) \\ & + \sum_{j=0}^{J(K)-1} \sum_{k_1, k_2 \in \{-L, \ldots, L\}^d} \sum_{e \in \bar{E}} \beta_{j;k_1,k_2}^{(c,e)} \Psi_{j;k_1,k_2}^{(e)}(x, y), \end{aligned} \tag{2.6}$$

where $\bar{E} := (E \times E) \cup (E \times \{0_d\}) \cup (\{0_d\} \times E)$,

$$\Psi_{j;k_1,k_2}^{(e)} := \begin{cases} \Psi_{j,k_1}^{(e_1)} \Psi_{j,k_2}^{(e_2)} & \text{for } (e_1', e_2')' \in E \times E, \\ \Psi_{j,k_1}^{(e_1)} \Phi_{j,k_2} & \text{for } (e_1', e_2')' \in E \times \{0_d\}, \\ \Phi_{j,k_1} \Psi_{j,k_2}^{(e_2)} & \text{for } (e_1', e_2')' \in \{0_d\} \times E, \end{cases}$$

$\alpha_{k_1,k_2}^{(c)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y)\Phi_{0,k_1}(x)\Phi_{0,k_2}(y)\,\mathrm{d}x\,\mathrm{d}y$ and $\beta_{j;k_1,k_2}^{(c,e)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y) \times \Psi_{j;k_1,k_2}^{(e)}(x,y)\,\mathrm{d}x\,\mathrm{d}y$. We refer to the degenerate version of $\widetilde{h}_c^{(K,L)}$ as $h_c^{(K,L)}$, given by

$$h_c^{(K,L)}(x,y) := \widetilde{h}_c^{(K,L)}(x,y) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(K,L)}(x,y) P_X(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(K,L)}(x,y) P_X(\mathrm{d}y) + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}_c^{(K,L)}(x,y) P_X(\mathrm{d}x) P_X(\mathrm{d}y).$$

The associated U-type statistic will be denoted by $U_{n,c}^{(K,L)}$.

Lemma 2.2. *Assume that* (A1), (A2), *and* (A3) *are fulfilled. Then the sequence of indices* $(J(K))_{K\in\mathbb{N}}$ *in* (2.6) *with* $J(K) \longrightarrow_{K\to\infty} \infty$ *can be chosen such that*

$$\lim_{K\to\infty} \limsup_{L\to\infty} \sup_{n\in\mathbb{N}} n^2 \mathbb{E}\big(U_{n,c} - U_{n,c}^{(K,L)}\big)^2 = 0.$$

Employing the CLT of Neumann and Paparoditis [26] and the continuous mapping theorem, we obtain the limit distribution of $nU_{n,c}^{(K,L)}$. Finally, based on this result, the asymptotics of the U-type statistic nU_n can be derived. Moreover, a weak law of large numbers (Lemma 5.1 in Section 5.2) allows for deducing the limit distribution of nV_n since $nV_n = nU_n + n^{-1}\sum_{k=1}^n h(X_k, X_k)$.

Before stating the main result of this section, we introduce constants $A_{k_1,k_2} := \operatorname{cov}(\Phi_{0,k_1}(X_1), \Phi_{0,k_2}(X_1))$ and

$$B_{j;k_1,k_2}^{(c,e)} := \begin{cases} \operatorname{cov}\big(\Psi_{j,k_1}^{(e_1)}(X_1), \Psi_{j,k_2}^{(e_2)}(X_1)\big) & \text{for } (e_1', e_2')' \in E\times E, \\ \operatorname{cov}\big(\Psi_{j,k_1}^{(e_1)}(X_1), \Phi_{j,k_2}(X_1)\big) & \text{for } (e_1', e_2')' \in E\times \{0_d\}, \\ \operatorname{cov}\big(\Phi_{j,k_1}(X_1), \Psi_{j,k_2}^{(e_2)}(X_1)\big) & \text{for } (e_1', e_2')' \in \{0_d\}\times E, \end{cases} \qquad j\in\mathbb{Z}, k_1,k_2\in\mathbb{Z}^d.$$

Theorem 2.1. *Suppose that the assumptions* (A1), (A2), *and* (A3) *are fulfilled. Then, as* $n\to\infty$,

$$nU_n \xrightarrow{d} Z$$

with

$$Z := \lim_{c\to\infty}\Bigg(\sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{k_1,k_2}^{(c)}[Z_{k_1}Z_{k_2} - A_{k_1,k_2}] + \sum_{j=0}^{\infty}\sum_{k_1,k_2\in\mathbb{Z}^d}\sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\big[Z_{j;k_1}^{(e_1)} Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(c,e)}\big]\Bigg).$$

Here, $(Z_k)_{k\in\mathbb{Z}^d}$ *as well as* $(Z_{j;k}^{(e)})_{j\geq 0, k\in\mathbb{Z}^d, e\in\{0,1\}^d}$ *are centered and jointly normally distributed random variables and the r.h.s. converges in the* L_2*-sense. If additionally* $\mathbb{E}|h(X_1,X_1)| < \infty$, *then*

$$nV_n \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1).$$

As in the case of i.i.d. random variables, the limit distributions of nU_n and nV_n are, up to a constant, weighted sums of products of centered normal random variables. In contrast to many other results in the literature, the prerequisites of this theorem, namely moment constraints and Lipschitz continuity of the kernel, can be checked fairly easily in many cases. Nevertheless, the asymptotic distribution has a complicated structure. Hence, quantiles can hardly be determined on the basis of the previous result. However, we show in the following section that the conditional distributions of the bootstrap counterparts of nU_n and nV_n, given $X_1, \ldots, X_n$, converge to the same limits in probability.

Of course, the assumption of Lipschitz continuous kernels is rather restrictive. Thus, we extend our theory to a more general class of kernel functions. The costs for enlarging the class of feasible kernels are additional moment constraints.

Besides (A1) and (A2), we assume

(A4) (i) The kernel function satisfies

$$|h(x, y) - h(\bar{x}, \bar{y})| \leq f(x, \bar{x}, y, \bar{y})[\|x - \bar{x}\|_{l_1} + \|y - \bar{y}\|_{l_1}] \qquad \forall x, \bar{x}, y, \bar{y} \in \mathbb{R}^d,$$

where $f : \mathbb{R}^{4d} \to \mathbb{R}$ is continuous. Moreover,

$$\sup_{Y_1, \ldots, Y_5 \sim P_X} \mathbb{E}\Big(\max_{a_1, a_2 \in [-A, A]^d} [f(Y_1, Y_2 + a_1, Y_3, Y_4 + a_2)]^\eta \|Y_5\|_{l_1} \Big) < \infty$$

for $\eta := 1/(1 - \delta)$ with δ satisfying (A2) and some $A > 0$.

(ii) $\sum_{r=1}^\infty r(\tau_r)^{\delta^2} < \infty$.

Even though the assumption (A4)(i) has a rather technical structure, it is satisfied for example, by polynomial kernel functions as long as the sample variables have sufficiently many finite moments. Analogous to Lemma 2.1 and Lemma 2.2, the following assertion holds.

Lemma 2.3. *Suppose that* (A1), (A2), *and* (A4) *are fulfilled. Then a family of bounded kernels* $(h_c)_c$ *satisfying* (A2) *and* (A4) *uniformly and the sequence of indices* $(J(K))_{K \in \mathbb{N}}$ *in* (2.6) *with* $J(K) \longrightarrow_{K \to \infty} \infty$ *can be chosen such that*

$$\lim_{c \to \infty} \limsup_{K \to \infty} \limsup_{L \to \infty} \sup_{n \in \mathbb{N}} \mathbb{E}\big(U_n - U_{n,c}^{(K,L)}\big)^2 = 0.$$

This auxiliary result implies the analogue of Theorem 2.1 for non-Lipschitz kernels.

Theorem 2.2. *Assume that* (A1), (A2), *and* (A4) *are satisfied. Then, as* $n \to \infty$,

$$nU_n \xrightarrow{d} Z,$$

where Z is defined as in Theorem 2.1. If additionally $\mathbb{E}|h(X_1, X_1)| < \infty$, then

$$nV_n \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1).$$

3. Consistency of general bootstrap methods

As we have seen in the previous section, the limit distributions of degenerate U- and V-statistics have a rather complicated structure. Therefore, in the majority of cases it is quite difficult to determine quantiles, which are required in order to derive asymptotic critical values of U- and V-type test statistics. The bootstrap offers a suitable way of approximating these quantities.

Given $X_1, \ldots, X_n$, let X^* and Y^* denote vectors of bootstrap random variables with values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$. In order to describe the dependence structure of the bootstrap sample, we introduce, in analogy to Definition 2.1,

$$\tau^*(Y^*, X^*, x_n) := \mathbb{E}\left(\sup_{f \in \Lambda_1(\mathbb{R}^{d_1})} \left| \int_{\mathbb{R}^{d_1}} f(x) P_{X^*|Y^*}(\mathrm{d}x) - \int_{\mathbb{R}^{d_1}} f(x) P_{X^*}(\mathrm{d}x) \right| \middle| \mathbb{X}_n = x_n \right)$$

provided that $\mathbb{E}(\|X^*\|_{l_1} | \mathbb{X}_n = x_n) < \infty$ with $\mathbb{X}_n := (X_1', \ldots, X_n')'$. We make the following assumptions:

(A1*) (i) The sequence of bootstrap variables is stationary with probability tending to one. Additionally, $(X_{t_1}^{*\prime}, X_{t_2}^{*\prime})' \xrightarrow{d} (X_{t_1}', X_{t_2}')'$, $\forall t_1, t_2 \in \mathbb{N}$, holds true in probability.

(ii) Conditionally on $X_1, \ldots, X_n$, the random variables $(X_k^*)_{k \in \mathbb{Z}}$ are τ-weakly dependent, that is, there exist a sequence of coefficients $(\bar{\tau}_r)_{r \in \mathbb{N}}$ with $\sum_{r=1}^{\infty} r(\bar{\tau}_r)^{\delta} < \infty$ for some $\delta \in (0, 1)$, a constant $C_1 < \infty$, and a sequence of sets $(\mathfrak{X}_n^{(1)})_{n \in \mathbb{N}}$ with $P(\mathbb{X}_n \in \mathfrak{X}_n^{(1)}) \longrightarrow_{n \to \infty} 1$ and the following property: For any sequence $(x_n)_{n \in \mathbb{N}}$ with $x_n \in \mathfrak{X}_n^{(1)}$, $n \in \mathbb{N}$, $\sup_{k \in \mathbb{N}} \mathbb{E}(\|X_k^*\|_{l_1} | \mathbb{X}_n = x_n) \leq C_1$ and

$$\begin{aligned} \tau_r^*(x_n) := \sup\{ &\tau^*((X_{s_1}^{*\prime}, \ldots, X_{s_u}^{*\prime})', (X_{t_1}^{*\prime}, X_{t_2}^{*\prime}, X_{t_3}^{*\prime})', x_n) | \\ &u \in \mathbb{N}, s_1 \leq \cdots \leq s_u < s_u + r \leq t_1 \leq t_2 \leq t_3 \in \mathbb{N}\} \end{aligned}$$

can be bounded by $\bar{\tau}_r$ for all $r \in \mathbb{N}$.

Remark 2.

(i) Neumann and Paparoditis [26] proved that in case of stationary Markov chains of finite order, the key for convergence of the finite-dimensional distributions is convergence of the conditional distributions, cf. their Lemma 4.2. In particular, they showed that AR(p) bootstrap and ARCH(p) bootstrap yield samples that satisfy (A1*)(i).

(ii) In Section 4.2, we present another example that satisfies (A1*), namely a residual-based bootstrap procedure for a Lipschitz contracting nonlinear AR(1) process, given by $X_t = g(X_{t-1}) + \varepsilon_t$. In particular, note that the bootstrap process there cannot be proved to be mixing according to the discreteness of the bootstrap innovations that are generated via

Efron's bootstrap from the empirical distribution of the recentered residuals of the original process.

Lemma 3.1. *Suppose that* (A1) *and* (A1*) *hold true. Further let* $h:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ *be a bounded, symmetric, Lipschitz continuous function such that* $\mathbb{E}h(X_1,y)=\mathbb{E}(h(X_1^*,y)|X_1,\dots,X_n)=0,\forall y\in\mathbb{R}^d$. *Then,*

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}h(X_j^*,X_k^*)\xrightarrow{d}Z\quad\text{and}\quad\frac{1}{n}\sum_{j,k=1}^{n}h(X_j^*,X_k^*)\xrightarrow{d}Z+\mathbb{E}h(X_1,X_1)$$

hold in probability as $n\to\infty$. *Here,* Z *is defined as in Theorem* 2.1.

In order to deduce bootstrap consistency, additionally, convergence in a certain metric ρ is required, that is,

$$\rho\left(P\left(\frac{1}{n}\sum_{j,k=1}^{n}h(X_j^*,X_k^*)\leq x|X_1,\dots,X_n\right),P\left(\frac{1}{n}\sum_{j,k=1}^{n}h(X_j,X_k)\leq x\right)\right)\xrightarrow{P}0.$$

(Here, $\xrightarrow{P}$ denotes convergence in probability.) Convergence in the uniform metric follows from Lemma 3.1 if the limit distribution has a continuous cumulative distribution function. The next assertion gives a necessary and sufficient condition for this.

Lemma 3.2. *The limit variable* Z, *derived in Theorem* 2.1*/Theorem* 2.2 *under* (A1), (A2), *and* (A3)/(A4), *has a continuous cumulative distribution function if* $\operatorname{var}(Z)>0$.

Kernels of statistics emerging from goodness-of-fit tests for composite hypotheses often depend on an unknown parameter. We establish bootstrap consistency for this setting, that is, when parameters have to be estimated. Moreover, the class of feasible kernels is enlarged. For this purpose, we additionally assume

(A2*) (i) $\widehat{\theta}_n\xrightarrow{P}\theta\in\Theta\subseteq\mathbb{R}^p$.
(ii) $\mathbb{E}(h(X_1^*,y,\widehat{\theta}_n)|\mathbb{X}_n)=0,\forall y\in\mathbb{R}^d$.
(iii) For some δ satisfying (A1*)(ii), $\nu>(2-\delta)/(1-\delta)$, and a constant $C_2<\infty$, there exists a sequence of sets $(\mathfrak{X}_n^{(2)})_{n\in\mathbb{N}}$ such that $P(\mathbb{X}_n\in\mathfrak{X}_n^{(2)})\longrightarrow_{n\to\infty}1$ and $\forall(x_n)_{n\in\mathbb{N}}$ with $x_n\in\mathfrak{X}_n^{(2)}$ the following moment constraint holds true:

$$\sup_{1\leq k<n}\mathbb{E}\big(|h(X_1^*,X_{1+k}^*,\widehat{\theta}_n)|^\nu+|h(X_1^*,\widetilde{X}_1^*,\widehat{\theta}_n)|^\nu|\mathbb{X}_n=x_n\big)\leq C_2,$$

where (conditionally on $\mathbb{X}_n$) $\widetilde{X}_1^*$ denotes an independent copy of X_1^*.

(A3*) (i) The kernel is continuous in its third argument in some neighbourhood $U(\theta)\subseteq\Theta$ of θ and satisfies

$$|h(x,y,\widehat{\theta}_n)-h(\bar{x},\bar{y},\widehat{\theta}_n)|\leq f(x,\bar{x},y,\bar{y},\widehat{\theta}_n)[\|x-\bar{x}\|_{l_1}+\|y-\bar{y}\|_{l_1}]$$

for all $x, \bar{x}, y, \bar{y} \in \mathbb{R}^d$, where $f: \mathbb{R}^{4d} \times \mathbb{R}^p \to \mathbb{R}$ is continuous on $\mathbb{R}^{4d} \times U(\theta)$. Moreover, for $\eta := 1/(1-\delta)$ and some constants $A > 0, C_3 < \infty$ there exists a sequence of sets $(\mathfrak{X}_n^{(3)})_{n\in\mathbb{N}}$ such that $P(\mathbb{X}_n \in \mathfrak{X}_n^{(3)}) \longrightarrow_{n\to\infty} 1$ and $\forall (x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathfrak{X}_n^{(3)}$ the following moment constraint holds true:

$$\mathbb{E}\Big(\max_{a_1,a_2\in[-A,A]^d} [f(Y_1^*, Y_2^* + a_1, Y_3^*, Y_4^* + a_2, \widehat{\theta}_n)]^\eta \|Y_5^*\|_{l_1} \big| \mathbb{X}_n = x_n\Big) \le C_3$$

for all $Y_1^*, \dots, Y_5^*$ with $Y_k^* \stackrel{d}{=} X_1^*, k \in \{1, \dots, 5\}$ (conditionally on $X_1, \dots, X_n$).

(ii) $\sum_{r=1}^\infty r(\bar{\tau}_r)^{\delta^2} < \infty$.

Under these assumptions a result concerning the asymptotic distributions of $nU_n^* = n^{-1} \times \sum_{j=1}^n \sum_{k\neq j} h(X_j^*, X_k^*, \widehat{\theta}_n)$ and $nV_n^* = n^{-1} \sum_{j,k=1}^n h(X_j^*, X_k^*, \widehat{\theta}_n)$ can be derived. To this end, we denote the U- and V-statistics with kernel $h(\cdot, \cdot, \theta)$ and arguments $X_1, \dots, X_n$ by U_n and V_n, respectively.

Theorem 3.1. *Suppose that the conditions* (A1), (A2), *and* (A4) *as well as* (A1*), (A2*), *and* (A3*) *are fulfilled.*

(i) *As $n \to \infty$,*

$$nU_n^* \xrightarrow{d} Z, \qquad \textit{in probability,}$$

where Z is defined as in Theorem 2.1. *If furthermore* $\operatorname{var}(Z) > 0$, *then*

$$\sup_{-\infty<x<\infty} |P(nU_n^* \le x | X_1, \dots, X_n) - P(nU_n \le x)| \xrightarrow{P} 0.$$

(ii) *If additionally* $\mathbb{E}|h(X_1, X_1, \theta)| < \infty$ *and* $\mathbb{E}(|h(X_1^*, X_1^*, \widehat{\theta}_n)||\mathbb{X}_n) \xrightarrow{P} \mathbb{E}|h(X_1, X_1, \theta)|$, *then as $n \to \infty$,*

$$nV_n^* \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1, \theta), \qquad \textit{in probability.}$$

Moreover, in case of $\operatorname{var}(Z) > 0$,

$$\sup_{-\infty<x<\infty} |P(nV_n^* \le x | X_1, \dots, X_n) - P(nV_n \le x)| \xrightarrow{P} 0.$$

Remark 3. Theorem 3.1 implies that bootstrap-based tests of U- or V-type have asymptotically a prescribed size α, that is, $P(nU_n > t_{u,\alpha}^*) \longrightarrow_{n\to\infty} \alpha$ and $P(nV_n > t_{v,\alpha}^*) \longrightarrow_{n\to\infty} \alpha$, where $t_{u,\alpha}^*$ and $t_{v,\alpha}^*$ denote the $(1-\alpha)$-quantiles of nU_n^* and nV_n^*, respectively, given $X_1, \dots, X_n$.

4. L_2-tests for weakly dependent observations

This section is dedicated to two applications in the field of hypothesis testing. For sake of simplicity, we restrict ourselves to real-valued random variables and consider simple null hypotheses only. The test for symmetry as well as the model-specification test can be extended to problems with composite hypotheses, cf. Leucht [23,24].

4.1. A test for symmetry

Answering the question whether a distribution is symmetric or not is interesting for several reasons. Often robust estimators of and robust tests for location parameters assume the observations to arise from a symmetric distribution, see, for example, Staudte and Sheather [28]. Consequently, it is important to check this assumption before applying those methods. Moreover, symmetry plays a central role in analyzing and modeling real-life phenomena. For instance, it is often presumed that an observed process can be described by an AR(p) process with Gaussian innovations which in turn implies a Gaussian marginal distribution. Rejecting the hypothesis of symmetry contradicts this type of marginal distribution. Furthermore, this result of the test excludes any kind of symmetric innovations in that context.

Suppose that we observe $X_1, \ldots, X_n$ from a sequence of real-valued random variables with common distribution P_X and satisfying (A1). For some $\mu \in \mathbb{R}$, we are given the problem

$$\mathcal{H}_0\colon\ P_{X-\mu} = P_{\mu-X} \quad \text{vs.} \quad \mathcal{H}_1\colon\ P_{X-\mu} \neq P_{\mu-X}.$$

Similar to Feuerverger and Mureika [18], who studied the problem for i.i.d. random variables, we propose the following test statistic:

$$S_n = n \int_{\mathbb{R}} \bigl[\Im\bigl(c_n(t)\mathrm{e}^{-\mathrm{i}\mu t}\bigr)\bigr]^2 w(t)\,\mathrm{d}t = \frac{1}{n} \sum_{j,k=1}^{n} \int_{\mathbb{R}} \sin\bigl(t(X_j - \mu)\bigr) \sin\bigl(t(X_k - \mu)\bigr) w(t)\,\mathrm{d}t$$

which makes use of the fact that symmetry of a distribution is equivalent to a vanishing imaginary part of the associated characteristic function. Here, $\Im(z)$ denotes the imaginary part of $z \in \mathbb{C}$, c_n denotes the empirical characteristic function and w is some positive measurable weight function with $\int_{\mathbb{R}} (1 + |t|) w(t)\,\mathrm{d}t < \infty$. Obviously, S_n is a V-type statistic whose kernel satisfies (A2) and (A3). Thus, its limit distribution can be determined by Theorem 2.1. Assuming that the observations come from a stationary AR(p) or ARCH(p) process, the validity of (A1*) is assured when the AR(p) or ARCH(p) bootstrap methods given by Neumann and Paparoditis [26] are used in order to generate the bootstrap counterpart of the sample. Hence, in these cases the prerequisites of Lemma 3.1 are satisfied excluding degeneracy. Inspired by Dehling and Mikosch [10], who discussed this problem for Efron's Bootstrap in the i.i.d. case, we propose a bootstrap statistic with the kernel

$$h_n^*(x, y) = h(x, y) - \int_{\mathbb{R}} h(x, y) P_n^*(\mathrm{d}x) - \int_{\mathbb{R}} h(x, y) P_n^*(\mathrm{d}y) + \int_{\mathbb{R}^2} h(x, y) P_n^*(\mathrm{d}x) P_n^*(\mathrm{d}y).$$

Here, h denotes the kernel function of S_n and P_n^* the distribution of X_1^* conditionally on $X_1, \ldots, X_n$. Similar to the proof of Theorem 3.1, the desired convergence property of S_n^* can be verified.

4.2. A model-specification test

Let $X_0, \ldots, X_n$ be observations resulting from a stationary real-valued nonlinear autoregressive process with centered i.i.d. innovations $(\varepsilon_k)_{k\in\mathbb{Z}}$, that is, $X_k = g(X_{k-1}) + \varepsilon_k$. Suppose that $\mathbb{E}|\varepsilon_0|^{4+\delta} < \infty$ for some $\delta > 0$ and that $g \in G := \{f : \mathbb{R} \to \mathbb{R} | f$ Lipschitz continuous with $\mathrm{Lip}(f) < 1\}$. Thus, the process $(X_k)_{k\in\mathbb{Z}}$ is τ-dependent with exponential rate, see Dedecker and Prieur [9], Example 4.2. We will present a test for the problem

$$\mathcal{H}_0\colon\ P\big(\mathbb{E}(X_1|X_0) = g_0(X_0)\big) = 1 \quad \text{vs.} \quad \mathcal{H}_1\colon\ P\big(\mathbb{E}(X_1|X_0) = g_0(X_0)\big) < 1$$

with $g_0 \in G$. For sake of simplicity, we stick to these small classes of functions G and of processes $(X_k)_{k\in\mathbb{Z}}$. An extension to a more comprehensive variety of model-specification tests is investigated in a forthcoming paper, cf. Leucht [24].

Similar to Fan and Li [16], we propose the following test statistic:

$$\begin{aligned} T_n &= \frac{1}{n\sqrt{h}} \sum_{j=1}^{n} \sum_{k\neq j} \big(X_j - g_0(X_{j-1})\big)\big(X_k - g_0(X_{k-1})\big) K\left(\frac{X_{j-1} - X_{k-1}}{h}\right) \\ &=: \frac{1}{n} \sum_{j=1}^{n} \sum_{k\neq j} H(Z_j, Z_k), \end{aligned}$$

that is, a kernel estimator (multiplied with $n\sqrt{h}$) of $\mathbb{E}([X_1 - g(X_0)]\mathbb{E}(X_1 - g(X_0)|X_0)p(X_0))$ that is equal to zero under $\mathcal{H}_0$. Here, $Z_k := (X_k, X_{k-1})', k \in \mathbb{Z}$, and p denotes the density of the distribution of X_0. Fan and Li [16], who considered β-mixing processes, used a similar test statistic with a vanishing bandwidth. In contrast, we consider the case of a fixed bandwidth. These tests are more powerful against Pitman alternatives $g_{1,n}(x) = g_0(x) + n^{-\beta} w(x) + \mathrm{o}(n^{-\beta}), \beta > 0, w \in G$. For a detailed discussion of this topic, see Fan and Li [17].

Obviously, T_n is degenerate under $\mathcal{H}_0$. If we assume K to be a bounded, even, and Lipschitz continuous function, then there exists a function $f : \mathbb{R}^8 \to \mathbb{R}$ with $|H(z_1, z_2) - H(\bar{z}_1, \bar{z}_2)| \le f(z_1, \bar{z}_1, z_2, \bar{z}_2)(\|z_1 - \bar{z}_1\|_{l_1} + \|z_2 - \bar{z}_2\|_{l_1})$ and such that (A4) is valid. Moreover, under these conditions H satisfies (A2). Hence, the assertion of Theorem 2.2 holds true. In order to determine critical values of the test, we propose the bootstrap procedure given by Franke and Wendel [19] (without estimating the regression function). The bootstrap innovations $(\varepsilon_t^*)_t$ are drawn with replacement from the set $\{\tilde{\varepsilon}_t = \varepsilon_t - n^{-1}\sum_{k=1}^n \varepsilon_k\}_{t=1}^n$, where $\varepsilon_t = X_t - g_0(X_{t-1}), t = 1, \ldots, n$. After choosing a starting value X_0^* independently of $(\varepsilon_t^*)_{t\geq 1}$, the bootstrap sample $X_t^* = g(X_{t-1}^*) + \varepsilon_t^*$ as well as the bootstrap counterpart $T_n^* = n^{-1}\sum_{j=1}^n \sum_{k\neq j} H(Z_j^*, Z_k^*)$ of the test statistic with $Z_k^* = (X_k^*, X_{k-1}^*)', k = 1, \ldots, n$, can be computed. In contrast to the previous subsection, the proposed bootstrap method leads to a degenerate kernel function. Obviously, the bootstrap sample is τ-dependent in the sense of (A1*) and satisfies $\mathbb{E}(|X_k^*||Z_1, \ldots, Z_n) < C$ for some $C < \infty$ with probability tending to one. Theorem 1 of Diaconis and Freedman [13] yields the existence of a stationary solution to $X_t^* = g(X_{t-1}^*) + \varepsilon_t^*$ and that the distribution of any "reasonably" started process converges to the stationary one with exponential rate. In order to apply our theory, X_0^* is assumed to be drawn from the stationary bootstrap distribution, conditionally on $X_1, \dots, X_n$. We employ Lemma 4.2 of Neumann and Paparoditis [26] to verify convergence of the finite dimensional distributions. The application of this result requires the convergence of the conditional distributions, that is, $\sup_{x \in K} d(P^{X_t^*|X_{t-1}^*=x}, P^{X_t|X_{t-1}=x}) \xrightarrow{P} 0$ for every compact $K \subset \mathbb{R}$ and $d(P, Q) = \inf_{X\sim P, Y\sim Q} \mathbb{E}(|X - Y| \wedge 1)$. In the present context, this can be confirmed similarly to the proof of Lemma 4.1 by Neumann and Paparoditis [26] if the innovations of the original process have a bounded density. Summing up, all prerequisites of Theorem 3.1 are satisfied. Hence, critical values of the above test can be determined using the proposed model-based bootstrap procedure.

5. Proofs

5.1. Proofs of the main theorems

Throughout this section, C denotes a positive finite generic constant.

Proof of Theorem 2.1. First, we derive the limit distribution of $nU_{n,c}^{(K,L)}$, defined before Lemma 2.2. Afterwards, the asymptotic distributions of nU_n and nV_n are deduced by means of Lemma 2.1, Lemma 2.2, and a weak law of large numbers.

The following modified representation of $\widetilde{h}_c^{(K,L)}$ will be useful in the sequel:

$$\widetilde{h}_c^{(K,L)}(x, y) = \sum_{k,l=1}^{M(K,L)} \gamma_{k,l}^{(c)} \tilde{q}_k(x)\tilde{q}_l(y),$$

where $(\tilde{q}_l)_{l=1}^{M(K,L)}$ is an ordering of $\bigcup_{k\in\{-L,\dots,L\}^d} \{\{\Phi_{j,k}\} \cup \{\Psi_{j,k}^{(e)}\}_{e\in E, j\in\{0,\dots,J(K)-1\}}\}$ and $\gamma_{k,l}^{(c)} = \gamma_{l,k}^{(c)}, k, l \in \{1, \dots, M(K, L)\}$, are the associated coefficients. Moreover, the introduction of $q_k(X_i) := \tilde{q}_k(X_i) - \mathbb{E}\tilde{q}_k(X_i), k \in \{1, \dots, M(K, L)\}, i \in \{1, \dots, n\}$, allows for the compact notation of $nU_{n,c}^{(K,L)}$,

$$nU_{n,c}^{(K,L)} = \sum_{k,l=1}^{M(K,L)} \gamma_{k,l}^{(c)} \left(\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q_k(X_i) \right] \left[\frac{1}{\sqrt{n}} \sum_{j=1}^{n} q_l(X_j) \right] - \frac{1}{n} \sum_{i=1}^{n} q_k(X_i) q_l(X_i) \right).$$

The latter summand in the round brackets converges to $-\mathbb{E}q_k(X_1)q_l(X_1)$ in probability by virtue of Lemma 5.1. In order to derive the limit distributions of the first summands, we consider $n^{-1/2}\sum_{i=1}^{n}(q_1(X_i), \dots, q_{M(K,L)}(X_i))'$. Due to the Cramér–Wold device, it suffices to investigate $\sum_{k=1}^{M(K,L)} t_k n^{-1/2} \sum_{i=1}^{n} q_k(X_i)$, $\forall (t_1, \dots, t_{M(K,L)})' \in \mathbb{R}^{M(K,L)}$. Asymptotic normality can be established by applying the CLT of Neumann and Paparoditis [26] to $Q_i := \sum_{k=1}^{M(K,L)} t_k q_k(X_i), i = 1, \dots, n$. To this end, the prerequisites of this tool have to be checked. Obviously, we are given a strictly stationary sequence of centered bounded random variables. This implies in conjunction with the dominated convergence theorem that the Lindeberg condition is fulfilled. In order to show

$$\frac{1}{n}\operatorname{var}(Q_1+\cdots+Q_n) \underset{n\to\infty}{\longrightarrow} \sigma^2 := \operatorname{var}(Q_1) + 2\sum_{k=2}^{\infty}\operatorname{cov}(Q_1, Q_k),$$

the validity of (A1) can be employed which moreover assures the existence of the limit σ^2. Then,

$$\begin{aligned}\left|\frac{1}{n}\operatorname{var}(Q_1+\cdots+Q_n) - \sigma^2\right| &= \left|\frac{2}{n}\sum_{r=2}^{n}(n-[r-1])\operatorname{cov}(Q_1, Q_r) - 2\sum_{k=2}^{\infty}\operatorname{cov}(Q_1, Q_k)\right| \\ &\le 2\sum_{r=2}^{\infty}\min\left\{\frac{r-1}{n}, 1\right\}|\operatorname{cov}(Q_1, Q_r)| \\ &\le 4\|Q_1\|_\infty \operatorname{Lip}(Q_1)\sum_{r=2}^{\infty}\min\left\{\frac{r-1}{n}, 1\right\}\tau_{r-1},\end{aligned}$$

where the latter inequality follows from (2.4). The summability condition of the dependence coefficients in connection with Lebesgue's dominated convergence theorem yields the desired result. Since $Q_{t_1}Q_{t_2}$ forms a Lipschitz continuous function, inequality (6.4) of Neumann and Paparoditis [26] holds true with $\theta_r = \operatorname{Lip}(Q_{t_1}Q_{t_2})\tau_r$. It is easy to convince oneself that their condition (6.3) is not needed if the involved random variables are uniformly bounded. Finally, we obtain

$$n^{-1/2}(Q_1+\cdots+Q_n) \stackrel{d}{\longrightarrow} N(0,\sigma^2)$$

and hence,

$$\begin{aligned} nU_{n,c}^{(K,L)} &\stackrel{d}{\longrightarrow} Z_c^{(K,L)} \\ &:= \sum_{k_1,k_2\in\{-L,\dots,L\}^d} \alpha_{k_1,k_2}^{(c)}[Z_{k_1}Z_{k_2} - A_{k_1,k_2}] \\ &\quad + \sum_{j=0}^{J(K)-1}\sum_{k_1,k_2\in\{-L,\dots,L\}^d}\sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\big[Z_{j;k_1}^{(e_1)}Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)}\big].\end{aligned}$$

Here, $(Z_k)_{k\in\{-L,\dots,L\}^d}$ and $(Z_{j;k}^{(e)})_{j\in\{0,\dots,J(K)-1\},e\in\{0,1\}^d,k\in\{-L,\dots,L\}^d}$, respectively, are centered and jointly normally distributed random variables.

By Lemma 2.1 and Lemma 2.2, we have

$$\lim_{c\to\infty}\limsup_{K\to\infty}\limsup_{L\to\infty}\sup_{n\in\mathbb{N}} n^2\mathbb{E}\big(U_{n,c}^{(K,L)} - U_n\big)^2 = 0.$$

Since $nU_{n,c}^{(K,L)} \stackrel{d}{\longrightarrow} Z_c^{(K,L)}$, it remains to show

$$\lim_{c\to\infty}\limsup_{K\to\infty}\limsup_{L\to\infty}\mathbb{E}\big(Z_c^{(K,L)} - Z\big)^2 = 0 \tag{5.1}$$

in order to prove that $nU_n \xrightarrow{d} Z$ due to Billingsley [4], Theorem 4.2. To this end, we first show that $(Z_c^{(K,L)})_L$ is a Cauchy sequence in L_2. Note that $n(U_{n,c}^{(K,L_1)} - U_{n,c}^{(K,L_2)}) \xrightarrow{d} Z_c^{(K,L_1)} - Z_c^{(K,L_2)}$. According to Theorem 5.3 of Billingsley [4], we obtain $\mathbb{E}(Z_c^{(K,L_1)} - Z_c^{(K,L_2)})^2 \le \liminf_{n\to\infty} n^2 \mathbb{E}(U_{n,c}^{(K,L_1)} - U_{n,c}^{(K,L_2)})^2$. The r.h.s. converges to zero as $L_1, L_2 \to \infty$ by virtue of (5.10) in the proof of Lemma 2.2. Denoting the corresponding limit by $Z_c^{(K)}$ similar arguments yield

$$\begin{aligned}\mathbb{E}\big(Z_c^{(K_1)} - Z_c^{(K_2)}\big)^2 &\le 4 \limsup_{L\to\infty} \mathbb{E}\big(Z_c^{(K_1,L)} - Z_c^{(K_2,L)}\big)^2 \\ &\le 4 \limsup_{L\to\infty} \liminf_{n\to\infty} n^2 \mathbb{E}\big(U_{n,c}^{(K_1,L)} - U_{n,c}^{(K_2,L)}\big)^2 \\ &\le 16 \liminf_{n\to\infty} n^2 \mathbb{E}\big(U_{n,c}^{(K_1)} - U_{n,c}^{(K_2)}\big)^2 \underset{K_1,K_2\to\infty}{\longrightarrow} 0\end{aligned}$$

according to (5.9) of the proof of Lemma 2.2. In view of Lemma 2.1, we obtain (5.1) by applying the above method once again. This in turn leads to the desired limit distribution of nU_n.

Based on the result concerning U-type statistics, the limit distribution of nV_n can be established. Since $V_n = U_n + n^{-2}\sum_{k=1}^n h(X_k, X_k)$, it remains to verify that $n^{-1}\sum_{k=1}^n h(X_k, X_k) \xrightarrow{P} \mathbb{E}h(X_1, X_1)$. This in turn is a consequence of Lemma 5.1. □

Proof of Theorem 2.2. On the basis of Lemma 2.3 similar arguments as in the proof of Theorem 2.1 yield $nU_n \xrightarrow{d} Z$. Moreover, Lemma 5.1 implies $n^{-1}\sum_{k=1}^n h(X_k, X_k) \xrightarrow{P} \mathbb{E}h(X_1, X_1)$. Thus, $nV_n \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1)$. □

Proof of Theorem 3.1. Due to Lemma 3.2, it suffices to verify distributional convergence. To this end, we introduce

$$\mathfrak{X}_n^\theta \subseteq \mathfrak{X}_n^{(1)} \cap \mathfrak{X}_n^{(2)} \cap \mathfrak{X}_n^{(3)} \cap \{\mathbb{X}_n \,|\, \|\widehat{\theta}_n - \theta\|_{l_1} < \delta_n\}$$

such that

$$\mathcal{L}\big((X_{t_1}^{*\prime}, \dots, X_{t_k}^{*\prime})' | \mathbb{X}_n = x_n\big) \;=\; \mathcal{L}\big((X_{t_1+l}^{*\prime}, \dots, X_{t_k+l}^{*\prime})' | \mathbb{X}_n = x_n\big), \tag{5.2}$$

$$\mathcal{L}\big((X_{t_1}^{*\prime}, X_{t_2}^{*\prime})' | \mathbb{X}_n = x_n\big) \Longrightarrow \mathcal{L}((X_{t_1}', X_{t_2}')') \tag{5.3}$$

uniformly for any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathfrak{X}_n^\theta$ and $t_1, \dots, t_k, k, l \in \mathbb{N}$. Moreover, the null sequence $(\delta_n)_{n\in\mathbb{N}}$ can be chosen such that on $\mathfrak{X}_n^\theta$, $\widehat{\theta}_n \in U(\theta)$ and $P(\mathbb{X}_n \in \mathfrak{X}_n^\theta) \longrightarrow_{n\to\infty} 1$ hold. Hence, to prove $nU_n^* \xrightarrow{d} Z$, in probability, it suffices to verify that nU_n^* converges to Z in distribution conditionally on $\mathbb{X}_n = x_n$ for any sequence $(x_n)_n$ with $x_n \in \mathfrak{X}_n^\theta$. Now, we take an arbitrary sequence $(x_n)_n$ with $x_n \in \mathfrak{X}_n^\theta$, $n \in \mathbb{N}$.

In order to show that it suffices to investigate statistics with bounded kernels, we consider the degenerate version h_c^* of

$$\widetilde{h}_c^*(x,y,\widehat{\theta}_n) := \begin{cases} h(x,y,\widehat{\theta}_n) & \text{for } |h(x,y,\widehat{\theta}_n)| \le c_h(\widehat{\theta}_n), \\ -c_h(\widehat{\theta}_n) & \text{for } h(x,y,\widehat{\theta}_n) < -c_h(\widehat{\theta}_n), \\ c_h(\widehat{\theta}_n) & \text{for } h(x,y,\widehat{\theta}_n) > c_h(\widehat{\theta}_n) \end{cases}$$

with $c_h(\widehat{\theta}_n) := \max_{x,y\in[-c,c]^d} |h(x,y,\widehat{\theta}_n)| \le \max_{x,y\in[-c,c]^d, \|\bar{\theta}\|_{l_1}\le\delta_1} |h(x,y,\bar{\theta})| < \infty$. The associated U-statistics are denoted by $U_{n,c}^*$. Now, imitating the proof of Lemma 2.1 results in

$$\limsup_{n\to\infty} n^2 \mathbb{E}[(U_n^* - U_{n,c}^*)^2 | \mathbb{X}_n = x_n] \underset{c\to\infty}{\longrightarrow} 0.$$

Within the calculations, the relation $\limsup_{n\to\infty} P(X_1^* \notin (-c,c)^d | \mathbb{X}_n = x_n) \le P(X_1 \notin (-c,c)^d) \longrightarrow_{c\to\infty} 0$ has to be invoked which follows from Portmanteau's theorem in conjunction with (5.3). Next, we approximate the bounded kernel by the degenerate version of

$$\widetilde{h}_c^{*(K,L)} := \sum_{k_1,k_2\in\{-L,\dots,L\}^d} \widehat{\alpha}_{k_1,k_2}^{(c)} \Phi_{0,k_1}\Phi_{0,k_2} + \sum_{j=0}^{J(K)-1} \sum_{k_1,k_2\in\{-L,\dots,L\}^d} \sum_{e\in\bar{E}} \widehat{\beta}_{j;k_1,k_2}^{(c,e)} \Psi_{j;k_1,k_2}^{(e)},$$

where $\widehat{\alpha}_{k_1,k_2}^{(c)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c^*(x,y,\widehat{\theta}_n)\Phi_{0,k_1}(x)\Phi_{0,k_2}(y)\,\mathrm{d}x\,\mathrm{d}y$ and $\widehat{\beta}_{j;k_1,k_2}^{(c,e)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c^*(x,y,\widehat{\theta}_n)\Psi_{j;k_1,k_2}^{(e)}(x,y)\,\mathrm{d}x\,\mathrm{d}y$. Denoting the associated U-statistic by $\widehat{U}_{n,c}^{*(K,L)}$ leads to

$$\lim_{K\to\infty} \limsup_{L\to\infty} \limsup_{n\to\infty} n^2 \mathbb{E}\big[\big(U_{n,c}^* - \widehat{U}_{n,c}^{*(K,L)}\big)^2 | \mathbb{X}_n = x_n\big] = 0$$

which can be proved by following the lines of the proof of Lemma 2.3. Here, $J(K)$ is chosen as follows: We first select some $b = b(K) < \infty$ such that $P(X_1 \notin (-b,b)^d) \le 1/K$. Afterwards, we choose $J(K)$ such that $\max_{x,y\in[-b,b]^d} |h_c(x,y,\theta) - \widetilde{h}_c^{(K)}(x,y,\theta)| \le 1/K$ and $S_\phi/2^{J(K)} < A$, where S_ϕ denotes the length of the support of the scale function ϕ. The index $J(K)$ can be determined independently of n on $(\mathfrak{X}_n^\theta)_n$ since $\max_{x,y\in[-b,b]^d} |h_c^*(x,y,\widehat{\theta}_n) - h_c(x,y,\theta)| \longrightarrow 0$ and $\max_{x,y\in[-b,b]^d} |\widetilde{h}_c^{(K)}(x,y,\theta) - \widetilde{h}_c^{*(K)}(x,y,\widehat{\theta}_n)| \longrightarrow 0$, as $n\to\infty$, due to the continuity assumptions on f. Here, $\widetilde{h}_c^{*(K)}$ is defined by the substitution of $\sum_{k_1,k_2\in\{-L,\dots,L\}^d}$ through $\sum_{k_1,k_2\in\mathbb{Z}^d}$ in the definition of $\widetilde{h}_c^{*(K,L)}$. Also note that

$$\widehat{\alpha}_{k_1,k_2}^{(c)} \underset{n\to\infty}{\longrightarrow} \alpha_{k_1,k_2}^{(c)} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y,\theta)\Phi_{0,k_1}(x)\Phi_{0,k_2}(y)\,\mathrm{d}x\,\mathrm{d}y,$$

$$\widehat{\beta}_{j;k_1,k_2}^{(c,e)} \underset{n\to\infty}{\longrightarrow} \beta_{j;k_1,k_2}^{(c,e)} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y,\theta)\Psi_{j;k_1,k_2}^{(e)}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

on $(\mathfrak{X}_n^\theta)_n$. Hence, $\lim_{n\to\infty} n^2 \mathbb{E}[(\widehat{U}_{n,c}^{*(K,L)} - U_{n,c}^{*(K,L)})^2 | \mathbb{X}_n = x_n] = 0$, where the kernel of $U_{n,c}^{*(K,L)}$ is obtained by substituting $\widehat{\alpha}_{k_1,k_2}^{(c)}$ and $\widehat{\beta}_{j;k_1,k_2}^{(c,e)}$ in the kernel of $\widehat{U}_{n,c}^{*(K,L)}$ through $\alpha_{k_1,k_2}^{(c)}$ and $\beta_{j;k_1,k_2}^{(c,e)}$, respectively.

Thus, the next step is the application of the CLT of Neumann and Paparoditis [26] to $nU_{n,c}^{*(K,L)}$. For this purpose, we introduce $Q_i^* := \sum_{k=1}^{M(K,L)} t_k q_k^*(X_i^*), t_1, \dots, t_{M(K,L)} \in \mathbb{R}$, where q_k^* denotes the centered version (w.r.t. $P_{X_1^*|\mathbb{X}_n=x_n}$) of $\widetilde{q}_k$ and $(\widetilde{q}_k)_k$ is defined as in the proof of Theorem 2.1. Obviously, given $X_1, \dots, X_n$, the sequence $(Q_i^*)_i$ is centered and has uniformly bounded second moments. Due to (A1*)(i), the Lindeberg condition is satisfied. In order to show that for arbitrary $\varepsilon > 0$ the inequalities $|\frac{1}{n}\operatorname{var}(Q_1^* + \cdots + Q_n^*|\mathbb{X}_n = x_n) - \sigma^2| < \varepsilon, \forall n \ge n_0(\varepsilon)$, hold true with σ^2 as in the proof of Theorem 2.1, the abbreviations $\operatorname{var}^*(\cdot) = \operatorname{var}(\cdot|\mathbb{X}_n = x_n)$ and $\operatorname{cov}^*(\cdot) = \operatorname{cov}(\cdot|\mathbb{X}_n = x_n)$ are used. Hence,

$$
\begin{aligned}
&\left|\frac{1}{n}\operatorname{var}^*[Q_1^* + \cdots + Q_n^*] - \sigma^2\right| \\
&\quad \le 2\sum_{r=2}^{\infty} \min\left\{\frac{r-1}{n}, 1\right\} |\operatorname{cov}^*(Q_1^*, Q_r^*)| + \left|\operatorname{var}^*(Q_1^*) + 2\sum_{r=2}^{\infty}\operatorname{cov}^*(Q_1^*, Q_r^*) - \sigma^2\right| \\
&\quad \le 2\sum_{r=2}^{\infty} \min\left\{\frac{r-1}{n}, 1\right\} |\operatorname{cov}^*(Q_1^*, Q_r^*)| + 2\left|\sum_{r=2}^{R-1}[\operatorname{cov}^*(Q_1^*, Q_r^*) - \operatorname{cov}(Q_1, Q_r)]\right| \\
&\qquad + |\operatorname{var}^*(Q_1^*) - \operatorname{var}(Q_1)| + 2\left|\sum_{r\ge R}\operatorname{cov}^*(Q_1^*, Q_r^*)\right| + 2\left|\sum_{r\ge R}\operatorname{cov}(Q_1, Q_r)\right|.
\end{aligned}
$$

By (A1) and (A1*), R can be chosen such that $|\sum_{r\ge R}\operatorname{cov}(Q_1, Q_r)| + |\sum_{r\ge R}\operatorname{cov}^*(Q_1^*, Q_r^*)| \le \varepsilon/4$. Moreover, (A1*) implies that the first summand can be bounded from above by $\varepsilon/4$ as well if $n \ge n_0(\varepsilon)$ for some $n_0(\varepsilon) \in \mathbb{N}$. According to the convergence of the two-dimensional distributions and the uniform boundedness of $(Q_k^*)_{k\in\mathbb{Z}}$, it is possible to pick $n_0(\varepsilon)$ such that additionally the two remaining summands are bounded by $\varepsilon/8$. For the validity of the CLT of Neumann and Paparoditis [26] in probability, it remains to verify their inequality (6.4). By Lipschitz continuity of $Q_{t_1}^* Q_{t_2}^*$ this holds with $\bar{\theta}_r = \operatorname{Lip}(Q_{t_1}^* Q_{t_2}^*)\bar{\tau}_r \le C\bar{\tau}_r$. The application of the continuous mapping theorem results in $nU_{n,c}^{*(K,L)} \xrightarrow{d} Z_c^{(K,L)}$, in probability. Invoking the same arguments as in the proof of Theorem 2.1, this implies $nU_n^* \xrightarrow{d} Z$, in probability.

In order to obtain the analogous result of convergence for nV_n^*, we define $\widetilde{\mathfrak{X}}_n^\theta \subseteq \mathfrak{X}_n^\theta, n \in \mathbb{N}$, such that $|\mathbb{E}(|h(X_1^*, X_1^*, \widehat{\theta}_n)||\mathbb{X}_n = x_n) - \mathbb{E}|h(X_1, X_1, \theta)|| \le \eta_n, \forall x_n \in \widetilde{\mathfrak{X}}_n^\theta$. Here, the null sequence $(\eta_n)_{n\in\mathbb{N}}$ is chosen in such a way that $P(\mathbb{X}_n \in \widetilde{\mathfrak{X}}_n^\theta) \longrightarrow_{n\to\infty} 1$. Now, additionally to our previous considerations,

$$
P\left(\left|\frac{1}{n}\sum_{i=1}^n h(X_i^*, X_i^*, \widehat{\theta}_n) - \mathbb{E}h(X_1, X_1, \theta)\right| > \varepsilon \middle| \mathbb{X}_n = x_n\right) \underset{n\to\infty}{\longrightarrow} 0
$$

has to be proved for arbitrary $\varepsilon > 0$ and any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \widetilde{\mathfrak{X}}_n^\theta, n \in \mathbb{N}$. According to the definition of the sets $(\widetilde{\mathfrak{X}}_n^\theta)_n$, we get $\mathbb{E}(h(X_1^*, X_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n) \longrightarrow_{n\to\infty} \mathbb{E}h(X_1, X_1, \theta)$.

Therefore, it suffices to prove

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\left[h(X_k^*, X_k^*, \widehat{\theta}_n) - \mathbb{E}\left(h(X_1^*, X_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n\right)\right]\right| > \frac{\varepsilon}{2}\middle|\mathbb{X}_n = x_n\right) \underset{n\to\infty}{\longrightarrow} 0.$$

This in turn is a consequence of Lemma 5.1 since under the assumptions of the theorem the sequence of functions $(g_n)_{n\in\mathbb{N}}$ with $g^{(n)}(\cdot) = h(\cdot,\cdot,\widehat{\theta}_n) - \mathbb{E}(h(X_1^*, X_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n)$ is uniformly integrable and satisfies the smoothness property presumed in Lemma 5.1. Finally, bootstrap consistency follows from Lemma 3.2. □

5.2. Proofs of auxiliary results

First, we derive a weak law of large numbers for smooth functions of triangular arrays of τ-dependent random variables.

Lemma 5.1 (Weak law of large numbers). *Let $(X_{n,k})_{k=1}^n$, $n \in \mathbb{N}$, be a triangular scheme of (row-wise) stationary, $\mathbb{R}^d$-valued, integrable random variables such that* $\lim_{K\to\infty} \sup_{n\in\mathbb{N}} P(\|X_{n,1}\|_{l_1} > K) = 0$. *Suppose that the coefficients $\bar{\tau}_r := \sup_{n>r} \tau_{r,n}$ satisfy $\bar{\tau}_r \longrightarrow_{r\to\infty} 0$, where*

$$\tau_{r,n} := \sup\{\tau(\sigma(X_{n,s_1},\ldots,X_{n,s_u}), (X'_{n,t_1}, X'_{n,t_2}, X'_{n,t_3})')|u \in \mathbb{N},$$
$$1 \le s_1 \le \cdots \le s_u < s_u + r \le t_1 \le t_2 \le t_3 \le n\}.$$

Moreover, suppose that the functions $g^{(n)}: \mathbb{R}^d \to \mathbb{R}^p$ with $\mathbb{E}g^{(n)}(X_{n,1}) = 0_p$ are uniformly Lipschitz continuous on any bounded interval. If additionally the sequence $(g^{(n)}(X_{n,1}))_{n\in\mathbb{N}}$ is uniformly integrable, then

$$\frac{1}{n}\sum_{k=1}^{n} g^{(n)}(X_{n,k}) \stackrel{P}{\longrightarrow} 0_p.$$

Proof. W.l.o.g. let $p = 1$. We prove that for arbitrary $\varepsilon, \eta > 0$ there exists an n_0 such that for all $n > n_0$ the inequality $P(|n^{-1}\sum_{k=1}^n g^{(n)}(X_{n,k})| > \varepsilon) \le \eta$ holds. To this end, a truncation argument is invoked. Let w_K denote a Lipschitz continuous, nonnegative function that is bounded from above by one such that $w_K(x) = 1$ for $x \in [-K, K]^d$ and $w_K(x) = 0$ for $x \notin [-K-1, K+1]^d$ with $K \in \mathbb{R}_+$. For a finite constant M, that is specified later, define functions $g_{M,K}^{(n)}: \mathbb{R}^d \to \mathbb{R}$ by

$$g_{M,K}^{(n)}(x) := \begin{cases} g^{(n)}(x)w_K(x) & \text{for } \left|g^{(n)}(x)w_K(x)\right| \le M, \\ -M & \text{for } g^{(n)}(x)w_K(x) < -M, \\ M & \text{for } g^{(n)}(x)w_K(x) > M \end{cases}$$

and $g_{M,K}^{(n,c)}$ by $g_{M,K}^{(n,c)}(x) = g_{M,K}^{(n)}(x) - \mathbb{E}g_{M,K}^{(n)}(X_{n,1})$. This allows for the estimation

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} g^{(n)}(X_{n,k})\right| > \varepsilon\right) \le P\left(\left|\frac{1}{n}\sum_{k=1}^{n} g^{(n)}(X_{n,k}) - g_{M,K}^{(n)}(X_{n,k})\right| > \frac{\varepsilon}{3}\right) + P\left(\left|\mathbb{E}g_{M,K}^{(n)}(X_{n,1})\right| > \frac{\varepsilon}{3}\right) + P\left(\left|\frac{1}{n}\sum_{k=1}^{n} g_{M,K}^{(n,c)}(X_{n,k})\right| > \frac{\varepsilon}{3}\right).$$

According to Markov's inequality, the first summand on the r.h.s. can be bounded by

$$\frac{3}{\varepsilon}\Big[\sup_{n\in\mathbb{N}} \mathbb{E}\big|g^{(n)}(X_{n,1})\big| \mathbb{1}_{|g^{(n)}(X_{n,1})|>M} + M \sup_{n\in\mathbb{N}} P(\|X_{n,1}\|_{l_1} > K)\Big].$$

Since the functions $g^{(n)}$, $n \in \mathbb{N}$, are centered, we additionally obtain

$$\begin{aligned}
&P\left(\left|\mathbb{E}g_{M,K}^{(n)}(X_{n,1})\right| > \frac{\varepsilon}{3}\right) \\
&\quad\le P\left(\sup_{n\in\mathbb{N}} \mathbb{E}\big|g_{M,K}^{(n)}(X_{n,1}) - g^{(n)}(X_{n,1})\big| > \frac{\varepsilon}{3}\right) \\
&\quad\le P\left(\sup_{n\in\mathbb{N}} \mathbb{E}\big|g^{(n)}(X_{n,1})\big| \mathbb{1}_{|g^{(n)}(X_{n,1})|>M} + M \sup_{n\in\mathbb{N}} P(\|X_{n,1}\|_{l_1} > K) > \frac{\varepsilon}{3}\right).
\end{aligned}$$

Therefore, by choosing M and $K = K(M)$ sufficiently large, we get

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} g^{(n)}(X_{n,k}) - g_{M,K}^{(n)}(X_{n,k})\right| > \frac{\varepsilon}{3}\right) + P\left(\left|\mathbb{E}g_{M,K}^{(n)}(X_{n,1})\right| > \frac{\varepsilon}{3}\right) \le \frac{\eta}{2}.$$

Concerning the remaining term, Chebyshev's inequality leads to

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} g_{M,K}^{(n,c)}(X_{n,k})\right| > \frac{\varepsilon}{3}\right) \le \frac{9M^2}{\varepsilon^2 n} + \frac{18}{\varepsilon^2 n^2}\sum_{j<k} \mathbb{E}g_{M,K}^{(n,c)}(X_{n,j}) g_{M,K}^{(n,c)}(X_{n,k}).$$

Thus, it remains to derive an upper bound for $n^{-2}\sum_{j<k} |\mathbb{E}g_{M,K}^{(n,c)}(X_{n,j}) g_{M,K}^{(n,c)}(X_{n,k})|$ that vanishes asymptotically. For this purpose, we introduce a copy $\widetilde{X}_{n,k}$ of $X_{n,k}$, that is independent of $X_{n,j}$ and such that $\mathbb{E}\|X_{n,k} - \widetilde{X}_{n,k}\|_{l_1} \le \tau_{k-j,n}$. Due to their construction, the functions $g_{M,K}^{(n,c)}$ are Lipschitz continuous uniformly in n and with a constant $C(M, K)$. This implies

$$\begin{aligned}
\frac{1}{n^2}\sum_{j<k} \big|\mathbb{E}g_{M,K}^{(n,c)}(X_{n,j}) g_{M,K}^{(n,c)}(X_{n,k})\big| &\le \frac{2M}{n^2}\sum_{j<k} \mathbb{E}\big|g_{M,K}^{(n,c)}(X_{n,k}) - g_{M,K}^{(n,c)}(\widetilde{X}_{n,k})\big| \\
&\le \frac{2MC(M,K)}{n}\sum_{r=1}^{n} \bar{\tau}_r,
\end{aligned}$$

where the remaining term converges to zero according to Cauchy's limit theorem, cf. Knopp [22]. □

In order to prove Lemma 2.1, Lemma 2.2, and Lemma 2.3, an approximation of terms of the structure

$$Z_n := \frac{1}{n^2} \sum_{\substack{i,j,k,l=1 \\ i \neq j; k \neq l}}^{n} \mathbb{E} H(X_i, X_j) H(X_k, X_l)$$

is required. Here, H denotes a symmetric, degenerate kernel function. Assuming that $(X_n)_{n\in\mathbb{N}}$ satisfies (A1), we obtain

$$Z_n \leq \frac{8}{n^2} \sum_{i<j; k<l; i \leq k}^{n} |\mathbb{E} H(X_i, X_j) H(X_k, X_l)| \leq 8 \sup_{1 \leq k < n} \mathbb{E}|H(X_1, X_{1+k})|^2 + \frac{8}{n^2} \sum_{r=1}^{n-1} \sum_{t=1}^{4} Z_{n,r}^{(t)}$$

with

$$Z_{n,r}^{(1)} := \sum_{\substack{1 \leq i < j; k < l; j \leq l \leq n \\ r := \min\{j,k\} - i \geq l - \max\{j,k\}}} \left| \mathbb{E} H(X_i, X_j) H(X_k, X_l) - \mathbb{E} H\big(X_i, \widetilde{X}_j^{(r)}\big) H\big(\widetilde{X}_k^{(r)}, \widetilde{X}_l^{(r)}\big) \right|,$$

$$Z_{n,r}^{(2)} := \sum_{\substack{1 \leq i < j; i \leq k; k < l \leq n \\ r := l - \max\{j,k\} > \min\{j,k\} - i}} \left| \mathbb{E} H(X_i, X_j) H(X_k, X_l) - \mathbb{E} H(X_i, X_j) H\big(X_k, \widetilde{X}_l^{(r)}\big) \right|,$$

$$Z_{n,r}^{(3)} := \sum_{\substack{1 \leq i \leq k < l < j \leq n \\ r := k - i \geq j - l}} \left| \mathbb{E} H(X_i, X_j) H(X_k, X_l) - \mathbb{E} H\big(X_i, \widetilde{X}_j^{(r)}\big) H\big(\widetilde{X}_k^{(r)}, \widetilde{X}_l^{(r)}\big) \right|,$$

$$Z_{n,r}^{(4)} := \sum_{\substack{1 \leq i \leq k < l < j \leq n \\ r := j - l > k - i}} \left| \mathbb{E} H(X_i, X_j) H(X_k, X_l) - \mathbb{E} H\big(X_i, \widetilde{X}_j^{(r)}\big) H(X_k, X_l) \right|.$$

Here, in every summand of $Z_{n,r}^{(1)}$ and $Z_{n,r}^{(3)}$ the vector $(\widetilde{X}_j^{(r)\prime}, \widetilde{X}_k^{(r)\prime}, \widetilde{X}_l^{(r)\prime})'$ is chosen such that it is independent of the random variable X_i, $(\widetilde{X}_j^{(r)\prime}, \widetilde{X}_k^{(r)\prime}, \widetilde{X}_l^{(r)\prime})' \stackrel{d}{=} (X_j', X_k', X_l')'$, and (2.3) holds. Within $Z_{n,r}^{(2)}$ (resp., $Z_{n,r}^{(4)}$), the random variable $\widetilde{X}_l^{(r)}$ (resp., $\widetilde{X}_j^{(r)}$) is chosen to be independent of the vector $(X_i', X_j', X_k')'$ (resp., $(X_i', X_k', X_l')'$) such that $\widetilde{X}_l^{(r)} \stackrel{d}{=} X_l$ (resp., $\widetilde{X}_j^{(r)} \stackrel{d}{=} X_j$) and (2.3) holds. This may possibly require an enlargement of the underlying probability space. Moreover, note that the subtrahends of these expressions vanish due to the degeneracy of H and that the number of summands of $Z_{n,r}^{(t)}, t = 1, \ldots, 4$, is bounded by $(r+1)n^2$. For sake of notational simplicity, the upper index r is omitted in the sequel.

Proof of Lemma 2.1. For $c > 0$, we define $c_h := \max_{x,y\in[-c,c]^d} |h(x, y)|$,

$$\widetilde{h}^{(c)}(x, y) := \begin{cases} h(x, y) & \text{for } |h(x, y)| \le c_h, \\ -c_h & \text{for } h(x, y) < -c_h, \\ c_h & \text{for } h(x, y) > c_h \end{cases}$$

and its degenerate version

$$\begin{aligned} h_c(x, y) := {}& \widetilde{h}^{(c)}(x, y) - \int_{\mathbb{R}^d} \widetilde{h}^{(c)}(x, y) P_X(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{h}^{(c)}(x, y) P_X(\mathrm{d}y) \\ & + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}^{(c)}(x, y) P_X(\mathrm{d}x) P_X(\mathrm{d}y). \end{aligned}$$

The approximation error $n^2\mathbb{E}(U_n - U_{n,c})^2$ can be reformulated in terms of Z_n with kernel $H = H^{(c)} := h - h^{(c)}$. Hence, it remains to verify that $\sup_{k\in\mathbb{N}} \mathbb{E}|H^{(c)}(H_1, X_{1+k})|^2$ and $\sup_{n\in\mathbb{N}} n^{-2} \sum_{r=1}^{n-1} \sum_{t=1}^{4} Z_{n,r}^{(t)}$ tend to zero as $c \to \infty$. First, we consider $\sup_{n\in\mathbb{N}} n^{-2} \sum_{r=1}^{n-1} Z_{n,r}^{(1)}$, the remaining quantities can be treated similarly. The summands of $Z_{n,r}^{(1)}$ are bounded as follows:

$$\begin{aligned} &\big|\mathbb{E}H^{(c)}(X_i, X_j)H^{(c)}(X_k, X_l) - \mathbb{E}H^{(c)}(X_i, \widetilde{X}_j)H^{(c)}(\widetilde{X}_k, \widetilde{X}_l)\big| \\ &\quad \le \mathbb{E}\big|H^{(c)}(X_k, X_l)\big[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big]\mathbb{1}_{(X_k', X_l')'\in[-c,c]^{2d}}\big| \\ &\qquad + \mathbb{E}\big|H^{(c)}(X_k, X_l)\big[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big]\mathbb{1}_{(X_k', X_l')'\notin[-c,c]^{2d}}\big| \\ &\qquad + \mathbb{E}\big|H^{(c)}(X_i, \widetilde{X}_j)\big[H^{(c)}(X_k, X_l) - H^{(c)}(\widetilde{X}_k, \widetilde{X}_l)\big]\mathbb{1}_{(X_i', \widetilde{X}_j')'\in[-c,c]^{2d}}\big| \\ &\qquad + \mathbb{E}\big|H^{(c)}(X_i, \widetilde{X}_j)\big[H^{(c)}(X_k, X_l) - H^{(c)}(\widetilde{X}_k, \widetilde{X}_l)\big]\mathbb{1}_{(X_i', \widetilde{X}_j')'\notin[-c,c]^{2d}}\big| \\ &\quad = E_1 + E_2 + E_3 + E_4. \end{aligned} \tag{5.4}$$

The functions $H^{(c)}$ are obviously Lipschitz continuous uniformly in c. Therefore, an iterative application of Hölder's inequality to E_2 yields

$$\begin{aligned} E_2 &\le \big(\mathbb{E}\big|H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big|\big)^{\delta} \\ &\quad \times \big(\mathbb{E}\big|H^{(c)}(X_k, X_l)\big|^{1/(1-\delta)}\big|H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big|\mathbb{1}_{(X_k', X_l')'\notin[-c,c]^{2d}}\big)^{1-\delta} \\ &\le C\tau_r^{\delta}\big\{\big(\mathbb{E}\big|H^{(c)}(X_k, X_l)\big|^{(2-\delta)/(1-\delta)}\mathbb{1}_{(X_k', X_l')'\notin[-c,c]^{2d}}\big)^{1/(2-\delta)} \\ &\quad \times \big(\mathbb{E}\big|H^{(c)}(X_i, X_j)\big|^{(2-\delta)/(1-\delta)} + \mathbb{E}\big|H^{(c)}(X_i, \widetilde{X}_j)\big|^{(2-\delta)/(1-\delta)}\big)^{(1-\delta)/(2-\delta)}\big\}^{1-\delta} \\ &\le C\tau_r^{\delta}\big(\mathbb{E}\big|H^{(c)}(X_k, X_l)\big|^{(2-\delta)/(1-\delta)}\mathbb{1}_{(X_k', X_l')'\notin[-c,c]^{2d}}\big)^{(1-\delta)/(2-\delta)}. \end{aligned} \tag{5.5}$$

As $\sup_{k\in\mathbb{N}} \mathbb{E}|h(X_1, X_{1+k})|^{\nu} < \infty$ for $\nu > (2 - \delta)/(1 - \delta)$, we obtain $E_2 \le \tau_r^{\delta}\varepsilon_1(c)$ with $\varepsilon_1(c) \longrightarrow_{c\to\infty} 0$ after employing Hölder's inequality once again. Analogous calculations yield

$E_4 \le \tau_r^\delta \varepsilon_2(c)$ with $\varepsilon_2(c) \longrightarrow_{c\to\infty} 0$. Likewise, the approximation methods for E_1 and E_3 are equal. Therefore, only E_1 is considered:

$$
\begin{aligned}
E_1 &\le \mathbb{E}\left|\int_{\mathbb{R}^d} \widetilde{h}^{(c)}(X_k, y) P_X(\mathrm{d}y)\big[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big]\mathbb{1}_{X_k\in[-c,c]^d}\right| \\
&\quad + \mathbb{E}\left|\int_{\mathbb{R}^d} \widetilde{h}^{(c)}(y, X_l) P_X(\mathrm{d}y)\big[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big]\mathbb{1}_{X_l\in[-c,c]^d}\right| \\
&\quad + \mathbb{E}\left|\iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}^{(c)}(x, y) P_X(\mathrm{d}x) P_X(\mathrm{d}y)\big[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\big]\right| \\
&= E_{1,1} + E_{1,2} + E_{1,3}.
\end{aligned}
$$

Analogous to (5.5), we obtain

$$
\begin{aligned}
E_{1,1} &\le C\tau_r^\delta\left\{\left(\mathbb{E}\left|\int_{\mathbb{R}^d} h(X_k, y) - \widetilde{h}^{(c)}(X_k, y) P_X(\mathrm{d}y)\right|^{(2-\delta)/(1-\delta)} \mathbb{1}_{X_k\in[-c,c]^d}\right)^{1/(2-\delta)}\right. \\
&\qquad \left.\times\left[\sup_{k\in\mathbb{N}} \mathbb{E}\big|H^{(c)}(X_1, X_{1+k})\big|^{(2-\delta)/(1-\delta)} + \mathbb{E}\big|H^{(c)}(X_i, \widetilde{X}_j)\big|^{(2-\delta)/(1-\delta)}\right]^{(1-\delta)/(2-\delta)}\right\}^{1-\delta} \\
&\le C\tau_r^\delta\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big|h(x, y) - \widetilde{h}^{(c)}(x, y)\big|^{(2-\delta)/(1-\delta)}\right. \\
&\qquad\qquad \left.\times P_X(\mathrm{d}y)\mathbb{1}_{x\in[-c,c]^d} P_X(\mathrm{d}x)\right)^{(1-\delta)/(2-\delta)} \\
&\le \tau_r^\delta \varepsilon_3(c)
\end{aligned}
$$

with $\varepsilon_3(c) \longrightarrow_{c\to\infty} 0$. The estimation of $E_{1,2}$ coincides with the previous one. The expression $E_{1,3}$ can be bounded as follows:

$$
\begin{aligned}
E_{1,3} &\le C\tau_r \iint_{\mathbb{R}^d\times\mathbb{R}^d} \big|h(x, y) - \widetilde{h}^{(c)}(x, y)\big| P_X(\mathrm{d}x) P_X(\mathrm{d}y) \\
&\le C\tau_r \iint_{\mathbb{R}^d\times\mathbb{R}^d} |h(x, y)|\mathbb{1}_{(x', y')'\notin[-c,c]^{2d}} P_X(\mathrm{d}x) P_X(\mathrm{d}y) \\
&\le \tau_r \varepsilon_4(c)
\end{aligned}
$$

with $\varepsilon_4(c) \longrightarrow_{c\to\infty} 0$. To sum up, we have $E_1 + E_2 + E_3 + E_4 \le \varepsilon_5(c)\tau_r^\delta$, where $\varepsilon_5(c) \longrightarrow_{c\to\infty} 0$ uniformly in n. This leads to

$$
\lim_{c\to\infty}\sup_{n\in\mathbb{N}} \frac{1}{n^2}\sum_{r=1}^{n-1} Z_{n,r}^{(1)} \le \lim_{c\to\infty}\sup_{n\in\mathbb{N}} \frac{1}{n^2}\sum_{r=1}^{n-1}(r+1)n^2\tau_r^\delta\varepsilon_5(c) = 0.
$$

It remains to examine

$$\sup_{k\in\mathbb{N}} \mathbb{E}\big[H^{(c)}(X_1, X_{1+k})\big]^2 \le C\Big(\sup_{k\in\mathbb{N}} \mathbb{E}\big[h(X_1, X_{1+k}) - \widetilde{h}^{(c)}(X_1, X_{1+k})\big]^2 + \mathbb{E}\big[h(X_1, \widetilde{X}_1) - \widetilde{h}^{(c)}(X_1, \widetilde{X}_1)\big]^2\Big).$$

Here, $\widetilde{X}_1$ denotes an independent copy of X_1. Similar arguments as before yield $\lim_{c\to\infty} \sup_{k\in\mathbb{N}} \mathbb{E}[H^{(c)}(X_1, X_{1+k})]^2 = 0$. □

The characteristics stated in the following two lemmas will be essential for a wavelet approximation of the kernel function h.

Lemma 5.2. *Given a Lipschitz continuous function* $g:\mathbb{R}^d \to \mathbb{R}$, *define a wavelet series approximation* g_j *by* $g_j(x) := \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}\Phi_{j,k}(x)$, $j\in\mathbb{Z}$, *where* $\alpha_{j,k} = \int_{\mathbb{R}^d} g(x)\Phi_{j,k}(x)\,\mathrm{d}x$. *Then* g_j *is Lipschitz continuous with a constant that is independent of* j.

Proof. In order to establish Lipschitz continuity, the function g_j is decomposed into two parts

$$\begin{aligned} g_j(x) &= \sum_{k\in\mathbb{Z}^d}\Big[\int_{\mathbb{R}^d}\Phi_{j,k}(u)g(x)\,\mathrm{d}u\Big]\Phi_{j,k}(x) + \sum_{k\in\mathbb{Z}^d}\Big[\int_{\mathbb{R}^d}\Phi_{j,k}(u)[g(u)-g(x)]\,\mathrm{d}u\Big]\Phi_{j,k}(x) \\ &= H_1(x) + H_2(x). \end{aligned}$$

According to the above choice of the scale function (with characteristics (1)–(3) of Section 2.2), the prerequisites of Corollary 8.1 of Härdle *et al.* [20] are fulfilled for $N=1$. This implies that $\int_{-\infty}^{\infty}\sum_{l\in\mathbb{Z}}\phi(y-l)\phi(z-l)\,\mathrm{d}z = 1$, $\forall y\in\mathbb{R}$. Based on this result, we obtain

$$\sum_{k\in\mathbb{Z}^d}\int_{\mathbb{R}^d}\Phi_{j,k}(u)\Phi_{j,k}(x)\,\mathrm{d}u = 2^{jd}\prod_{i=1}^{d}\int_{-\infty}^{\infty}\sum_{l\in\mathbb{Z}}\phi(2^j u_i - l)\phi(2^j x_i - l)\,\mathrm{d}u_i = 1 \qquad \forall x\in\mathbb{R}^d,$$

by applying an appropriate variable substitution. To this end, note that for every fixed x, the number of non-vanishing summands can be bounded by a finite constant uniformly in j because of the finite support of ϕ. Therefore, the order of summation and integration is interchangeable. Hence, $H_1 = g$ which in turn immediately implies the desired continuity property for H_1.

In order to investigate H_2, we define a sequence of functions $(\kappa_k)_{k\in\mathbb{Z}}$ by

$$\kappa_k(x) = \int_{\mathbb{R}^d}\Phi_{j,k}(u)[g(u)-g(x)]\,\mathrm{d}u.$$

These functions are Lipschitz continuous with a constant decreasing in j:

$$|\kappa_k(x) - \kappa_k(\bar{x})| \le \mathrm{Lip}(g)\mathrm{O}(2^{-jd/2})\|x-\bar{x}\|_{l_1}. \tag{5.6}$$

Moreover, boundedness and Lipschitz continuity of ϕ yield

$$\|\Phi_{j,k}\|_\infty = \mathrm{O}(2^{jd/2}) \quad \text{and} \quad |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| = \mathrm{O}\big(2^{j(d/2+1)}\big)\|x-\bar{x}\|_{l_1}. \tag{5.7}$$

Thus,

$$\begin{aligned}|H_2(x) - H_2(\bar{x})| &\le \sum_{k\in\mathbb{Z}^d} |\Phi_{j,k}(x)||\kappa_k(x) - \kappa_k(\bar{x})| \\ &\quad + \sum_{k\in\mathbb{Z}^d} |\kappa_k(\bar{x})||\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \\ &\le C\|x - \bar{x}\|_{l_1} + \sum_{k\in\mathbb{Z}^d} |\kappa_k(\bar{x})||\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})|.\end{aligned}$$

Now, it has to be distinguished whether or not $\bar{x} \in \operatorname{supp}(\Phi_{j,k})$ in order to approximate the second summand. (Here, supp denotes the support of a function.) In the first case, it is helpful to illuminate $|\kappa_k(\bar{x})| = |\int_{\mathbb{R}^d} \Phi_{j,k}(u)[g(u) - g(\bar{x})]\,\mathrm{d}u|$. The integrand is non-trivial only if $u \in \operatorname{supp}(\Phi_{j,k})$. In these situations, $|g(u) - g(\bar{x})| = \mathrm{O}(2^{-j})$ by Lipschitz continuity. Consequently, we get

$$|\kappa_k(\bar{x})| \le \mathrm{O}(2^{-j}) \int_{\mathbb{R}^d} |\Phi_{j,k}(u)|\,\mathrm{d}u = \mathrm{O}\big(2^{-j(d/2+1)}\big)$$

which leads to

$$\sum_{k\in\mathbb{Z}^d} |\kappa_k(\bar{x})||\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \le C\|x - \bar{x}\|_{l_1}$$

as the number of nonvanishing summands is finite, independently of the values of x and $\bar{x}$. Therefore, Lipschitz continuity of H_2 is obtained as long as $\bar{x} \in \operatorname{supp}(\Phi_{j,k})$.

In the opposite case, we only have to consider the situation of $x \in \operatorname{supp}(\Phi_{j,k})$ since the setting $\bar{x}, x \notin \operatorname{supp}(\Phi_{j,k})$ is trivial. With the aid of (5.6) and (5.7), the first term of the r.h.s. of

$$|\kappa_k(\bar{x})[\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})]| \le |\kappa_k(\bar{x}) - \kappa_k(x)||\Phi_{j,k}(x)| + |\kappa_k(x)||\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \quad (5.8)$$

can be estimated from above by $C\|x - \bar{x}\|_{l_1}$. The investigation of the second summand is identical to the analysis of the case $\bar{x} \in \operatorname{supp}(\Phi_{j,k})$.

Finally, we obtain $|H_2(x) - H_2(\bar{x})| \le C\|x - \bar{x}\|_{l_1}$, where $C < \infty$ is a constant that is independent of j. This yields the assertion of the lemma. □

Lemma 5.3. *Let $g:\mathbb{R}^d \to \mathbb{R}$ be a function that is continuous on some interval $(-c, c)^d$. For arbitrary $b \in (0, c)$ and $K \in \mathbb{N}$ there exists a $J(K, b, c) \in \mathbb{N}$ such that for g and its approximation g_J given by $g_J(x) = \sum_{k\in\mathbb{Z}^d} \alpha_{J,k}\Phi_{J,k}(x)$ it holds*

$$\max_{x\in[-b,b]^d} |g(x) - g_J(x)| \le 1/K \qquad \forall J \ge J(K, b, c).$$

Proof. Given $b \in (0, c)$, we define $\bar{g}^{(b,c)}(x) := g(x)w_{b,c}(x)$, where $w_{b,c}$ is a Lipschitz continuous and nonnegative weight function with compact support $S_w \subset (-c, c)^d$. Moreover, $w_{b,c}$ is assumed to be bounded from above by 1 and $w_{b,c}(x) := 1$ for $x \in (-b - \delta, b + \delta)^d$ for some

$\delta > 0$ with $b + \delta < c$. Additionally, we set $\alpha_{J,k}^{(b,c)} := \int_{\mathbb{R}^d} \bar{g}^{(b,c)}(u)\Phi_{J,k}(u)\,\mathrm{d}u$. Hence,

$$\begin{aligned}
&\max_{x\in[-b,b]^d} |g(x) - g_J(x)| \\
&\quad \leq \max_{x\in[-b,b]^d} \left| \bar{g}^{(b,c)}(x) - \sum_{k\in\mathbb{Z}^d} \alpha_{J,k}^{(b,c)} \Phi_{J,k}(x) \right| + \max_{x\in[-b,b]^d} \left| \sum_{k\in\mathbb{Z}^d} \alpha_{J,k}^{(b,c)} \Phi_{J,k}(x) - g_J(x) \right| \\
&\quad = \max_{x\in[-b,b]^d} A^{(J)}(x) + \max_{x\in[-b,b]^d} B^{(J)}(x).
\end{aligned}$$

Since $\bar{g}^{(b,c)} \in C_0(\mathbb{R}^d)$, Theorem 8.4 of Wojtaszczyk [29] implies that there exists a $J_0(K,b,c) \in \mathbb{N}$ such that $\max_{x\in[-b,b]^d} A^{(J)}(x) \leq 1/K$ for all $J \geq J_0(K,b,c)$. Moreover, the introduction of the finite set of indices

$$\bar{Z}(J) := \{k \in \mathbb{Z}^d | \Phi_{J,k}(x) \neq 0 \text{ for some } x \in [-b,b]^d\}$$

leads to

$$\max_{x\in[-b,b]^d} B^{(J)}(x) = \max_{x\in[-b,b]^d} \left| \sum_{k\in\bar{Z}(J)} \bigl(\alpha_{J,k} - \alpha_{J,k}^{(b,c)}\bigr)\Phi_{J,k}(x) \right|.$$

This term is equal to zero for all $J \geq J(K,b,c)$ and some $J(K,b,c) \geq J_0(K,b,c)$ since the definition of $\bar{g}^{(b,c)}$ implies $\alpha_{J,k} = \alpha_{J,k}^{(b,c)}$, $\forall k \in \bar{Z}$, for all sufficiently large J. □

Proof of Lemma 2.2. The assertion of the lemma is verified in two steps. First, the bounded kernel h_c, constructed in the proof of Lemma 2.1, is approximated by $\widetilde{h}_c^{(K)}$ which is defined by $\widetilde{h}_c^{(K)}(x,y) = \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{J(K);k_1,k_2}^{(c)} \Phi_{J(K),k_1}(x)\Phi_{J(K),k_2}(y)$ with $\alpha_{J(K);k_1,k_2}^{(c)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y)\Phi_{J(K),k_1}(x)\Phi_{J(K),k_2}(y)\,\mathrm{d}x\,\mathrm{d}y$. Here, the indices $(J(K))_{K\in\mathbb{N}}$ with $J(K) \longrightarrow_{K\to\infty} \infty$ are chosen such that the assertion of Lemma 5.3 holds true for $b = b(K) \in \mathbb{R}$ with $P(X_1 \notin [-b,b]^d) \leq K^{-1}$ and $c = 2b$. Since the function $\widetilde{h}_c^{(K)}$ is not degenerate in general, we introduce its degenerate counterpart

$$\begin{aligned}
h_c^{(K)}(x,y) = \widetilde{h}_c^{(K)}(x,y) &- \int_{\mathbb{R}^d} \widetilde{h}_c^{(K)}(x,y) P_X(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(K)}(x,y) P_X(\mathrm{d}y) \\
&+ \iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}_c^{(K)}(x,y) P_X(\mathrm{d}x) P_X(\mathrm{d}y)
\end{aligned}$$

and denote the corresponding U-statistic by $U_{n,c}^{(K)}$.

Now, the structure of the proof is as follows. First, we prove

$$\sup_{n\in\mathbb{N}} n^2 \mathbb{E}\bigl(U_{n,c} - U_{n,c}^{(K)}\bigr)^2 \underset{K\to\infty}{\longrightarrow} 0. \tag{5.9}$$

In a second step, it remains to show that for every fixed K

$$\sup_{n\in\mathbb{N}} n^2 \mathbb{E}\bigl(U_{n,c}^{(K)} - U_{n,c}^{(K,L)}\bigr)^2 \underset{L\to\infty}{\longrightarrow} 0. \tag{5.10}$$

In order to verify (5.9), we rewrite $n^2\mathbb{E}(U_{n,c} - U_{n,c}^{(K)})^2$ in terms of Z_n with kernel function $H := H^{(K)} = h_c - h_c^{(K)}$. Hence, it remains to verify that $\sup_{n\in\mathbb{N}} n^{-2}\sum_{r=1}^{n-1}\sum_{t=1}^{4} Z_{n,r}^{(t)}$ and $\sup_{k\in\mathbb{N}}\mathbb{E}|H^{(K)}(H_1, X_{1+k})|^2$ tend to zero as $K \to \infty$. Exemplarily, we investigate $\sup_{n\in\mathbb{N}} n^{-2}\sum_{r=1}^{n-1} Z_{n,r}^{(1)}$. The summands of $Z_{n,r}^{(1)}$ can be bounded as follows:

$$\begin{aligned}
&\big|\mathbb{E}H^{(K)}(X_i, X_j)H^{(K)}(X_k, X_l) - H^{(K)}(X_i, \widetilde{X}_j)H^{(K)}(\widetilde{X}_k, \widetilde{X}_l)\big| \\
&\quad\le \mathbb{E}\big|H^{(K)}(X_k, X_l)\big[H^{(K)}(X_i, X_j) - H^{(K)}(X_i, \widetilde{X}_j)\big]\big| \\
&\qquad + \mathbb{E}\big|H^{(K)}(X_i, \widetilde{X}_j)\big[H^{(K)}(X_k, X_l) - H^{(K)}(\widetilde{X}_k, \widetilde{X}_l)\big]\big|.
\end{aligned}$$

Since further approximations are similar for both summands, we concentrate on the first one. Note that boundedness of h_c implies uniform boundedness of $(H^{(K)})_K$ due to the compact support of the function ϕ. Moreover, the constant $\mathrm{Lip}(H^{(K)})$ does not depend on K in consequence of Lemma 5.2. Therefore, the application of Hölder's inequality leads to

$$\mathbb{E}\big|H^{(K)}(X_k, X_l)\big[H^{(K)}(X_i, X_j) - H^{(K)}(X_i, \widetilde{X}_j)\big]\big| \le C\tau_r^{\delta}\big[\mathbb{E}\big|H^{(K)}(X_k, X_l)\big|^{1/(1-\delta)}\big]^{1-\delta}.$$

The construction of the sequence $(b(K))_K$ above allows for the following estimation:

$$\begin{aligned}
&\mathbb{E}\big|H^{(K)}(X_k, X_l)\big|^{1/(1-\delta)} \\
&\quad = \mathbb{E}\big|H^{(K)}(X_k, X_l)\big|^{1/(1-\delta)}\mathbb{1}_{X_k, X_l\in[-b(K),b(K)]^d} + \mathrm{O}\big(P\big(X_1 \notin [-b(K), b(K)]^d\big)\big) \\
&\quad \le \sup_{x,y\in[-b(K),b(K)]^d}\big|H^{(K)}(x, y)\big|^{1/(1-\delta)} + \frac{C}{K}.
\end{aligned}$$

According to Lemma 5.3 and the above choice of the sequence $(b(K))_K$, we obtain

$$\begin{aligned}
&\sup_{x,y\in[-b(K),b(K)]^d}\big|H^{(K)}(x, y)\big| \\
&\quad \le \frac{1}{K} + 2\sup_{x,y\in[-b(K),b(K)]^d}\mathbb{E}\big|h_c(x, X_1) - \widetilde{h}_c^{(K)}(x, X_1)\big| \\
&\qquad + \bigg|\iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x, y) - \widetilde{h}_c^{(K)}(x, y)P_X(\mathrm{d}x)P_X(\mathrm{d}y)\bigg| \\
&\quad \le \frac{4}{K} + 2\sup_{x\in[-b(K),b(K)]^d}\mathbb{E}\big|h_c(x, X_1) - \widetilde{h}_c^{(K)}(x, X_1)\big|\mathbb{1}_{X_1\notin[-b(K),b(K)]^d} \\
&\qquad + 2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d\setminus[-b(K),b(K)]^d}\big|h_c(x, y) - \widetilde{h}_c^{(K)}(x, y)\big|P_X(\mathrm{d}x)P_X(\mathrm{d}y) \\
&\quad \le \frac{C}{K}.
\end{aligned}$$

Consequently,

$$\big|\mathbb{E}H^{(K)}(X_i, X_j)H^{(K)}(X_k, X_l) - \mathbb{E}H^{(K)}(X_i, \widetilde{X}_j)H^{(K)}(\widetilde{X}_k, \widetilde{X}_l)\big| \le C\varepsilon_K \tau_r^\delta$$

for some null sequence $(\varepsilon_K)_K$. This implies that $\sup_{n\in\mathbb{N}} n^{-2}\sum_{r=1}^n Z_{n,r}^{(1)}$ tends to zero as K increases. Furthermore, one obtains $\sup_{k\in\mathbb{N}} \mathbb{E}[H^{(K)}(X_1, X_{1+k})]^2 = \mathrm{O}(K^{-1})$ similarly to the consideration of $\mathbb{E}|H^{(K)}(X_k, X_l)|^{1/(1-\delta)}$ above. Thus, we get $\sup_n n^2\mathbb{E}(U_{n,c} - U_{n,c}^{(K)})^2 \longrightarrow_{K\to\infty} 0$.

The main goal of the previous step was the multiplicative separation of the random variables which are cumulated in h_c. The aim of the second step is the approximation of $h_c^{(K)}$, whose representation is given by an infinite sum, by a function consisting of only finitely many summands. Similar to the foregoing part of the proof the approximation error $n^2\mathbb{E}(U_{n,c}^{(K)} - U_{n,c}^{(K,L)})^2$ is reformulated in terms of Z_n with kernel $H := H^{(L)} = h_c^{(K)} - h_c^{(K,L)}$. As before, we exemplarily take $n^{-2}\sum_{r=1}^{n-1} Z_{n,r}^{(1)}$ and $\sup_{k\in\mathbb{N}} \mathbb{E}|H^{(L)}(X_1, X_{1+k})|^2$ into further consideration. Concerning the summands of $Z_{n,r}^{(1)}$, we obtain

$$\begin{aligned}
&\big|\mathbb{E}H^{(L)}(X_i, X_j)H^{(L)}(X_k, X_l) - \mathbb{E}H^{(L)}(X_i, \widetilde{X}_j)H^{(L)}(\widetilde{X}_k, \widetilde{X}_l)\big| \\
&\quad\le \mathbb{E}\big|H^{(L)}(X_k, X_l)\big[H^{(L)}(X_i, X_j) - H^{(L)}(X_i, \widetilde{X}_j)\big]\mathbb{1}_{(X_k', X_l')' \in [-B,B]^{2d}}\big| \\
&\qquad + \mathbb{E}\big|H^{(L)}(X_k, X_l)\big[H^{(L)}(X_i, X_j) - H^{(L)}(X_i, \widetilde{X}_j)\big]\mathbb{1}_{(X_k', X_l')' \notin [-B,B]^{2d}}\big| \\
&\qquad + \mathbb{E}\big|H^{(L)}(X_i, \widetilde{X}_j)\big[H^{(L)}(X_k, X_l) - H^{(L)}(\widetilde{X}_k, \widetilde{X}_l)\big]\mathbb{1}_{(X_i', \widetilde{X}_j')' \in [-B,B]^{2d}}\big| \\
&\qquad + \mathbb{E}\big|H^{(L)}(X_i, \widetilde{X}_j)\big[H^{(L)}(X_k, X_l) - H^{(L)}(\widetilde{X}_k, \widetilde{X}_l)\big]\mathbb{1}_{(X_i', \widetilde{X}_j')' \notin [-B,B]^{2d}}\big| \\
&\quad= E_1 + E_2 + E_3 + E_4
\end{aligned}$$

for arbitrary $B > 0$. Obviously, it suffices to take the first two summands into further considerations. The both remaining terms can be treated similarly. First, note that $(H^{(L)})_L$ is uniformly bounded. Since ϕ and ψ have compact support, the number of overlapping functions within $(\Phi_{0,k})_{k\in\{-L,\dots,L\}^d}$ and $(\Psi_{j,k}^{(e)})_{k\in\{-L,\dots,L\}^d, 0\le j<J(K), e\in E}$ can be bounded by a constant that is independent of L. By Lipschitz continuity of ϕ and ψ, this leads to uniform Lipschitz continuity of $(h_c^{(K,L)})_{L\in\mathbb{N}}$. Due to the reformulation

$$\widetilde{h}_c^{(K)}(x, y) = \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{k_1,k_2}^{(c)} \Phi_{0,k_1}(x)\Phi_{0,k_2}(y) + \sum_{j=0}^{J(K)-1} \sum_{k_1,k_2\in\mathbb{Z}^d} \sum_{e\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)} \Psi_{j;k_1,k_2}^{(e)}(x, y)$$

one can choose $(B = B(K, L))_{L\in\mathbb{N}}$ such that $\max_{x,y\in[-B,B]^d} |\widetilde{h}_c^{(K)}(x, y) - \widetilde{h}_c^{(K,L)}(x, y)| = 0$ and $B(K, L) \longrightarrow_{L\to\infty} \infty$. This setting allows for the approximations

$$\begin{aligned}
E_1 &\le C\tau_r^\delta\big[\mathbb{E}\big|H^{(L)}(X_k, X_l)\big|^{1/(1-\delta)}\mathbb{1}_{(X_k', X_l')' \in [-B,B]^d}\big]^{1-\delta} \le C\tau_r^\delta\big[P(X_1 \notin [-B, B]^d)\big]^{1-\delta}, \\
E_2 &\le C\tau_r^\delta\big[P(X_1 \notin [-B, B]^{2d})\big]^{1-\delta}.
\end{aligned}$$

Analogously, it can be shown that $\sup_{k\in\mathbb{N}} \mathbb{E}[H^{(L)}(X_1, X_{1+k})]^2 \le C P(X_1 \notin [-B, B]^d)$. Finally, we obtain

$$\sup_{n\in\mathbb{N}} n^2 \mathbb{E}\big(U_{n,c}^{(K)} - U_{n,c}^{(K,L)}\big)^2 \le C\big[P(X_1 \notin [-B, B]^d)\big]^{1-\delta} \left[\sup_{n\in\mathbb{N}} \sum_{r=1}^{n-1} (r+1)\tau_r^\delta\right] \underset{L\to\infty}{\longrightarrow} 0.$$

Hence, the relations (5.9) and (5.10) hold. □

Proof of Lemma 2.3. In order to prove the assertion, we follow the lines of the proofs of Lemma 2.1, Lemma 2.2, and Lemma 5.2 and carry out some modifications.

In a first step, we reduce the problem to statistics with bounded kernels h_c defined in the proof of Lemma 2.1. To this end, we use the modified approximation

$$\begin{aligned}\big|H^{(c)}(x, y) - H^{(c)}(\bar{x}, \bar{y})\big| &\le [2f(x, \bar{x}, y, \bar{y}) + g(x, \bar{x}) + g(y, \bar{y})][\|x - \bar{x}\|_{l_1} + \|y - \bar{y}\|_{l_1}] \\ &=: f_1(x, \bar{x}, y, \bar{y})[\|x - \bar{x}\|_{l_1} + \|y - \bar{y}\|_{l_1}],\end{aligned}$$

where g is given by $g(x, \bar{x}) := \int_{\mathbb{R}^d} f(x, \bar{x}, z, z) P_X(\mathrm{d}z)$. Under (A4)(i) Hölder's inequality yields

$$\begin{aligned}&\mathbb{E}\big|H^{(c)}(Y_{k_1}, Y_{k_2}) - H^{(c)}(Y_{k_3}, Y_{k_4})\big| \\ &\quad\le \left(\mathbb{E}[f_1(Y_{k_1}, Y_{k_2}, Y_{k_3}, Y_{k_4})]^{1/(1-\delta)} \sum_{i=1}^{4} \|Y_{k_i}\|_{l_1}\right)^{1-\delta} (\mathbb{E}\|Y_{k_1} - Y_{k_3}\|_{l_1} + \mathbb{E}\|Y_{k_2} - Y_{k_4}\|_{l_1})^\delta\end{aligned}$$

for Y_{k_i} ($k_i = 1, \dots, 5, i = 1, \dots, 4$), as defined in (A4). Plugging in this inequality into the calculations of the proof of Lemma 2.1 yields $\sup_{n\in\mathbb{N}} n^2 \mathbb{E}(U_n - U_n^{(c)})^2 \longrightarrow_{c\to\infty} 0$.

The next step contains the wavelet approximation of the bounded kernel h_c. Defining $h_c^{(K)}$ and $U_{n,c}^{(K)}$ as in the proof of Lemma 2.2, analogous to the proof of Lemma 5.2 there exists a $C > 0$ such that

$$\begin{aligned}&\big|\widetilde{h}_c^{(K)}(\bar{x}, \bar{y}) - \widetilde{h}_c^{(K)}(x, y)\big| \\ &\quad\le f_1(x, \bar{x}, y, \bar{y})[\|x - \bar{x}\|_{l_1} + \|y - \bar{y}\|_{l_1}] + |H_2(\bar{x}, \bar{y}) - H_2(x, y)| \\ &\quad\le C f_1(x, \bar{x}, y, \bar{y})[\|x - \bar{x}\|_{l_1} + \|y - \bar{y}\|_{l_1}] \\ &\qquad + \sum_{k_1,k_2\in\mathbb{Z}^d} \big(|\kappa_{k_1,k_2}(\bar{x}, \bar{y})|\big|\Phi_{J(K),k_1}(x)\Phi_{J(K),k_2}(y) - \Phi_{J(K),k_1}(\bar{x})\Phi_{J(K),k_2}(\bar{y})\big|\big),\end{aligned} \tag{5.11}$$

where κ_{k_1,k_2} is given by

$$\kappa_{k_1,k_2}(x, y) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Phi_{J(K),k_1}(u)\Phi_{J(K),k_2}(v)[h_c(u, v) - h_c(x, y)]\,\mathrm{d}u\,\mathrm{d}v$$

and H_2 is defined as in the proof of Lemma 5.2. In order to approximate the last summand of (5.11), we distinguish again between the cases whether or not $(\bar{x}', \bar{y}')' \in \operatorname{supp}(\Phi_{J(K),k_1} \times$

$\Phi_{J(K),k_2})$. In the first case, an upper bound of order

$$\mathrm{O}\Big(\max_{a_1,a_2\in[-S_\phi/2^{J(K)},S_\phi/2^{J(K)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2)\Big)(\|\bar{x}-x\|_{l_1}+\|\bar{y}-y\|_{l_1})$$

can be obtained since

$$\begin{aligned}|\kappa_{k_1,k_2}(\bar{x},\bar{y})| &\le \frac{S_\phi}{2^{J(K)}}\max_{a_1,a_2\in[-S_\phi/2^{J(K)},S_\phi/2^{J(K)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2)\\ &\quad\times\iint_{\mathbb{R}^d\times\mathbb{R}^d}\big|\Phi_{J(K),k_1}(u)\Phi_{J(K),k_2}(v)\big|\,\mathrm{d}u\,\mathrm{d}v\\ &\le \mathrm{O}\big(2^{-J(K)(d+1)}\big)\max_{a_1,a_2\in[-S_\phi/2^{J(K)},S_\phi/2^{J(K)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2).\end{aligned}$$

Here, S_ϕ denotes the length of the support of ϕ. In the second case, a decomposition similar to (5.8) can be employed which leads to the upper bound

$$\mathrm{O}\Big(f_1(x,\bar{x},y,\bar{y})+\max_{a_1,a_2\in[-S_\phi/2^{J(K)},S_\phi/2^{J(K)}]^d} f_1(x,x+a_1,y,y+a_2)\Big)(\|\bar{x}-x\|_{l_1}+\|\bar{y}-y\|_{l_1}).$$

Consequently, we get

$$\begin{aligned}\big|\widetilde{h}_c^{(K)}(\bar{x},\bar{y})-\widetilde{h}_c^{(K)}(x,y)\big| &\le \mathrm{O}\Big(\max_{a_1,a_2\in[-S_\phi/2^{J(K)},S_\phi/2^{J(K)}]^d} f_1(x,x+a_1,y,y+a_2)\\ &\quad+\max_{a_1,a_2\in[-S_\phi/2^{J(K)},S_\phi/2^{J(K)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2)\\ &\quad+f_1(x,\bar{x},y,\bar{y})\Big)\times(\|\bar{x}-x\|_{l_1}+\|\bar{y}-y\|_{l_1})\\ &=: f_2(x,\bar{x},y,\bar{y})(\|\bar{x}-x\|_{l_1}+\|\bar{y}-y\|_{l_1}).\end{aligned}$$

This yields $|H^{(K)}(x,y)-H^{(K)}(\bar{x},\bar{y})|\le f_3(x,\bar{x},y,\bar{y})(\|x-\bar{x}\|_{l_1}+\|y-\bar{y}\|_{l_1})$ with $f_3(x,\bar{x},y,\bar{y})=2f_2(x,\bar{x},y,\bar{y})+\int_{\mathbb{R}^d}f_2(x,\bar{x},z,z)P_X(\mathrm{d}z)+\int_{\mathbb{R}^d}f_2(z,z,\bar{y},y)P_X(\mathrm{d}z)$. Note that under (A4)(i), $\mathbb{E}[f_3(Y_i,Y_j,Y_k,Y_l)]^\eta(\|Y_i\|_{l_1}+\|Y_j\|_{l_1}+\|Y_k\|_{l_1}+\|Y_l\|_{l_1})<\infty$ if $J(K)$ is sufficiently large. Thus, we have

$$\mathbb{E}\big|H^{(K)}(Y_{k_1},Y_{k_2})-H^{(K)}(Y_{k_3},Y_{k_4})\big|\le C(\mathbb{E}\|Y_{k_1}-Y_{k_3}\|_{l_1}+\mathbb{E}\|Y_{k_2}-Y_{k_4}\|_{l_1})^\delta$$

for Y_{k_i} ($k_i=1,\dots,5$, $i=1,\dots,4$), as defined in (A4). Moreover, Lemma 5.3 remains valid with $g=h_c$. Therefore, one can follow the lines of the proof of Lemma 2.3 and plug in the inequality above. This procedure leads to $\sup_{n\in\mathbb{N}}n^2\mathbb{E}(U_{n,c}-U_{n,c}^{(K)})^2\longrightarrow_{K\to\infty}0$.

In the third step of the proof, we verify $\sup_{n\in\mathbb{N}}n^2\mathbb{E}(U_{n,c}^{(K)}-U_{n,c}^{(K,L)})^2\longrightarrow_{L\to\infty}0$. For this purpose, it suffices to plug in a modified approximation of $H^{(L)}(x,y)-H^{(L)}(\bar{x},\bar{y})$ into the second part of the proof of Lemma 2.2. Lipschitz continuity of $h_c^{(K,L)}$ implies

$$\big|H^{(L)}(x,y)-H^{(L)}(\bar{x},\bar{y})\big|\le f_4(x,\bar{x},y,\bar{y})[\|x-\bar{x}\|_{l_1}+\|y-\bar{y}\|_{l_1}]$$

with $f_4(x, \bar{x}, y, \bar{y}) = C + f_3(x, \bar{x}, y, \bar{y})$. Since, f_4 satisfies the moment assumption of (A4)(i) with $A = 0$ for sufficiently large $J(K)$, we obtain

$$\mathbb{E}\left|H^{(L)}(Y_{k_1}, Y_{k_2}) - H^{(L)}(Y_{k_3}, Y_{k_4})\right| \leq C[\mathbb{E}(\|Y_{k_1} - Y_{k_3}\|_{l_1} + \|Y_{k_2} - Y_{k_4}\|_{l_1})]^{\delta}.$$

Hence, $\sup_{n\in\mathbb{N}} n^2 \mathbb{E}(U_{n,c}^{(K)} - U_{n,c}^{(K,L)})^2 \longrightarrow_{L\to\infty} 0$. Summing up the three steps yields

$$\lim_{c\to\infty} \limsup_{K\to\infty} \limsup_{L\to\infty} \sup_{n\in\mathbb{N}} n^2 \mathbb{E}\big(U_n - U_{n,c}^{(K,L)}\big)^2 = 0. \qquad \square$$

Proof of Lemma 3.2. A positive variance of Z implies the existence of constants $V > 0$ and $c_0 > 0$ such that for every $c \geq c_0$ we can find a $K_0 \in \mathbb{N}$ such that for every $K \geq K_0$ there is an L_0 with $\operatorname{var}(Z_c^{(K,L)}) \geq V, \forall L \geq L_0$. Moreover, uniform equicontinuity of the distribution functions of $(((Z_c^{(K,L)})_L)_K)_c$ yields the desired property of Z. By matrices-based notation of $Z_c^{(K,L)}$, we obtain

$$Z_c^{(K,L)} = C^{(K,L)} + \sum_{k_1,k_2=1}^{M(K,L)} \gamma_{k_1,k_2}^{(c,K,L)} Z_{k_1}^{(K,L)} Z_{k_2}^{(K,L)} = C^{(K,L)} + \big[\bar{Z}^{(K,L)}\big]' \Gamma_c^{(K,L)} \bar{Z}^{(K,L)},$$

with a constant $C^{(K,L)}$, a symmetric matrix of coefficients $\Gamma_c^{(K,L)}$, and a normal vector $\bar{Z}^{(K,L)} = (Z_1^{(K,L)}, \dots, Z_{M(K,L)}^{(K,L)})'$. Hence, $Z_c^{(K,L)} - C^{(K,L)}$ can be rewritten as follows:

$$\begin{aligned} Z_c^{(K,L)} - C^{(K,L)} &\stackrel{d}{=} \bar{Y}'\big[U_c^{(K,L)}\big]' \Lambda_c^{(K,L)} U_c^{(K,L)} \bar{Y} = Y' \Lambda_c^{(K,L)} Y \\ &= \sum_{k=1}^{M(K,L)} \lambda_k^{(c,K,L)} Y_k^2. \end{aligned}$$

Here $U_c^{(K,L)}$ is a certain orthogonal matrix, $\Lambda_c^{(K,L)} := \operatorname{diag}(\lambda_1^{(c,K,L)}, \dots, \lambda_{M(K,L)}^{(c,K,L)})$ with $|\lambda_1^{(c,K,L)}| \geq \cdots \geq |\lambda_{M(K,L)}^{(c,K,L)}|$, and $\bar{Y}$ as well as Y are multivariate standard normally distributed random vectors. For notational simplicity, we suppress the upper index (c, K, L) in the sequel. Due to the above choice of the triple (c, K, L), either $\sum_{k=1}^{4} (\lambda_k)^2$ or $\sum_{k=5}^{M(K,L)} (\lambda_k)^2$ is bounded from below by $V/4$. In the first case, $\lambda_1 \geq \sqrt{V/16}$ holds true which implies

$$P\big(Z_c^{(K,L)} \in [x - \varepsilon, x + \varepsilon]\big) \leq \int_0^{2\varepsilon} f_{\lambda_1 Y_1^2}(t)\, \mathrm{d}t \leq P(Y_1^2 \leq 2\varepsilon) \max\left\{1, \frac{4}{\sqrt{V}}\right\} \qquad \forall x \in \mathbb{R}.$$

Here, the first inequality results from the fact that convolution preserves the continuity properties of the smoother function. In the opposite case, that is, $\sum_{k=5}^{M(K,L)} (\lambda_k)^2 \geq V/4$, it is possible to bound the uniform norm of the density function of $Z_c^{(K,L)}$ by means of its variance. To this end, we first consider the characteristic function $\varphi_{Z_c^{(K,L)}}$ of $Z_c^{(K,L)}$ and assume w.l.o.g. that $M(K, L)$ is divisible by 4. Defining a sequence $(\mu_k)_{k=1}^{M(K,L)/4}$ by $\mu_k = \lambda_{4k}$ for $k \in \{1, \dots, M(K, L)/4\}$

allows for the approximation:

$$\left|\varphi_{Z_c^{(K,L)}}(t)\right| = \left\{\prod_{j=1}^{M(K,L)} (1+[2\lambda_j t]^2)\right\}^{-1/4} \leq \left\{\prod_{j=1}^{M(K,L)/4} (1+[2\mu_j t]^2)\right\}^{-1}$$
$$\leq \frac{1}{1+4(\mu_1^2+\cdots+\mu_{M(K,L)/4}^2)t^2}.$$

By inverse Fourier transform, we obtain the following result concerning the density function of $Z_c^{(K,L)}$:

$$\|f_{Z_c^{(K,L)}}\|_\infty \leq \frac{1}{2\pi}\|\varphi_{Z_c^{(K,L)}}\|_1 \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{1+(2\sqrt{\mu_1^2+\cdots+\mu_{M(K,L)/4}^2}t)^2}\,\mathrm{d}t$$
$$= \frac{1}{\sqrt{\mu_1^2+\cdots+\mu_{M(K,L)/4}^2}}\frac{1}{2\pi}\int_0^\infty \frac{1}{1+u^2}\,\mathrm{d}u$$
$$\leq \frac{1}{2\sqrt{4(\mu_1^2+\cdots+\mu_{M(K,L)/4-1}^2)}}$$
$$\leq \frac{1}{2\sqrt{\lambda_5^2+\cdots+\lambda_{M(K,L)}^2}} \leq \frac{1}{\sqrt{V}}.$$

Thus, $P(Z_c^{(K,L)} \in [x-\varepsilon, x+\varepsilon]) \leq 2\varepsilon/\sqrt{V}$ which completes the studies of the case $\sum_{k=5}^{M(K,L)}(\lambda_k)^2 > V/4$ and finally yields the assertion. □

Proof of Lemma 3.1. This result is an immediate consequence of Theorem 3.1. □

Acknowledgements

The author is grateful to Michael H. Neumann for his constructive advice and fruitful discussions. She also thanks an anonymous referee for helpful comments that led to an improvement of the paper. This research was funded by the German Research Foundation DFG (project: NE 606/2-1).

References

[1] Andrews, D.W.K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. MR0766830

[2] Arcones, M.A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586

[3] Babbel, B. (1989). Invariance principles for U-statistics and von Mises functionals. *J. Statist. Plann. Inference* **22** 337–354. MR1006168

[4] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[5] Carlstein, E. (1988). Degenerate U-statistics based on nonindependent observations. *Calcutta Statist. Assoc. Bull.* **37** 55–65. MR0964306
[6] Daubechies, I. (1992). *Ten Lectures on Wavelets*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). MR1162107
[7] de Wet, T. (1987). Degenerate U- and V-statistics. *South African Statist. J.* **21** 99–129. MR0936366
[8] Dedecker, J. and Prieur, C. (2004). Couplage pour la distance minimale. *C. R. Math. Acad. Sci. Paris* **338** 805–808. MR2059491
[9] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132** 203–236. MR2199291
[10] Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.* **51** 392–413. MR1321305
[11] Denker, M. (1982). Statistical decision procedures and ergodic theory. In *Ergodic Theory and Related Topics* (*Vitte*, 1981) (H. Michael, ed.). *Math. Res.* **12** 35–47. Berlin: Akademie-Verlag. MR0730767
[12] Dewan, I. and Prakasa Rao, B.L.S. (2001). Asymptotic normality for U-statistics of associated random variables. *J. Statist. Plann. Inference* **97** 201–225. MR1861150
[13] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737
[14] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342. MR1719345
[15] Eagleson, G.K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.* **21** 221–237. MR0561949
[16] Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Statist.* **10** 245–271. MR1708583
[17] Fan, Y. and Li, Q. (2000). Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* **16** 1016–1041. MR1803718
[18] Feuerverger, A. and Mureika, R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5** 88–97. MR0428584
[19] Franke, J. and Wendel, M. (1992). A bootstrap approach for nonlinear autoregressions – some preliminary results. In *Bootstrapping and Related Techniques* (K.-H. Jöckel, G. Rothe, and W. Sendler, eds.). *Lecture Notes in Economics and Mathematical Systems* **376** 101–106. Berlin: Springer.
[20] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. *Lecture Notes in Statistics* **129**. New York: Springer. MR1618204
[21] Huang, W. and Zhang, L.X. (2006). Asymptotic normality for U-statistics of negatively associated random variables. *Statist. Probab. Lett.* **76** 1125–1131. MR2269283
[22] Knopp, K. (1956). *Infinite Sequences and Series*. New York: Dover Publications Inc. MR0079110
[23] Leucht, A. (2010). Characteristic function-based goodness-of-fit tests under weak dependence. Report of the Dept. Mathematics and Computer Science 10-06, Friedrich Schiller Univ. Jena.
[24] Leucht, A. (2010). Consistent model-specification tests based on parametric bootstrap. Report of the Dept. Mathematics and Computer Science 10-07, Friedrich Schiller Univ. Jena.
[25] Leucht, A. and Neumann, M.H. (2009). Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *J. Multivariate Anal.* **100** 1622–1633. MR2535374
[26] Neumann, M.H. and Paparoditis, E. (2008). Goodness-of-fit tests for Markovian time series models: Central limit theory and bootstrap approximations. *Bernoulli* **14** 14–46. MR2401652
[27] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165

[28] Staudte, R.G. and Sheather, S.J. (1990). *Robust Estimation and Testing*. New York: Wiley. MR1061581

[29] Wojtaszczyk, P. (1997). *A Mathematical Introduction to Wavelets*. *London Mathematical Society Student Texts* **37**. Cambridge: Cambridge Univ. Press. MR1436437

Received February 2010 and revised November 2010