Dependent wild bootstrap for degenerate U- and V-statistics

Anne Leucht
Technische Universität Braunschweig
Institut für Mathematische Stochastik
Pockelsstr. 14
D-38106 Braunschweig
Germany
E-mail: a.leucht@tu-braunschweig.de

Michael H. Neumann
Friedrich-Schiller-Universität Jena
Institut für Stochastik
Ernst-Abbe-Platz 2
D-07743 Jena
Germany
E-mail: michael.neumann@uni-jena.de

Abstract

Degenerate U- and V-statistics play an important role in the field of hypothesis testing since numerous test statistics can be formulated in terms of these quantities. Therefore, consistent bootstrap methods for U- and V-statistics can be applied in order to approximate critical values of those tests. We prove a new asymptotic result for degenerate U- and V-statistics of weakly dependent random variables. As our main contribution, we propose a new model-free bootstrap method for U- and V-statistics of dependent random variables. Our method is a modification of the dependent wild bootstrap recently proposed by Shao (2010, *JASA* **105**, 218–235), where we do not directly bootstrap the underlying random variables but the summands of the U- and V-statistics. Asymptotic theory for the original and the bootstrap statistics is derived under simple and easily verifiable conditions. We discuss applications to a Cramér-von Mises-type test and a two sample test for the marginal distribution of a time series in detail. The finite sample behavior of the Cramér-von Mises test is explored in a small simulation study. While the empirical size is reasonably close to the nominal one, we see nontrivial empirical power in all cases considered.

2010 Mathematics Subject Classification. Primary 62M07, 62G09; secondary 62E20.
Keywords and Phrases. Bootstrap, weak dependence, U-statistic, V-statistic, Cramér-von Mises test, two-sample test.
Short title. Dependent wild bootstrap for U-statistics.

version: December 5, 2011

1. Introduction

U- and related von Mises- (V-)statistics play an important role in mathematical statistics. In the case of hypothesis testing, major interest is on degenerate statistics of this type since they approximate important test quantities under the null hypothesis. Well-known examples are the Cramér-von Mises and the χ^2-statistics. Especially in the case of dependent random variables, the distribution of such a statistic has quite an involved form and depends on characteristics of the underlying process in a complicated manner. For the determination of critical values, we propose new versions of a model-free bootstrap method that can be viewed as variants of the dependent wild bootstrap recently proposed by Shao (2010) for smooth functions of the mean.

In Section 2 we derive the limit distributions of degenerate U- and V-statistics under a condition of weak dependence introduced by Dedecker and Prieur (2004). The classical approach to derive the limit distributions of such statistics is based on a spectral decomposition of the kernel function and was first taken by Gregory (1977) in the case of i.i.d. random variables. Later it has been used for mixing random variables by Eagleson (1979), Carlstein (1988), and Borisov and Volodko (2008) as well as for associated random variables by Dewan and Prakasa Rao (2001) and Huang and Zhang (2006). This method works actually perfectly well in the case of independent random variables, however, it has to be taken with care in the dependent case. The authors mentioned above imposed additional conditions on the corresponding eigenvalues and eigenfunctions that can hardly be checked in practice. Making use of the observation that typical test statistics of L_2-type can be approximated by V-statistics with positive semidefinite kernel functions, Leucht and Neumann (2011) could simplify technical issues and derived the asymptotics for U- and V-statistics for ergodic processes. They imposed a stricter form of degeneracy that is satisfied by V-statistics resulting from model-specification tests for the conditional mean function or from goodness-of-fit tests for the conditional distribution of time series data. However, it is violated by the classical Cramér-von Mises test statistic, see also their Remark 2. Here we derive the limit distribution under the usual degeneracy and under easily verifiable conditions imposed directly on the kernel function.

For bootstrapping U- or V-statistics of dependent random variables, there are so far only consistency results tailor-made for model-based bootstrap methods; see Leucht (2011) and Leucht and Neumann (2011). However, goodness-of-fit tests based on the empirical distribution are most appropriate if no particular model class is available. To the best of our knowledge, the literature does not provide consistency results on model-free bootstrap methods for degenerate U- and V-statistics under dependence. In Section 3 we introduce new variants of the dependent wild bootstrap that are suitable for degenerate U- and V-statistics. For dependent random variables $X_1, \ldots, X_n$ satisfying certain conditions, and a smooth function H, Shao (2010) proposed to approximate the distribution of $H(\bar{X}_n) - H(EX_1)$ by that of $H(n^{-1}\sum_{t=1}^n X_t^*) - H(\bar{X}_n)$, where $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$, $X_t^* = \bar{X}_n + (X_t - \bar{X}_n)W_{t,n}^*$ $(t = 1, \ldots, n)$ and $(W_{t,n}^*)_{t=1}^n$ are weakly dependent random variables independent of $X_1, \ldots, X_n$. Because of the obvious similarity to Wu's (1986) wild bootstrap for independent random variables, Shao called this method the dependent wild bootstrap. The role played by the triangular scheme $(W_{t,n}^*)_{t=1}^n$, $n \in \mathbb{N}$, is easily explained: While a nondegenerate distribution of $W_{t,n}^*$ with $E^*W_{t,n}^* = 0$ introduces the necessary randomness, the condition of $\mathrm{cov}^*(W_{s,n}^*, W_{t,n}^*) \longrightarrow_{n\to\infty} 1$ takes care that the dependence structure of the original process $X_1, \ldots, X_n$ is asymptotically captured. Knowledge of the mechanism producing the limit distribution of a V-statistic $V_n = n^{-1}\sum_{s,t=1}^n h(X_s, X_t)$ helps us to devise a variant of the dependent wild bootstrap appropriate for U- and V-statistics. Under the conditions imposed below, the V-statistic can be rewritten as $V_n = \sum_k \lambda_k \left(n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)\right)^2$, where $(\lambda_k)_k$ are the nonzero eigenvalues and $(\Phi_k)_k$ the corresponding eigenfunctions of a certain eigenvalue problem. The random variables $(\Phi_k(X_t))_{t=1}^n$ have zero mean and inherit the property of weak dependence from the original process. The limit distribution of V_n comes from joint asymptotic normality of $n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)$, $k \in \mathbb{N}$. Therefore, it is quite a natural attempt to approximate the distribution of V_n by that of $V_n^* = n^{-1}\sum_{s,t=1}^n h(X_s, X_t)W_{s,n}^* W_{t,n}^*$. We prove that this approximation is actually consistent under easily verifiable conditions. In particular, we require only a very weak condition on the tuning parameter of this method that is obviously a necessary one. An analogous result holds also for U-statistics.

In Section 4 we apply our general results to two particular test problems. We consider a goodness-of-fit test of Cramér-von Mises type and a test of equality of the marginal distribution of two samples. Even in the former case where we restrict our attention to tests of simple hypotheses, we cannot simply use simulations to determine an appropriate critical value since the dependence structure is left unspecified. We can also not use tabulated critical values from the independent case since the effect of dependence is not negligible here. It was already pointed out by Gleser and Moore (1983) that the null is rejected to often if the quantiles of the i.i.d. setting are used as an approximation for the corresponding quantities under positive dependence. Both test statistics can be rewritten as V-statistics that are degenerate under the null hypothesis and it can be seen that the conditions imposed in the Sections 2 and 3 are fulfilled. Hence, our version of the dependent wild bootstrap method provides asymptotically correct critical values.

Section 5 contains a numerical analysis of the proposed bootstrap method for the classical Cramér-von Mises test in the time series context. It can be seen from Figure 1 that the bootstrap distributions of the test statistic are much more accurate approximations than the distribution from the independent case. Moreover, it turns out that the actual size of the test is not too far from the prescribed one. Some impression on the power is also given by a few examples.

The proofs of the theorems and some auxiliary results are deferred to a final Section 6. Besides many approximations, a key tool for the asymptotic theory is a multivariate central limit theorem (CLT) for weakly dependent random variables. In case of the original process, a univariate CLT in conjunction with the Cramér-Wold device would obviously do the job. However, on the bootstrap side we have to deal with a triangular scheme and all conditions required for a CLT are only fulfilled in probability. This fact makes a direct application of the Cramér-Wold device much more cumbersome. Therefore, we establish as a by-product a multivariate generalisation of the CLT of Neumann (2011) that implies then a multivariate bootstrap CLT as a direct consequence. We think that these technical tools are of interest beyond this work.

2. Asymptotic distributions of U- and V-statistics

Suppose that observations $X_1, \ldots, X_n$ from a strictly stationary process are available. In view of Kolmogorov's consistency theorem there exists a two-sided process with the corresponding finite-dimensional distributions. To simplify the presentation, we will assume in the sequel that our observations stem from such a two-sided process $(X_t)_{t\in\mathbb{Z}}$.

In this section, we derive the limit distributions of

$$U_n = \frac{1}{n} \sum_{\substack{s,t=1 \\ s\neq t}}^{n} h(X_s, X_t) \qquad \text{and} \qquad V_n = \frac{1}{n} \sum_{s,t=1}^{n} h(X_s, X_t),$$

where $h\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a symmetric function that is degenerate, i.e. $Eh(X_0, y) = 0 \ \forall\, y \in \operatorname{supp}(P^{X_0})$. To impose a restriction on the dependence of the underlying process $(X_t)_{t\in\mathbb{Z}}$, we invoke the concept of τ-dependence introduced by Dedecker and Prieur (2005).

Definition 2.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary sequence of integrable $\mathbb{R}^d$-valued random variables. The process is called τ-dependent if

$$\tau(r) \;=\; \sup_{l\in\mathbb{N}} \frac{1}{l} \sup_{r\leq k_1<\cdots<k_l} \{\tau\left(\sigma(X_i, i\leq 0), (X_{k_1}, \ldots, X_{k_l})\right)\} \underset{r\to\infty}{\longrightarrow} 0,$$

where

$$\tau(\mathcal{M}, X) = E\left(\sup_{f\in\Lambda_1(\mathbb{R}^p)} \left| \int_{\mathbb{R}^p} f(x) dP_{X|\mathcal{M}}(x) - \int_{\mathbb{R}^p} f(x) dP_X(x) \right| \right).$$

Here, $\mathcal{M}$ is a sub-σ-algebra of $\mathcal{A}$, $P_{X|\mathcal{M}}$ denotes the conditional distribution of the $\mathbb{R}^p$-valued random variable X given $\mathcal{M}$, and $\Lambda_1(\mathbb{R}^p)$ denotes the set of 1-Lipschitz functions from $\mathbb{R}^p$ to $\mathbb{R}$, i.e. $f \in \Lambda_1(\mathbb{R}^p)$ if $|f(x) - f(y)| \leq \|x - y\|_1 = \sum_{j=1}^{p} |x_j - y_j| \ \forall\, x, y \in \mathbb{R}^p$.

If Ω is rich enough, there exists a random variable $\widetilde{X}$ that is independent of $\mathcal{M} = \sigma(X_i, i \leq 0)$ such that $\widetilde{X} \overset{d}{=} X$ and

$$\tau(\mathcal{M}, X) = E\big\|X - \widetilde{X}\big\|_1. \tag{2.1}$$

Actually, $\tau(\mathcal{M}, X)$ is the minimal L_1-distance between X and any $Y \stackrel{d}{=} X$ that is independent of $\mathcal{M}$, cf. Dedecker and Prieur (2004). This L_1-coupling property will be an essential device for all our proofs below. Dedecker and Prieur (2005) discussed relations of the τ-coefficient to ordinary mixing coefficients. Additionally, they provided an extensive list of examples for τ-dependent processes including causal linear and functional autoregressive processes. Moreover, ARMA processes with weakly dependent innovations and GARCH processes are τ-dependent; see Shao and Wu (2007).

We assume

(A1) $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary, τ-dependent process with $\sum_{r=1}^{\infty} \sqrt{\tau(r)} < \infty$.

Moreover, we impose the following conditions regarding the kernel h:

(A2) (i) $h\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is symmetric and degenerate, i.e. $h(x,y) = h(y,x)\ \forall\, x, y \in \mathbb{R}^d$ and $Eh(X_0, y) = 0$ for all $y \in \operatorname{supp}(P^{X_0}) := \{x \in \mathbb{R}^d \mid \forall \text{open } O\colon x \in O \Rightarrow P^{X_0}(O) > 0\}$.

(ii) h is a positive semidefinite function, i.e., for all $c_1, \ldots, c_m \in \mathbb{R}$, $x_1, \ldots, x_m \in \mathbb{R}^d$ and $m \in \mathbb{N}$, $\sum_{i,j=1}^m c_i\, c_j\, h(x_i, x_j) \;\geq\; 0$.

(iii) $Eh(X_0, X_0) < \infty$.

(iv) h is Lipschitz continuous, i.e. $\operatorname{Lip}(h) := \sup_{x,\bar{x},y\in\mathbb{R}^d,\, x\neq\bar{x}} |h(x,y) - h(\bar{x},y)|/\|x-\bar{x}\|_1 < \infty$.

Theorem 2.1. *Suppose that (A1) and (A2) hold. Then, as* $n \longrightarrow \infty$,

$$V_n \stackrel{d}{\longrightarrow} Z := \sum_k \lambda_k\, Z_k^2 \qquad \text{and} \qquad U_n \stackrel{d}{\longrightarrow} Z - Eh(X_0, X_0),$$

where $(Z_k)_k$ *is a sequence of centered, jointly normal random variables with* $\operatorname{cov}(Z_j, Z_k) = \sum_{r=-\infty}^{\infty} \operatorname{cov}(\Phi_j(X_0), \Phi_k(X_r))$, *and* $(\lambda_k)_k$ *and* $(\Phi_k)_k$ *are the sequences of non-zero eigenvalues and the corresponding orthonormal eigenfunctions of* $E[h(x, X_0)\, \Phi(X_0)] \;=\; \lambda\, \Phi(x)$.

There are already several papers on the asymptotic distributions of U- and V-statistics of dependent random variables. The classical approach of deriving and representing the limit by means of a spectral decomposition of the kernel function was adapted for mixing random variables by Eagleson (1979), Carlstein (1988), and Borisov and Volodko (2008) and for associated random variables by Dewan and Prakasa Rao (2001) and Huang and Zhang (2006). One essential drawback of these results is that regularity conditions are required to hold uniformly for all eigenfunctions of the eigenvalue problem $E[h(x, X_0)\Phi(X_0)] = \lambda\Phi(x)$. However, for the majority of test statistics the eigenfunctions of the kernel of the (approximating) V-statistic are unknown and, even worse, these conditions are not satisfied in general. For example, the smoothness assumptions of Dewan and Prakasa Rao (2001) and Huang and Zhang (2006) are not even satisfied in the case of the classical Cramér-von Mises-statistic. To manage with smoothness conditions directly for the kernel function, Babbel (1989) and Leucht (2011) used wavelet expansions of the kernel and obtained a representation of the weak limit associated with the chosen wavelet basis. Compared to the present paper, Babbel (1989) assumed a stricter form of degeneracy to hold while Leucht (2011) imposed stronger assumptions concerning the moments of h and assumed a faster decay of the dependence coefficients than we do. A spectral decomposition of the kernel was also employed by Leucht and Neumann (2011) for degenerate U- and V-statistics of random variables from ergodic processes. While their additional condition of a positive semidefinite kernel is typically satisfied in applications to tests of L_2-type, they also imposed the condition $E(h(x, X_t) \mid X_{t-1}, \ldots, X_1) = 0$ that is stricter than ordinary degeneracy and not satisfied in our applications in Section 4. Having these particular applications in mind, we have relaxed this assumption and assume ordinary degeneracy instead.

So far, we assumed the kernel h to be Lipschitz continuous, which is too restrictive for some applications. It turns out that it suffices to postulate local Lipschitz continuity under some slightly stricter assumption concerning the decay of the dependence coefficients.

(A3) With some function $f\colon \mathbb{R}^{4d} \to \mathbb{R}$, that is bounded on any compact interval,

$$|h(x,y) - h(\bar{x},\bar{y})| \leq f(x,\bar{x},y,\bar{y})\ \{\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1\}\,,$$

where

$$\sup_{Y_1,\ldots,Y_5 \sim P^{X_0}} E\left([f(Y_1,Y_2,Y_3,Y_4)]^{1/(1-\delta)}\ \|Y_5\|_1\right) < \infty$$

for some $\delta \in (0,1)$ with $\sum_{r=1}^{\infty} [\tau(r)]^{\delta/2} < \infty$.

Corollary 2.1. *Under the assumptions (A1), (A2)(i) - (iii), and (A3) the assertion of Theorem 2.1 holds true.*

3. Dependent wild bootstrap for U- and V-statistics

For a test statistic that can be approximated by a V-statistic, Theorem 2.1 and Corollary 2.1 identify the asymptotic distribution under the null as the sample size tends to infinity. This limit depends on the eigenvalues of the problem $E[h(x,X_0)\Phi(X_0)] = \lambda\Phi(x)$ that are typically unknown in applications and, even more seriously, depend on the whole dependence structure of the underlying process via sums of covariances. Some knowledge of the limit distribution is necessary when critical values have to be determined. Since a direct approximation of these quantities seems to be very cumbersome, we suggest to apply an appropriate bootstrap method here. Tests as those considered in Section 4 below are most convenient if no parametric model for the time series is available. While asymptotic correctness of certain model-based bootstrap methods was proved in Leucht (2011) and Leucht and Neumann (2011), we will devise a model-free bootstrap. We propose new variants of the dependent wild bootstrap, which was introduced by Shao (2010) for smooth functions of the mean. Originally, the idea of the dependent wild bootstrap is to construct the pseudo-observations as follows:

$$X_t^* = \bar{X}_n + \big(X_t - \bar{X}_n\big)\, W_t^*, \quad t = 1, \dots, n.$$

Here, $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ and $(W_t^*)_{t=1}^n = (W_{t,n}^*)_{t=1}^n$ is a triangular scheme of weakly dependent random variables that is independent of $X_1, \dots, X_n$. Shao (2010) verified that under certain regularity conditions

$$\sup_{x\in\mathbb{R}} \left| P\left(\sqrt{n}\left[H(\bar{X}_n) - H(EX_1)\right] \le x\right) - P^*\left(\sqrt{n}\left[H\left(\frac{1}{n}\sum_{t=1}^n X_t^*\right) - H(\bar{X}_n)\right] \le x\right)\right| \xrightarrow{P} 0,$$

where H is a smooth function. We adapt his approach in order to construct a wild bootstrap procedure for degenerate U- and V-statistics of dependent data. However, we do not generate a bootstrap counterpart of the observations themselves, but of $h(X_s, X_t),\ s, t = 1, \dots, n$. More precisely, we consider

$$U_{n,1}^* \;=\; \frac{1}{n} \sum_{\substack{s,t=1,\\ s\neq t}}^n W_s^*\, h(X_s, X_t)\, W_t^* \qquad \text{and} \qquad V_{n,1}^* \;=\; \frac{1}{n} \sum_{s,t=1}^n W_s^*\, h(X_s, X_t)\, W_t^*.$$

We will show that these statistics consistently mimic the behavior of U_n and V_n, respectively. Tests considered in Section 4 with a critical value chosen by this bootstrap method have asymptotically the correct size. Moreover, under a fixed alternative, the test statistic divided by n will converge to some positive constant while its bootstrap version is of oder $o_P(n)$. This implies consistency of the test. However, we have learned from our simulations reported in Section 5 that the power of such a test might be low in some cases, at least for moderate sample sizes. For good power properties, it would be ideal if the bootstrap statistic mimics under the alternative some null scenario, i.e., degeneracy of the corresponding kernel is important. To achieve this, we need that W_t^* has zero mean. On the other hand, this effect is mitigated by the fact that condition (B2) implies $\operatorname{cov}^*(W_s^*, W_t^*) \longrightarrow_{n\to\infty} 1$. Therefore, we propose the following modified versions that are improved by an empirical degeneration:

$$U_{n,2}^* \;=\; \frac{1}{n} \sum_{\substack{s,t=1,\\ s\neq t}}^n W_s^*\, \bar{h}(X_s, X_t)\, W_t^* \qquad \text{and} \qquad V_{n,2}^* \;=\; \frac{1}{n} \sum_{s,t=1}^n W_s^*\, \bar{h}(X_s, X_t)\, W_t^*,$$

where $\bar{h}(x, y) = h(x, y) - n^{-1} \sum_{k=1}^n h(X_k, y) - n^{-1} \sum_{k=1}^n h(x, X_k) + n^{-2} \sum_{k,l=1}^n h(X_k, X_l)$. For the purpose of an efficient implementation, we note that the V-statistic can be rewritten as

$$V_{n,2}^* \;=\; \frac{1}{n} \sum_{s,t=1}^n h(X_s, X_t)(W_s^* - \bar{W}_n^*)(W_t^* - \bar{W}_n^*),$$

where $\bar{W}_n^* = n^{-1} \sum_{t=1}^n W_t^*$.

In order to verify bootstrap validity, we impose a slightly stricter condition on the dependence structure of the underlying process than (A1) in the previous section.

(B1) $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary, τ-dependent process with $\sum_{r=1}^{\infty} r^2\sqrt{\tau_r}<\infty$.

Regarding the variables $(W_t^*)_t$ we make the following assumptions:

(B2) The row-wise strictly stationary triangular array $(W_t^*)_{t=1}^n = (W_{t,n}^*)_{t=1}^n$ is independent of $X_1,\ldots,X_n$. Moreover, $E^*W_1^* = 0$, $\sup_n E^*|W_{1,n}^*|^{2+\delta} < \infty$ for some $\delta > 0$, and $\operatorname{cov}(W_s^*,W_t^*) = \rho(|s-t|/l_n)$, where $\rho(u) \longrightarrow_{u\to 0} 1$, $\sum_{r=1}^{n-1}\rho(|r|/l_n) = O(l_n)$ with $l_n \longrightarrow_{n\to\infty} \infty$ and $l_n = o(n)$. The variables $(W_{t,n}^*)_{t=1}^n$ are τ-weakly dependent with coefficients $\tau^*(r) \le K\,\zeta^{r/l_n}$ for $r=1,\ldots,n$, some $\zeta\in(0,1)$ and a $K<\infty$.

Remark 1. A simple way to construct the process $(W_{t,n}^*)_{t=1}^n$ is to take a Gaussian process $(U_t)_{t\in[0,\infty)}$ with zero mean, unit variance, continuous sample paths and satisfying appropriate mixing properties, and to define $W_{t,n}^* := U_{t/l_n},\ t=1,\ldots,n$. For example, an Ornstein-Uhlenbeck process would be appropriate. Then the practical implementation becomes easy since a discrete sample of an Ornstein-Uhlenbeck process forms an AR(1) process.

Theorem 3.1. *Under the assumptions (A2), (B1), and (B2), for* $i=1,2$,

$$V_{n,i}^* \xrightarrow{d} Z \qquad \text{and} \qquad U_{n,i}^* \xrightarrow{d} Z - Eh(X_0,X_0) \qquad \text{in probability.}$$

If additionally the limiting distribution function is continuous, then

$$\sup_{x\in\mathbb{R}}|P^*(U_{n,i}^*\le x) - P(U_n\le x)| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{x\in\mathbb{R}}|P^*(V_{n,i}^*\le x) - P(V_n\le x)| \xrightarrow{P} 0.$$

Remark 2. In analogy to Corollary 2.1, we can weaken the assumption of h to be Lipschitz continuous to local Lipschitz continuity in the sense of (A3) if we assume a faster decay of the dependence coefficients, i.e. $\sum_{r=1}^{\infty} r^2\,[\tau(r)]^{\delta/2}<\infty$ for some $\delta\in(0,1)$.

A necessary and sufficient condition for the continuity of the limiting distribution function is stated in the following lemma.

Lemma 3.1. *Suppose that (A1) and (A2) hold true. Z has a continuous distribution function if and only if* $EZ=\sum_{r\in\mathbb{Z}} Eh(X_0,X_r)\ne 0$.

The latter condition in this lemma simplifies to $Eh(X_0,X_0)>0$ in all the applications considered by Leucht and Neumann (2011).

Remark 3. We conjecture that consistency of the stationary block bootstrap, proposed by Politis and Romano (1994), can be proved using similar tools as in our proof of Theorem 3.1. This method can be interpreted as an extension of Efron's bootstrap to dependent data. It has been shown by Arcones and Giné (1992) and Dehling and Mikosch (1994) that the application of Efron's bootstrap to degenerate U- and V-statistics necessarily requires an artificial degeneration of the statistics on the bootstrap side. Some preliminary calculations show that this problem carries over to block bootstrap methods. In this respect, the first variant of our bootstrap proposal is at least computationally less intensive.

4. Applications

4.1. **A generalized Cramér-von Mises test for dependent data.**

Let $X_1, \ldots, X_n$ be $\mathbb{R}^d$-valued observations from a strictly stationary process with unknown marginal distribution function F. We are interested in testing

$$\mathcal{H}_0\colon \quad F = F_0 \qquad \text{vs.} \qquad \mathcal{H}_1\colon \quad F \neq F_0.$$

Below we consider a test of generalized Cramér-von Mises type, i.e. based on the test statistic

$$T_n = n \int_{\mathbb{R}^d} [F_n(z) - F_0(z)]^2 \, w(z) \, \lambda^d(dz),$$

where F_n denotes the empirical distribution function based on $X_1, \ldots, X_n$ and w is a weight function.

While there is a great variety of tests when the underlying observations are i.i.d., the number of consistent tests is limited in the dependent case. Ignaccolo (2004) derived asymptotic theory for Pearson's χ^2-test and Neyman's smooth test with a fixed number of components for α-mixing variables. These tests are also covered by our more general approach, however, they are only consistent for alternatives that are not orthogonal to the finitely many basis functions involved in the test statistic. Munk *et al.* (2011) proposed a modified version of Neyman's smooth test where the number of components was chosen by Schwarz's selection rule and they showed its consistency for essentially all alternatives. It is clear that a test based on T_n is also consistent against all fixed alternatives. Moreover, distinct choices of the weight function w allow to direct the power towards different alternatives; see e.g. Anderson and Darling (1954). For β-mixing data Fan and Ullah (1999) as well as Neumann and Paparoditis (2000) considered tests based on the L_2-difference between a smoothed version of the parametric density estimate and a nonparametric estimator. The relative merits of smoothing-based tests based on local characteristics like densities versus non-smoothing tests based on cumulative features are discussed by Rosenblatt (1975) and Ghosh and Huang (1991). While the first-mentioned tests are more suitable to detect local alternatives characterized by densities with sharp peaks, they suffer from a loss of power against so-called Pitman alternatives. Here, we obtain a consistent testing procedure of the second type based on the theory of the foregoing part of the paper. To this end, first note that our test statistic can be reformulated as

$$T_n = \frac{1}{n} \sum_{s,t=1}^{n} h(X_s, X_t),$$

where $h(x,y) = \int [\mathbb{1}_{x \preceq z} - F_0(z)][\mathbb{1}_{y \preceq z} - F_0(z)] w(z) \lambda^d(dz)$ ($x \preceq y$ means that $x_i \leq y_i \ \forall i$). The weight function w is assumed to be non-negative, integrable and, for sake of simplicity, bounded. This yields Lipschitz continuity of the kernel h. Moreover, it is obvious that the V-statistic T_n is degenerate under the null. As can be observed from Figure 1 in Section 5 below, we cannot use critical values that are tabulated for the independent case. Moreover, although we deal with a simple hypothesis, we can also not obtain a critical value via simulations since the dependence structure of the underlying process is unspecified. Therefore, we will apply our version of the dependent wild bootstrap which can be implemented as follows.

Algorithm:

(1) Generate $W_1^*, \ldots, W_n^*$ such that condition (B2) is satisfied.

(2) Compute the bootstrap counterpart of the test statistic,

$$T_{n,1}^* = \frac{1}{n} \sum_{s,t=1}^{n} h(X_s, X_t)\, W_s^*\, W_t^* \qquad \text{resp.} \qquad T_{n,2}^* = \frac{1}{n} \sum_{s,t=1}^{n} h(X_s, X_t)\, (W_s^* - \bar{W}_n^*)\, (W_t^* - \bar{W}_n^*).$$

(3) Repeat steps (1) and (2) B times and determine the $(1-\alpha)$-quantile $t_{\alpha,i}^*$ of the empirical distribution of $T_{n,i}^*$.

(4) Reject $\mathcal{H}_0$ if $T_n > t_{\alpha,i}^*$.

In view of Theorem 3.1, $P(T_n > t_{\alpha,i}^*) \longrightarrow \alpha$ under $\mathcal{H}_0$ as the sample size increases if (B1) is satisfied and if $\sum_{r=-\infty}^{\infty} Eh(X_0, X_r) \neq 0$. The next proposition characterizes consistency of the test under the alternative.

Proposition 4.1. (i) *If (A1) is fulfilled, then*

$$n^{-1}\, T_n \xrightarrow{P} E[h(X_0, \widetilde{X}_0)] = \int_{\mathbb{R}^d} (F(z) - F_0(z))^2 w(z)\, \lambda^d(dz).$$

(Here $\widetilde{X}_0$ denotes an independent copy of X_0.)

(ii) *If (B1) and (B2) are fulfilled, then*

$$n^{-1} T_{n,1}^* \xrightarrow{P^*} 0 \qquad \text{in probability}$$

and

$$P\left(P^*\left(T_{n,2}^* > K(\epsilon)\right) < \epsilon\right) \underset{n\to\infty}{\longrightarrow} 1$$

for all $\epsilon > 0$ and appropriate $K(\epsilon) < \infty$.

(iii) *If (B1) and (B2) are fulfilled and if w is positive almost everywhere w.r.t. Lebesgue measure, then*

$$P\left(T_n > t^*_{\alpha,i}\right) \underset{n\to\infty}{\longrightarrow} 1 \qquad (i = 1, 2).$$

Remark 4. A model-based bootstrap-aided L_2-type test based on the empirical characteristic function was proposed by Leucht (2010). However, model-free bootstrap procedures are desirable in the present context since time series models characterize conditional rather than the marginal distribution. The validity of the dependent wild bootstrap for characteristic function based L_2-tests can be verified similarly to Proposition 4.1. Moreover, counterparts of the tests by Fan and Ullah (1999) and by Neumann and Paparoditis (2000) based on fixed kernel estimates can also be constructed along these lines.

4.2. **A two-sample test for time series.** Suppose that we observe $\mathbb{R}^d$-valued random variables $X_1, \ldots, X_n$ with cumulative distribution function F_X and $Y_1, \ldots, Y_n$ with cumulative distribution function F_Y from a process $((X_t', Y_t')')_t$ that satisfies (B1). Here the case of two independent processes that are both weakly dependent is contained as an important special case. We are concerned with the question whether the distributions of the two samples coincide, i.e.

$$\widetilde{\mathcal{H}}_0\colon \quad F_X = F_Y \qquad \text{vs.} \qquad \widetilde{\mathcal{H}}_1\colon \quad F_X \neq F_Y.$$

This problem has been intensively studied when the underlying variables are independent. Exemplarily, we mention Anderson *et al.* (1994), whose test is based on fixed kernel density estimates, Li (1996), whose test relies on kernel density estimates with vanishing bandwidths, and the characteristic function-based approach of Alba-Fernández, Jiménez-Gamero, and Muñoz-García (2008). For the dependent case, we are only aware of the work of Fan and Ullah (1999) based on an L_2-type comparison of nonparametric density estimators with vanishing bandwidth. Here, we propose a test statistic of Cramér-Mises type

$$\widetilde{T}_n = n \int_{\mathbb{R}^d} \left(F_n^{(X)}(z) - F_n^{(Y)}(z)\right)^2 w(z)\, \lambda^d(dz) = \frac{1}{n} \sum_{s,t=1}^{n} \widetilde{h}((X_s', Y_s)', (X_t', Y_t')'),$$

where $\widetilde{h}((x_1', y_1')', (x_2', y_2')') = \int (\mathbb{1}_{x_1 \preceq z} - \mathbb{1}_{y_1 \preceq z})(\mathbb{1}_{x_2 \preceq z} - \mathbb{1}_{y_2 \preceq z})\, w(z)\, \lambda^d(dz)$ and $F_n^{(X)}$ and $F_n^{(Y)}$ denote the empirical distribution functions based on $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. $\widetilde{T}_n$ is a V-statistic in $((X_t', Y_t')')_t$ that is degenerate under $\widetilde{\mathcal{H}}_0$. If we assume w to be non-negative, integrable and bounded, critical values of the tests can again be determined using our theory of Section 3. Moreover, consistency of this bootstrap-based test can be obtained in the same manner as in the previous subsection if w is positive almost everywhere with respect to the Lebesgue measure.

5. Simulation study

We illustrate the finite sample behavior of the dependent wild bootstrap for V-statistics by some numerical examples. We revisit the Cramér-von Mises type test of Section 4.1. The statistic T_n is applied to test

$$\mathcal{H}_0: \quad P_X = \mathcal{N}(0,1) \qquad \text{vs.} \qquad \mathcal{H}_1: \quad P_X \neq \mathcal{N}(0,1)\,.$$

The weight function w is chosen to be the density of $\mathcal{N}(0,1)$, i.e. we consider the classical Cramér-von Mises statistic $T_n = n\int_{\mathbb{R}^d}[F_n(z) - F_0(z)]^2 F_0(dz)$. In order to simulate the performance of our test under the null, we draw samples of size $n = 200$ from the stationary version of an AR(1) process

$$X_t = 0.5\, X_{t-1} + \varepsilon_t, \tag{5.1}$$

where the innovations $(\varepsilon_t)_t$ are i.i.d. standard normal random variables. Under the null hypothesis the variables $X_1, \ldots, X_n$ are positive dependent in the sense of Gleser and Moore (1983). They defined a stochastic process, whose bivariate distributions are exchangeable, to be positive dependent if

$$Eh(X_i)h(X_j) \geq 0 \quad \forall\, h \quad \text{ with } \quad E|h(X_i)h(X_j)| < \infty \quad \forall\, i,j. \tag{5.2}$$

Since the autoregression parameter is positive in the present situation, we have $cov(X_i, X_j) \geq 0\; \forall\, i,j$ which implies (5.2) under normality; see also Gleser and Moore (1983, Remark (2), Sect. 2). They showed that the limiting level of the Cramér-von Mises tests under positive dependence is at least as large as in the i.i.d. case; cf. Gleser and Moore (1983, Remark (4), Sect. 4). Thus, in general, the null hypothesis might be rejected too often if one uses quantiles of the test statistic tabulated for the i.i.d. setting when the underlying variables are positive dependent. It can be seen from Figure 1 that both our bootstrap approximations are much more accurate.

For power investigations of the test we draw samples according to (5.1) with $\mathcal{N}(0,2)$, $\mathcal{N}(0.3,1)$, and $\mathcal{N}(0.5,1)$ marginals. The wild bootstrap variables $(W_t^*)_{t=1}^n$ are generated via the procedure described in Remark 1 with $l_{200} = 10$. We replicate the simulations $N = 200$ times each with $B = 500$ bootstrap resamplings. The implementations are carried out with the aid of the statistical software package R; see R Development Core Team (2007).

The rejection frequencies of our test for nominal significance levels $\alpha = 0.05$ and $\alpha = 0.1$ are summarized in Table 1. Under the null scenario, the distribution of $T_{n,1}^*$ mimics the one of T_n better than the artificially degenerated version $T_{n,2}^*$. In contrast, the power properties of the test based on the second bootstrap statistic clearly outperfoms the first one. This is however not surprising, since $T_{n,1}^*$ tends to infinity with increasing sample size while $T_{n,2}^*$ remains bounded in the sense of $P\left(P^*\left(T_{n,2}^* > K(\epsilon)\right) < \epsilon\right) \longrightarrow_{n\to\infty} 1$.

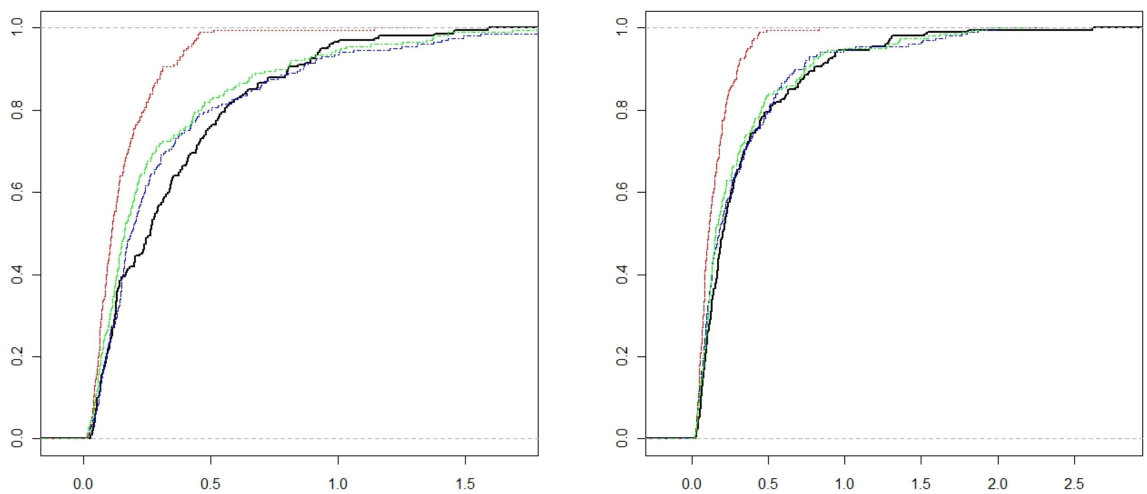

FIGURE 1. Simulated cumulative distribution functions of T_{200} (black), $T^{(*)}_{200,1}$ (blue), $T^{(*)}_{200,2}$ (green), and $T^{(ind)}_{200}$ (red) under $\mathcal{H}_0$, l.h.s n=200, r.h.s. n=500

Table 1. Rejection frequencies

	$\mathcal{N}(0,1)$		$\mathcal{N}(0,2)$		$\mathcal{N}(0.3,1)$		$\mathcal{N}(0.5,1)$	
	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
$T^*_{200,1}$	0.050	0.080	0.205	0.445	0.560	0.780	0.945	0.985
$T^*_{200,2}$	0.080	0.155	0.360	0.625	0.770	0.850	0.980	0.985

6. PROOFS

6.1. **Proofs of the main theorems.**

Proof of Theorem 2.1. *(i) Approximation of the V-statistic*

We denote by $(\lambda_k)_k$ an enumeration of the positive eigenvalues of $\lambda\Phi(x) = Eh(x, X_0)\Phi(X_0)$ in decreasing order and according to their multiplicity. The corresponding orthonormal eigenfunctions are denoted by $(\Phi_k)_k$. To avoid an explicit distinction of the cases whether the number of nonzero eigenvalues is finite or not, we set $\lambda_k := 0$ and $\Phi_k \equiv 0 \ \forall\, k > L$ if the number L of nonzero eigenvalues is finite. It follows from a version of Mercer's theorem (see Theorem 2 of Sun (2005) with $\mathcal{X} = \operatorname{supp}(P^{X_0})$) that

$$h^{(K)}(x,y) = \sum_{k=1}^{K} \lambda_k \, \Phi_k(x) \, \Phi_k(y) \underset{K\to\infty}{\longrightarrow} h(x,y) \qquad \forall\, x, y \in \operatorname{supp}(P^{X_0}). \tag{6.1}$$

The convergence of the series in (6.1) is absolute and uniform on compact subsets of $\operatorname{supp}(P^{X_0})$. The prerequisites of this result can be checked fairly easily here: Clearly, P^{X_0} is non-degenerate on

$\mathrm{supp}(P^{X_0})$ and there are compact sets $A_1 \subseteq A_2 \subseteq \cdots$ such that $\mathrm{supp}(P^{X_0}) = \bigcup_{n=1}^{\infty} A_n$. Moreover, h is a Mercer kernel (i.e. continuous, symmetric, positive semidefinite), $\int h^2(x,y)P^{X_0}(dy) \leq h(x,x)Eh(X_0,X_0) < \infty \quad \forall\, x \in \mathrm{supp}(P^{X_0})$ and $\iint h^2(x,y)P^{X_0}(dx)P^{X_0}(dy) \leq (Eh(X_0,X_0))^2 < \infty$. Thus, assumptions 1,2 and 3 of Sun (2005) are fulfilled, see also his Proposition 1.

We will approximate V_n by a V-statistic with a kernel having a finite spectral decomposition,

$$V_n^{(K)} \;=\; \frac{1}{n}\sum_{s,t=1}^{n} h^{(K)}(X_s,X_t).$$

Due to the positive semidefiniteness of h, all eigenvalues are non-negative which implies that $V_n - V_n^{(K)} \geq 0$. Hence,

$$\begin{aligned}
&E\big|V_n \,-\, V_n^{(K)}\big| \\
&\quad = \quad E\left[V_n \,-\, V_n^{(K)}\right] \\
&\quad = \quad E\left[h(X_0,X_0) \,-\, h^{(K)}(X_0,X_0)\right] \,+\, \sum_{r=1}^{n-1} 2\,(1-r/n)\,E\left[h(X_0,X_r) \,-\, h^{(K)}(X_0,X_r)\right].
\end{aligned}$$

The first term on the right-hand side converges to zero as $K \to \infty$ by majorized convergence. A repeated application of the Cauchy-Schwarz inequality yields for the second term

$$\begin{aligned}
&\sum_{r=1}^{n-1} 2\,(1-r/n)\,E\left[h(X_0,X_r) \,-\, h^{(K)}(X_0,X_r)\right] \\
&\quad \leq \quad 2\sum_{r=1}^{\infty}\left|E\left[\sum_{k=K+1}^{\infty}\lambda_k\Phi_k(X_0)\Phi_k(X_r)\right]\right| \\
&\quad = \quad 2\sum_{r=1}^{\infty}\left|E\left[\sum_{k=K+1}^{\infty}\lambda_k\Phi_k(X_0)\left(\Phi_k(X_r)-\Phi_k(\widetilde{X}_r)\right)\right]\right| \\
&\quad \leq \quad 2\sum_{r=1}^{\infty}\sqrt{E\left[\sum_{k=K+1}^{\infty}\lambda_k\Phi_k^2(X_0)\right]}\sqrt{E\left[\sum_{k=K+1}^{\infty}\lambda_k\left(\Phi_k(X_r)-\Phi_k(\widetilde{X}_r)\right)^2\right]} \\
&\quad \leq \quad 2\sqrt{\sum_{k=K+1}^{\infty}\lambda_k}\sum_{r=1}^{\infty}\sqrt{E\left[\sum_{k=1}^{\infty}\lambda_k\left(\Phi_k(X_r)-\Phi_k(\widetilde{X}_r)\right)^2\right]} \\
&\quad \leq \quad 2\sqrt{\sum_{k=K+1}^{\infty}\lambda_k}\;\sum_{r=1}^{\infty}\sqrt{E\left[h(X_r,X_r)-h(X_r,\widetilde{X}_r)-h(\widetilde{X}_r,X_r)+h(\widetilde{X}_r,\widetilde{X}_r)\right]} \\
&\quad \leq \quad 2\sqrt{\sum_{k=K+1}^{\infty}\lambda_k}\sum_{r=1}^{\infty}\sqrt{2\,\mathrm{Lip}(h)}\sqrt{\tau(r)},
\end{aligned}$$

where $\widetilde{X}_r$ denotes a copy of X_r that is independent of X_0 and satisfies $E|X_r - \widetilde{X}_r| \leq \tau(r)$. Since $\sum_{k=1}^{\infty} \lambda_k = Eh(X_0, X_0) < \infty$ and thus $\sum_{k=K}^{\infty} \lambda_k \to 0$ as $K \to \infty$, we obtain

$$\sup_n E\left|V_n \;-\; V_n^{(K)}\right| \underset{K\to\infty}{\longrightarrow} 0. \tag{6.2}$$

(ii) CLT for partial sums

We will show that for $K \leq L$

$$V_n^{(K)} \;=\; \sum_{k=1}^{K} \lambda_k \left(n^{-1/2} \sum_{t=1}^{n} \Phi_k(X_t)\right)^2 \;\stackrel{d}{\longrightarrow}\; \sum_{k=1}^{K} \lambda_k Z_k^2. \tag{6.3}$$

By the continuous mapping theorem, this follows from

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \;\stackrel{d}{\longrightarrow}\; (Z_1, \ldots, Z_K)', \tag{6.4}$$

with $Y_t = (\Phi_1(X_t), \ldots, \Phi_K(X_t))'$. To prove this, we will apply a multivariate version of the central limit theorem of Neumann (2011) to the random variables $n^{-1/2} Y_t$; see Theorem 6.1 of the present paper. We will check the conditions of this theorem in the following.

First, $(Y_t)_{t\in\mathbb{N}}$ is a strictly stationary sequence of centered random vectors whose components have unit variance. This implies in particular that the Lindeberg condition (6.23) is fulfilled.

According to (A1), for any $s, r \in \mathbb{N}$, there exists a random variable $\widetilde{X}_{s+r}$ that is independent of $X_s, X_{s-1}, \ldots$, has the same distribution as X_{s+r} and satisfies $E\|X_{s+r} - \widetilde{X}_{s+r}\|_1 \leq \tau(r)$. This implies that

$$\begin{aligned}
&\big(\operatorname{cov}(\Phi_j(X_s), \Phi_k(X_{s+r})\big)^2 \\
&\quad= \left(E\left[\Phi_j(X_s)\left(\Phi_k(X_{s+r}) \;-\; \Phi_k(\widetilde{X}_{s+r})\right)\right]\right)^2 \\
&\quad\leq E\left(\Phi_k(X_{s+r}) \;-\; \Phi_k(\widetilde{X}_{s+r})\right)^2 \\
&\quad\leq \frac{1}{\lambda_k}\, E\left[h(X_{s+r}, X_{s+r}) - h(X_{s+r}, \widetilde{X}_{s+r}) - h(\widetilde{X}_{s+r}, X_{s+r}) + h(\widetilde{X}_{s+r}, \widetilde{X}_{s+r})\right] \\
&\quad\leq \frac{2\ \operatorname{Lip}(h)}{\lambda_k}\, \tau(r).
\end{aligned} \tag{6.5}$$

Hence, we obtain by majorized convergence

$$\begin{aligned}
&\left|\frac{1}{n} \operatorname{cov}\left(\sum_{t=1}^{n} \Phi_j(X_t), \sum_{t=1}^{n} \Phi_k(X_t)\right) \;-\; \sum_{r=-\infty}^{\infty} \operatorname{cov}(\Phi_j(X_0), \Phi_k(X_r))\right| \\
&\quad\leq \sum_{r\in\mathbb{Z}} \min\{|r|/n, 1\}\, |\operatorname{cov}(\Phi_j(X_0), \Phi_k(X_r))| \underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}$$

i.e., condition (6.22) is fulfilled.

Note that it follows from (6.5) that for all $k = 1, \ldots, K$,

$$E\left(\Phi_k(X_{s+r}) \,-\, \Phi_k(\widetilde{X}_{s+r})\right)^2 \,=\, O(\tau(r)). \tag{6.6}$$

Therefore, we obtain by the Cauchy-Schwarz inequality, for $1 \leq s_1 < s_2 < \ldots < s_u < s_u + r = t_1 \leq t_2 \leq n$ and for any measurable function $g\colon \mathbb{R}^u \to \mathbb{R}$ with $\|g\|_\infty = \sup_{x \in \mathbb{R}^u} |g(x)| \leq 1$ that

$$\frac{1}{n}\left|\operatorname{cov}(g(n^{-1/2}\,Y_{s_1}, \ldots, n^{-1/2}\,Y_{s_u})Y_{s_u,j}, Y_{t_1,k})\right| \,=\, O\left(\frac{\sqrt{\tau(r)}}{n}\right)$$

and

$$\frac{1}{n}\left|\operatorname{cov}(g(n^{-1/2}\,Y_{s_1}, \ldots, n^{-1/2}\,Y_{s_u}), Y_{t_1,j}Y_{t_2,k})\right| \,=\, O\left(\frac{\sqrt{\tau(r)}}{n}\right),$$

i.e., conditions (6.24) and (6.25) are also fulfilled. Hence, (6.4) follows from Theorem 6.1.

(iii) Conclusion

The assertion of the theorem concerning the V-type statistics follows from (6.2), (6.3), and Theorem 2 of Dehling, Durieu, and Volny (2009), which is an improved variant of Theorem 4.2 of Billingsley (1968) for complete spaces. Moreover, note that $U_n \,=\, V_n \,-\, n^{-1}\sum_{t=1}^n h(X_t, X_t)$. Therefore, the limit distribution of U_n can be deduced from the one of V_n and a weak law of large numbers. The version of Leucht (2011, Lemma 5.1) for sequences of random variables $(g(X_t))_t$, where $g\colon \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz-continuous function and $(X_t)_t$ is a stationary sequence of τ-dependent random variables with $E|g(X_1)| < \infty$, can be applied here. □

Proof of Corollary 2.1. In the proof of Theorem 2.1, Lipschitz continuity of h is invoked to obtain the inequality

$$\sum_{r \in \mathbb{N}} \sqrt{E[h(X_r, X_r) - h(X_r, \widetilde{X}_r) - h(\widetilde{X}_r, X_r) + h(\widetilde{X}_r, \widetilde{X}_r)]} < \infty,$$

where $\widetilde{X}_r$ denotes a copy of X_r that is independent of $X_s, s \leq 0$. Under (A3) one obtains a similar result by Hölder's inequality,

$$\begin{aligned}
&\sum_{r \in \mathbb{N}} \sqrt{E[h(X_r, X_r) - h(X_r, \widetilde{X}_r) - h(\widetilde{X}_r, X_r) + h(\widetilde{X}_r, \widetilde{X}_r)]} \\
&\leq \sum_{r \in \mathbb{N}} \sqrt{E\left([f(X_r, X_r, X_r, \widetilde{X}_r) + f(\widetilde{X}_r, \widetilde{X}_r, X_r, \widetilde{X}_r)]\|X_r - \widetilde{X}_r\|_1\right)} \\
&\leq \sum_{r \in \mathbb{N}} \sqrt{\left(E[f(X_r, X_r, X_r, \widetilde{X}_r) + f(\widetilde{X}_r, \widetilde{X}_r, X_r, \widetilde{X}_r)]^{1/(1-\delta)}(\|X_r\|_1 + \|\widetilde{X}_r\|_1)\right)^{1-\delta} (\tau(r))^\delta} \\
&< \infty.
\end{aligned}$$

Moreover, since h is Lipschitz continuous on any compact set due to the assumptions on the function f, the limit distribution of U_n can be deduced by the WLLN as before. □

Proof of Theorem 3.1. *(i) V-statistics*

(a) Asymptotic equivalence of $V_{n,1}^*$ and $V_{n,2}^*$

First we show that the effect of the empirical degeneration is asymptotically negligible if the kernel is already degenerate. We have

$$\begin{aligned} & V_{n,2}^* - V_{n,1}^* \\ = & -\frac{2}{n^2}\sum_{s,t,k=1}^n h(X_s,X_k)W_s^*W_t^* + \frac{1}{n^3}\sum_{k,l,s,t=1}^n h(X_k,X_l)W_s^*W_t^* \\ = & \frac{1}{2n^2}\sum_{s,t=1}^n \{(W_s^*-1)h(X_s,X_t)(W_t^*-1) - (W_s^*+1)h(X_s,X_t)(W_t^*+1)\}\sum_{k=1}^n W_k^* \\ & + V_n\left(\frac{1}{n}\sum_{s=1}^n W_s^*\right)^2. \end{aligned}$$

$V_{n,-} := n^{-1}\sum_{s,t=1}^n (W_s^*-1)h(X_s,X_t)(W_t^*-1)$ and $V_{n,+} := n^{-1}\sum_{s,t=1}^n (W_s^*+1)h(X_s,X_t)(W_t^*+1)$ are non-negative V-statistics and we get $EE^*[V_{n,-}] = O(1)$ and $EE^*[V_{n,+}] = O(1)$. Moreover, $E^*(\sum_{k=1}^n W_k^*)^2 = O(n\, l_n)$ and $E[V_n] = O(1)$. Therefore, we obtain that

$$P\left(P^*\left(|V_n^{*(2)} - V_n^{*(1)}| > \epsilon\right) > \delta\right) \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall \epsilon > 0, \delta > 0,$$

i.e., the difference between $V_{n,1}^*$ and $V_{n,2}^*$ is asymptotically neglibible. Therefore, we only consider $V_{n,1}^*$ below.

(b) Approximation of the V-statistic

We define

$$V_{n,1}^{(K)*} := \frac{1}{n}\sum_{s,t=1}^n h^{(K)}(X_s,X_t)\, W_s^*W_t^* = \sum_{k=1}^K \lambda_k \left(\frac{1}{\sqrt{n}}\sum_{s=1}^n \Phi_k(X_s)W_s^*\right)^2.$$

Since $h(\cdot,\cdot) - h^{(K)}(\cdot,\cdot)$ is a positive semidefinite function, we have $V_{n,1}^* \geq V_{n,1}^{(K)*}$ for all K. In a first step we will show that

$$\limsup_{n\to\infty} P\left(P^*(|V_{n,1}^* - V_{n,1}^{(K)*}| > \varepsilon) > \delta\right) \underset{K\to\infty}{\longrightarrow} 0 \qquad \forall \delta,\varepsilon > 0. \tag{6.7}$$

This follows from Markov's inequality and the subsequent approximation:

$$\begin{aligned}
& EE^*(V_{n,1}^* - V_{n,1}^{(K)*}) \\
&= \frac{1}{n} \sum_{k=K+1}^{\infty} \lambda_k \sum_{s,t=1}^{n} E[\Phi_k(X_s)\Phi_k(X_t)]\, \rho(|s-t|/l_n) \\
&\le \sum_{k=K+1}^{\infty} \lambda_k \left\{ 1 + \sum_{r=1}^{n-1} \frac{2(n-r)}{n} \left| E\left[\Phi_k(X_0) \left(\Phi_k(X_r) - \Phi_k(\widetilde{X}_r) \right) \right] \right| \, |\rho(r/l_n)| \right\} \\
&\le \sum_{k=K+1}^{\infty} \lambda_k + 2 \sum_{r=1}^{n-1} \sqrt{\sum_{k=K+1}^{\infty} \lambda_k} \sqrt{E \sum_{k=K+1}^{\infty} \lambda_k \left[\Phi_k(X_r) - \Phi_k(\widetilde{X}_r) \right]^2} \\
&\le \sum_{k=K+1}^{\infty} \lambda_k + 2 \sqrt{\sum_{k=K+1}^{\infty} \lambda_k} \sqrt{2 \operatorname{Lip}(h)} \sum_{r=1}^{\infty} \sqrt{\tau(r)},
\end{aligned}$$

where $\widetilde{X}_r$ denotes a copy of X_r that is independent of X_0 and such that $E\|X_r - \widetilde{X}_r\|_1 \le \tau_r$.

(c) CLT for the partial sums

We define the vectors $Y_t^* = (\Phi_1(X_t), \ldots, \Phi_K(X_t))' W_t^*$. We will show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t^* \xrightarrow{d} (Z_1, \ldots, Z_K)' \sim \mathcal{N}(0_K, \Sigma_K) \qquad \text{in probability} \tag{6.8}$$

where $(\Sigma_K)_{j,k} = \sum_{r=-\infty}^{\infty} \operatorname{cov}(\Phi_j(X_0), \Phi_k(X_r))$. To this end, we decompose the index set $\{1, \ldots, n\}$ into disjoint blocks of length L_n, $I_{n,s} = \{t \in \mathbb{N}\colon (s-1)L_n < t \le sL_n\}$, for $s = 1, \ldots, k_n - 1$ with $k_n - 1 = [n/L_n]$, and a block of the remaining indices, $I_{n,k_n} = \{t \in \mathbb{N}\colon (k_n - 1)L_n < t \le n\}$. We define $X_{n,s}^* = n^{-1/2} \sum_{t \in I_{n,s}} Y_t^*$ and show that the conditions of Corollary 6.1 below are fulfilled by the triangular scheme $(X_{n,s}^*)_{s=1,\ldots,k_n}$, $n \in \mathbb{N}$.

(c.1) Convergence of variances and covariances

We first show that

$$P\left(\sum_{s=1}^{k_n} E^*(X_{n,s,j}^*)^2 \le v_0 \right) \underset{n\to\infty}{\longrightarrow} 1.$$

According to the definition of the blocks, we obtain

$$\begin{aligned}
\sum_{s=1}^{k_n} E^*(X_{n,s,j}^*)^2 &= \sum_{s=1}^{k_n - 1} \frac{1}{n} \sum_{u.v \in I_{n,s}} \Phi_j(X_u)\Phi_j(X_v)\rho(|u-v|/l_n) + o_P(1) \\
&\le \frac{1}{L_n} \sum_{u,v \in I_{n,1}} E[\Phi_j(X_u)\Phi_j(X_v)]\, \rho(|u-v|/l_n) + o_P(1).
\end{aligned}$$

The latter inequality follows from Lemma 6.1 after a substitution of n by L_n when the block length L_n is chosen such that $l_n = o(L_n)$. The remaining term can be bounded by some finite constant uniformly in n due to the τ-dependence condition (B1) on $(X_t)_t$.

Next we will show that

$$\mathrm{Cov}^*\left(X^*_{n,1} + \cdots + X^*_{n,k_n}\right) \stackrel{P}{\longrightarrow} \Sigma_K. \tag{6.9}$$

We have that

$$\begin{aligned}
&\left(\mathrm{Cov}^*\left(X^*_{n,1} + \cdots + X^*_{n,k_n}\right)\right)_{j,k} \\
&= \mathrm{cov}^*\left(n^{-1/2}\sum_{s=1}^n \Phi_j(X_s)W_s^*, n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)W_t^*\right) \\
&= \frac{1}{n}\sum_{s,t=1}^n \Phi_j(X_s)\Phi_k(X_t)\rho(|s-t|/l_n) \\
&= \frac{1}{n}\sum_{s,t=1}^n \left\{\Phi_j(X_s)\Phi_k(X_t) - E[\Phi_j(X_s)\Phi_k(X_t)]\right\}\rho(|s-t|/l_n) \\
&\quad + \sum_{r=-\infty}^{\infty} E[\Phi_j(X_0)\Phi_k(X_r)]\,\rho(|r|/l_n)\,\max\{1-|r|/n, 0\}.
\end{aligned}$$

The first term on the right-hand side converges to zero by Lemma 6.1 below while the second one converges to $(\Sigma_K)_{j,k}$ by majorized convergence. Hence, (6.9) holds true.

(c.2) Lindeberg condition

We want to show that, for arbitrary $\epsilon > 0$,

$$L^*_{n,j}(\epsilon) := \frac{1}{n}\sum_{s=1}^{k_n} E^*\left[(X^*_{n,s,j})^2 \mathbb{1}(|X^*_{n,s,j}| > \epsilon\sqrt{n})\right] \stackrel{P}{\longrightarrow} 0. \tag{6.10}$$

For this, it suffices to show that

$$E[L^*_{n,j}(\epsilon)] \underset{n\to\infty}{\longrightarrow} 0. \tag{6.11}$$

We obtain from the stationarity of the involved processes that

$$E[L^*_{n,j}(\epsilon)] \le \frac{k_n - 1}{n}\, EE^*\left[(X^*_{n,1,j})^2 \mathbb{1}(|X^*_{n,1,j}| > \epsilon\sqrt{n})\right] + \frac{1}{n} EE^*\left[(X^*_{n,k_n,j})^2\right].$$

While the second term on the right-hand side is obviously negligible, we show convergence to zero of the first one in the following. It is easy to see that the random variables $(L_n^{-1/2}\Phi_j(X_t)W_t^*)_{t\in I_{n,1}}$, when expectations with respect to both random mechanisms are taken, satisfy the conditions of Theorem 6.1. Hence, we obtain that

$$n^{1/2}L_n^{-1/2}X^*_{n,1,j} = L_n^{-1/2}\sum_{t\in I_{n,1}} \Phi_j(X_t)W_t^* \stackrel{d}{\longrightarrow} N(0, v_j) \tag{6.12}$$

with some $v_j \in [0,\infty)$. On the other hand, we have that

$$EE^*\left[n/L_n\,(X^*_{n,1,j})^2\right] \underset{n\to\infty}{\longrightarrow} v_j. \tag{6.13}$$

(6.12) and (6.13) imply uniform integrability of $(n/L_n\ (X^*_{n,1,j})^2)_{n\in\mathbb{N}}$, which also yields that

$$EE^*\left[n/L_n\ (X^*_{n,1,j})^2\mathbb{1}(|\sqrt{n/L_n}\ X^*_{n,1,j}|>c)\right]\underset{n\to\infty}{\longrightarrow} E[Y_j^2\mathbb{1}(|Y_j|>c)], \tag{6.14}$$

where $Y_j\sim N(0,v_j)$. Since $E[Y_j^2\mathbb{1}(|Y_j|>c)]\longrightarrow_{c\to\infty}0$, we obtain from (6.14) that (6.11) holds true.

(c.3) Weak dependence

The conditions (6.28) and (6.29) can be proved along the same lines. Therefore, we only state the proof of (6.28) in Lemma 6.3 below.

(d) Conclusion

Now we can apply Corollary 6.1 and we obtain (6.8), which in turn implies by the continuous mapping theorem that

$$V_{n,1}^{(K)^*}\stackrel{d}{\longrightarrow}\sum_{k=1}^K\lambda_k Z_k^2\qquad\text{in probability.} \tag{6.15}$$

Invoking Theorem 2 of Dehling, Durieu, and Volny (2009), we obtain from (6.8) and (6.15) that

$$V_{n,1}^*\stackrel{d}{\longrightarrow}Z\qquad\text{in probability.}$$

(ii) U-statistics

In order to deduce the corresponding result for $U^*_{n,1}$, we have to show that

$$P^*\left(\left|\frac{1}{n}\sum_{t=1}^n h(X_t,X_t)(W_t^*)^2-Eh(X_0,X_0)\right|>\epsilon\right)\stackrel{P}{\longrightarrow}0\quad\forall\epsilon>0. \tag{6.16}$$

With $M>0$, we decompose:

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^n h(X_t,X_t)(W_t^*)^2\ -\ Eh(X_0,X_0)\right|\\
&\leq\ \left|\frac{1}{n}\sum_{t=1}^n\left[h(X_t,X_t)\ -\ h(X_t,X_t)\wedge M\right](W_t^*)^2\right|\\
&\qquad+\left|\frac{1}{n}\sum_{t=1}^n h(X_t,X_t)\wedge M\left[(W_t^*)^2\ -\ 1\right]\right|\\
&\qquad+\left|\frac{1}{n}\sum_{t=1}^n h(X_t,X_t)\wedge M\quad-\quad E[h(X_0,X_0)\wedge M]\right|\\
&\qquad+\left|E[h(X_0,X_0)\wedge M]\ -\ E[h(X_0,X_0)]\right|\\
&=\ R_{n,1}+\cdots+R_{n,4}.
\end{aligned}$$

We obtain from monotone convergence that

$$R_{n,4} \underset{M\to\infty}{\longrightarrow} 0$$

and

$$EE^*[R_{n,1}] \underset{M\to\infty}{\longrightarrow} 0.$$

Moreover, it follows from the WLLN of Leucht (2011, Lemma 5.1) that

$$R_{n,3} \stackrel{P}{\longrightarrow} 0.$$

Thus, it remains to prove that

$$P^*\left(\left|\frac{1}{n}\sum_{t=1}^{n}[h(X_t,X_t)\wedge M][(W_t^*)^2-1]\right| > \epsilon\right) \stackrel{P}{\longrightarrow} 0 \quad \forall\, \epsilon>0. \tag{6.17}$$

To this end, we truncate the bootstrap variables and obtain

$$\begin{aligned}
&P^*\left(\left|\frac{1}{n}\sum_{t=1}^{n}[h(X_t,X_t)\wedge M][(W_t^*)^2-1]\right| > \epsilon\right)\\
&\le P^*\left(\left|\frac{1}{n}\sum_{t=1}^{n}[h(X_t,X_t)\wedge M][(W_t^*)^2\wedge K - E^*[(W_t^*)^2\wedge K]]\right| > \epsilon/3\right)\\
&\quad + \frac{3M}{\epsilon}(1-E^*[(W_1^*)^2\wedge K])\\
&\quad + \mathbb{1}\left(M(1-E^*[(W_1^*)^2\wedge K])>\epsilon/3\right).
\end{aligned}$$

The first term on the right-hand side is of order $o_P(1)$ for all $K\in\mathbb{N}$ by Chebyshev's inequality and the τ-dependence of the bootstrap variables. For fixed M, the two remaining summands are less than any $\delta>0$ if $K=K(\delta,M)$ is chosen sufficiently large, which then implies (6.17) and thus (6.16).

In the case of $U_{n,2}^*$, we have to prove that

$$P^*\left(\left|\frac{1}{n}\sum_{t=1}^{n}\bar{h}(X_t,X_t)(W_t^*)^2 - Eh(X_0,X_0)\right| > \epsilon\right) \stackrel{P}{\longrightarrow} 0 \quad \forall\, \epsilon>0. \tag{6.18}$$

We split up

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\bar{h}(X_t,X_t)(W_t^*)^2\\
&= \frac{1}{n}\sum_{t=1}^{n}h(X_t,X_t)(W_t^*)^2 - \frac{2}{n}\sum_{t,k=1}^{n}h(X_t,X_k)(W_t^*)^2 + V_n\,\frac{1}{n^2}\sum_{t=1}^{n}(W_t^*)^2.
\end{aligned}$$

According to (6.16), the first term on the right-hand side converges to $Eh(X_0,X_0)$. The third term is obviously negligible since $V_n=O_P(1)$ and $E[n^{-2}\sum_t (W_t^*)^2]=n^{-1}$. As to the second term, note first that

$$EE^* \left| \frac{1}{n^2} \sum_{t,k=1}^{n} h(X_t, X_t) \left((W_t^*)^2 \; - \; (W_t^*)^2 \wedge K \right) \right| \underset{K \to \infty}{\longrightarrow} 0.$$

Moreover, since h is positive semidefinite, we have that

$$\left| \frac{1}{n^2} \sum_{t,k=1}^{n} h(X_t, X_k)(W_t^*)^2 \wedge K \right| \leq \frac{\sqrt{V_n}}{n} \sqrt{\frac{1}{n} \sum_{t,k=1}^{n} h(X_t, X_k)(W_t^*)^2 \wedge K(W_k^*)^2 \wedge K}.$$

Since $EE^*[n^{-1} \sum_{t,k=1}^{n} h(X_t, X_k)(W_t^*)^2 \wedge K(W_k^*)^2 \wedge K] = O(1)$ we obtain (6.18).

(iii) Convergence in the uniform norm

Convergence of the distribution functions in the uniform norm can be deduced from the distributional convergence in conjunction with the continuity of the limiting distribution function. □

Proof of Lemma 3.1. (a) *Expectation of Z*

We have

$$EZ = \sum_{k=1}^{\infty} \sum_{t=-\infty}^{\infty} \lambda_k E\Phi_k(X_0)\Phi_k(X_t).$$

Since the double sum is absolutely convergent, we can interchange the order of summation and obtain

$$EZ = \sum_{t=-\infty}^{\infty} \lim_{K \to \infty} E \sum_{k=1}^{K} \lambda_k \Phi_k(X_0)\Phi_k(X_t).$$

The quantity $|\sum_{k=1}^{K} \lambda_k \Phi_k(X_0)\Phi_k(X_t)|$ can be bounded from above by

$$\sum_{k=1}^{\infty} \lambda_k [\Phi_k^2(X_0) + \Phi_k^2(X_t)] = h(X_0, X_0) + h(X_t, X_t)$$

which is integrable. Lebesgue's dominated convergence theorem finally yields

$$EZ = \sum_{t=-\infty}^{\infty} E \sum_{k=1}^{\infty} \lambda_k \Phi_k(X_0)\Phi_k(X_t) = \sum_{t=-\infty}^{\infty} Eh(X_0, X_t).$$

(b) *Continuity of the limit distribution function*

We deduce the continuity of the limit distribution function from $EZ > 0$: Let $Z^{(K)} = \sum_{k=1}^{K} \lambda_k Z_k^2$. Hence,

$$E\big|Z - Z^{(K)}\big| \; = \; E\big[Z - Z^{(K)}\big] \; = \; \sum_{k=K+1}^{\infty} \lambda_k E Z_k^2 \; \leq \; C \sqrt{\sum_{k=K+1}^{\infty} \lambda_k} \underset{K \to \infty}{\longrightarrow} 0,$$

which implies that

$$Z^{(K)} \; \stackrel{d}{\longrightarrow} \; Z,$$

as $K \to \infty$. It follows from Portmanteau's theorem (see Theorem 2.1 in Billingsley (1968)) for each fixed $x_0 \in \mathbb{R}$ and $\delta > 0$ that

$$P\left(Z \in (x_0 - \varepsilon, x_0 + \varepsilon)\right) \le \liminf_{K\to\infty} P(Z^{(K)} \in (x_0 - \varepsilon, x_0 + \varepsilon)).$$

Hence, it suffices to show that, for arbitrary $\delta > 0$,

$$\liminf_{K\to\infty} P(Z^{(K)} \in (x_0 - \varepsilon, x_0 + \varepsilon)) \le \delta \tag{6.19}$$

whenever $\varepsilon > 0$ is chosen sufficiently small. To this end, first note that $Z^{(K)}$ has the same distribution as $N_K' \Sigma_K^{1/2} \Lambda_K \Sigma_K^{1/2} N_K$, where $\Lambda_K = \mathrm{Diag}(\lambda_1, \ldots, \lambda_K)$, $\Sigma_K = \mathrm{Cov}((Z_1, \ldots, Z_K)')$ and $N_K \sim \mathcal{N}(0_K, I_K)$. For any symmetric matrix M, denote by $\alpha_i(M)$ the ith largest eigenvalue of this matrix. Then $\alpha_i(\Sigma_K^{1/2} \Lambda_K \Sigma_K^{1/2}) = \alpha_i(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2})$; see Lütkepohl (1996, Section 5.2.1, page 65). Since $\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}$ is a principle submatrix of $\Lambda_{K+1}^{1/2} \Sigma_{K+1} \Lambda_{K+1}^{1/2}$ we obtain by the inclusion principle for principle submatrices (see Lütkepohl (1996, Section 9.13.4, page 160)) that $\alpha_k(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}) \le \alpha_k(\Lambda_{K+1}^{1/2} \Sigma_{K+1} \Lambda_{K+1}^{1/2})$. Moreover, it follows from $EZ^{(K)} = \sum_{k=1}^K \alpha_k(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}) > EZ/2$, $\forall\, K > K_0$ that $\alpha_1(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}) \ge c_0 > 0$, $\forall K > K_0$. As convolution preserves the continuity properties of the smoother function, the latter inequality implies continuity of the distribution functions of $Z^{(K)} = \sum_{k=1}^K \alpha_k(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2})\, Y_k^2$, uniformly for all $K > K_0$, where $Y_1, \ldots, Y_k$ are i.i.d. standard normal random variables. $\square$

Proof of Proposition 4.1. We obtain under (A1) that

$$\begin{aligned} & E[n^{-1} T_n] \;-\; E[h_1(X_0, \widetilde{X}_0)] \\ & \quad = \; \frac{1}{n} \sum_{r=-(n-1)}^{n-1} (1 \,-\, |r|/n)\; E[h(X_0, X_r) \,-\, h(X_0, \widetilde{X}_0)] \\ & \quad = \; O\left(\frac{1}{n}\right) \end{aligned}$$

and

$$\begin{aligned} & \mathrm{var}(n^{-1} T_n) \\ & \quad = \; \frac{1}{n^4} \sum_{s,t,u,v=1}^{n} \{E[h(X_s, X_t)\, h(X_u, X_v)] \,-\, E[h(X_s, X_t)]\; E[h(X_u, X_v)]\} \\ & \quad \underset{n\to\infty}{\longrightarrow} \; 0. \end{aligned}$$

This implies (i).

Furthermore, we have

$$EE^*[n^{-1} T_{n,1}^*] \;=\; \frac{1}{n} \sum_{r=-(n-1)}^{n-1} (1 \,-\, |r|/n)\; E[h(X_0, X_r)]\; \rho(|r|/l_n) \;=\; O\left(\frac{l_n}{n}\right),$$

which implies that, for all $\epsilon, \delta > 0$,

$$P\left(P^*\left(n^{-1}T^*_{n,1} > \epsilon\right) \leq \delta\right) \underset{n\to\infty}{\longrightarrow} 1,$$

i.e., the first part of (ii) holds true. To treat the second part with the empirically degenerated kernel, we introduce the exactly degenerated kernel h_{deg} as

$$h_{deg}(x,y) = h(x,y) - \int h(x,y)\, P^{X_0}(dx) - \int h(x,y)\, P^{X_0}(dy) + \iint h(x,y)\, P^{X_0}(dx)\, P^{X_0}(dy).$$

It follows that

$$T^*_{n,2} = \frac{1}{n}\sum_{s,t=1}^{n}\left[h_{deg}(X_s,X_t) - n^{-1}\sum_{k=1}^{n} h_{deg}(X_s,X_k) - n^{-1}\sum_{k=1}^{n} h_{deg}(X_k,X_t) + n^{-2}\sum_{k,l=1}^{n} h_{deg}(X_k,X_l)\right] W^*_s W^*_t,$$

i.e., under empirical degeneration, we can replace h by h_{deg}. This means that $T^*_{n,2}$ can be written an empirically degenerated V-statistic with the kernel h_{deg} that is degenerate under P^{X_0}. Hence, we obtain from the part concerning $V^*_{n,2}$ of Theorem 3.1 that $T^*_{n,2}$ converges to some random variable in probability. As a consequence, we get the second part of (ii).

Finally, (iii) is an immediate consequence of (i) and (ii). □

6.2. **A multivariate bootstrap CLT.** The Cramér-Wold device is a very useful tool when asymptotic normality for a sequence of random vectors has to be proved and only a univariate CLT is available. However, this approach might be problematic in the context of bootstrap processes, where usually asymptotic normality only with the qualification "in probability" can be proved. To see what could happen, assume that there is a sequence of $\mathbb{R}^d$-valued bootstrap random vectors $Z^*_1, Z^*_2, \ldots$, where the distribution of Z^*_n depends on random variables $X_1, \ldots, X_n$. Suppose further, that we can exploit some univariate CLT to show that, for any arbitrary $c \in \mathbb{R}^d$,

$$c'Z^*_n \xrightarrow{d} c'Z \qquad \text{in probability,} \tag{6.20}$$

where $Z \sim N(0_d, \Sigma)$. This is, however, not sufficient in general for a proof of the asymptotic normality of Z^*_n, i.e. of

$$Z^*_n \xrightarrow{d} Z \qquad \text{in probability.} \tag{6.21}$$

To see why, note that (6.20) can be reformulated in such a way that there exist "bad sets" $\Omega_1(c), \Omega_2(c), \ldots$ such that

$$P\left((X_1,\ldots,X_n)' \in \Omega_n(c)\right) \underset{n\to\infty}{\longrightarrow} 0$$

and, for any sequence $(\omega_n)_{n\in\mathbb{N}}$ with $\omega_n \notin \Omega_n(c)$,

$$P^{c'Z^*_n|(X_1,\ldots,X_n)'=\omega_n} \Longrightarrow P^{c'Z}.$$

If the universal sets $\Omega_n := \bigcup_{c\in\mathbb{R}^d}\Omega_n(c)$ were measurable and satisfy $P((X_1,\ldots,X_n)'\in\Omega_n)\longrightarrow_{n\to\infty}0$, then we could actually conclude from (6.20) that (6.21) holds true. However, if this fails, then this conclusion is no longer correct in general. To overcome this difficulty, we formulate first an appropriate multivariate CLT for triangular arrays of weakly dependent random variables and present then, as an immediate consequence, a version tailor-made for bootstrap processes.

Theorem 6.1. *Suppose that $(X_{n,k})_{k=1,\ldots,k_n}$, $n\in\mathbb{N}$, is a triangular scheme of $\mathbb{R}^d$-valued random vectors with $EX_{n,k}=0_d$ for all n,k and $\sum_{k=1}^{k_n}EX_{n,k,j}^2\le v_0$, for all $n\in\mathbb{N}, j\in\{1,\ldots,d\}$ and some $v_0<\infty$. We assume that*

$$\Sigma_n := \operatorname{Cov}(X_{n,1}+\cdots+X_{n,k_n}) \underset{n\to\infty}{\longrightarrow} \Sigma \tag{6.22}$$

for some positive semidefinite matrix Σ, and that

$$\sum_{k=1}^{k_n} E[X_{n,k,j}^2\mathbb{1}(|X_{n,k,j}|>\epsilon)] \underset{n\to\infty}{\longrightarrow} 0 \tag{6.23}$$

holds for all $\epsilon>0$, $j\in\{1,\ldots,d\}$. Furthermore, we assume that there exists a summable sequence $(\theta_r)_{r\in\mathbb{N}}$ such that, for all $u\in\mathbb{N}$, all indices $1\le s_1<s_2<\cdots<s_u<s_u+r=t_1\le t_2\le k_n$ and all $j_1,j_2\in\{1,\ldots,d\}$, the following upper bounds for covariances hold true: for all measurable functions $g\colon\mathbb{R}^{du}\longrightarrow\mathbb{R}$ with $\|g\|_\infty=\sup_{x\in\mathbb{R}^{du}}|g(x)|\le 1$,

$$|\operatorname{cov}\left(g(X_{n,s_1},\ldots,X_{n,s_u})X_{n,s_u,j_1},X_{n,t_1,j_2}\right)| \le \left(EX_{n,s_u,j_1}^2 + EX_{n,t_1,j_2}^2 + k_n^{-1}\right)\theta_r \tag{6.24}$$

and

$$|\operatorname{cov}\left(g(X_{n,s_1},\ldots,X_{n,s_u}),X_{n,t_1,j_1}X_{n,t_2,j_2}\right)| \le \left(EX_{n,t_1,j_1}^2 + EX_{n,t_2,j_2}^2 + k_n^{-1}\right)\theta_r. \tag{6.25}$$

Then

$$X_{n,1}+\cdots+X_{n,k_n} \stackrel{d}{\longrightarrow} \mathcal{N}(0_d,\Sigma).$$

Proof. Let $c\in\mathbb{R}^d$ be arbitrary. It can be easily seen that the triangular scheme $(c'X_{n,k})_{k=1,\ldots,k_n}$, $n\in\mathbb{N}$, satisfies the conditions of the univariate CLT of Neumann (2011). (This theorem remains true if we have k_n rather than n summands in the n-th row of the triangular scheme. Note in particular that $k_n\not\to\infty$ implies in conjunction with the Lindeberg condition (6.23) that Σ is the zero matrix.) Therefore, we conclude from Theorem 2.1 of Neumann (2011) that

$$c'(X_{n,1}+\cdots+X_{n,k_n}) \stackrel{d}{\longrightarrow} \mathcal{N}(0,c'\Sigma c),$$

which implies the assertion by the Cramér-Wold device. □

The following bootstrap version of this CLT is an immediate consequence.

Corollary 6.1. *Suppose that for a given sample $X_{n,1},\dots,X_{n,n}$ $\mathbb{R}^d$-valued bootstrap variables $X^*_{n,1},\dots,X^*_{n,k_n}$ are available with $E^*X^*_{n,k}=0_d$ for all n,k,*

$$P\left(\sum_{k=1}^{k_n} E^*{X^*_{n,k,j}}^2 \,\le\, v_0\right) \underset{n\to\infty}{\longrightarrow} 1$$

for all $j\in\{1,\dots,d\}$ and some $v_0<\infty$. We assume that

$$\Sigma_n^* \,:=\, \operatorname{Cov}^*(X^*_{n,1}+\cdots+X^*_{n,k_n}) \stackrel{P}{\longrightarrow} \Sigma \tag{6.26}$$

for some positive semidefinite matrix Σ, and that

$$\sum_{k=1}^{k_n} E^*[{X^*_{n,k,j}}^2 \mathbb{1}(|X^*_{n,k,j}|>\epsilon)] \stackrel{P}{\longrightarrow} 0 \tag{6.27}$$

holds for all $\epsilon>0$, $j\in\{1,\dots,d\}$. Furthermore, we assume that there exists a summable sequence $(\theta_r)_{r\in\mathbb{N}}$ such that, for all $u\in\mathbb{N}$, the following upper bounds for covariances hold true: for all measurable functions $g\colon\mathbb{R}^{du}\longrightarrow\mathbb{R}$ with $\|g\|_\infty=\sup_{x\in\mathbb{R}^{du}}|g(x)|\le 1$,

$$\begin{aligned} P\Big(\left|\operatorname{cov}^*\left(g(X^*_{n,s_1},\dots,X^*_{n,s_u})X^*_{n,s_u,j_1}, X^*_{n,t_1,j_2}\right)\right| \,\le\, \left(E^*{X^*_{n,s_u,j_1}}^2 \,+\, E^*{X^*_{n,t_1,j_2}}^2 \,+\, k_n^{-1}\right)\,\theta_r \\ \forall u\in\mathbb{N},\ 1\le s_1<s_2<\cdots<s_u<s_u+r=t_1\le k_n,\ j_1,j_2\in\{1,\dots,d\}\Big) \underset{n\to\infty}{\longrightarrow} 1 \end{aligned} \tag{6.28}$$

and

$$\begin{aligned} P\Big(\left|\operatorname{cov}^*\left(g(X^*_{n,s_1},\dots,X^*_{n,s_u}), X^*_{n,t_1,j_1}X^*_{n,t_2,j_2}\right)\right| \,\le\, \left(E^*{X^*_{n,t_1,j_1}}^2 \,+\, E^*{X^*_{n,t_2,j_2}}^2 \,+\, k_n^{-1}\right)\,\theta_r \\ \forall u\in\mathbb{N},\ 1\le s_1<s_2<\cdots<s_u<s_u+r=t_1\le t_2\le k_n,\ j_1,j_2\in\{1,\dots,d\}\Big) \underset{n\to\infty}{\longrightarrow} 1. \end{aligned} \tag{6.29}$$

Then

$$X^*_{n,1} \,+\,\cdots\,+\, X^*_{n,k_n} \stackrel{d}{\longrightarrow} \mathcal{N}(0_d,\Sigma) \qquad \textit{in probability.}$$

Proof. Since the P-probability of the event that the conditions of Theorem 6.1 are satisfied by the triangular scheme $(X^*_{n,k})_{k=1,\dots,k_n}$, $n\in\mathbb{N}$, tends to one, the assertion is a direct consequence of the above theorem. □

6.3. **Some auxiliary lemmas.**

Lemma 6.1. *Suppose that assumptions (A2), (B1), and (B2) are fufilled. Then, for any fixed j,k,*

$$\frac{1}{n}\sum_{s,t=1}^{n}\left\{\Phi_j(X_s)\Phi_k(X_t)\rho(|s-t|/l_n) \,-\, E[\Phi_j(X_s)\Phi_k(X_t)\rho(|s-t|/l_n)]\right\} \stackrel{P}{\longrightarrow} 0.$$

Proof. Since we intend to compute second moments of the double sum we truncate and re-center the involved random variables $\Phi_j(X_s)$ and $\Phi_k(X_t)$ properly. We choose a sequence $(M_n)_{n\in\mathbb{N}}$ such that

$$M_n = o(\sqrt{n}),$$
$$l_n/M_n^2 = o(1)$$

and

$$P\left(|\Phi_l(X_0)| > M_n\right) \le \frac{E[\Phi_l^2(X_0)\mathbb{1}(|\Phi_l(X_0)| > M_n)]}{M_n^2} = o(n^{-1})$$

for $l = j,k$ are fufilled. (Since $P(|\Phi_l(X_0)| > \sqrt{n}/\delta) \le n^{-1}E[\Phi_l^2(X_0)\mathbb{1}(|\Phi_l(X_0)| > \sqrt{n}/\delta)]\delta^2 = o(n^{-1})$ holds for arbitrary $\delta > 0$, we can actually find a sequence $(M_n)_{n\in\mathbb{N}}$ with the above properties.)

Let $Y_{l,t} := (\Phi_l(X_t) \wedge M_n) \vee (-M_n)$ and $\bar{Y}_{l,t} = Y_{l,t} - E[Y_{l,t}]$. We split up as follows:

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{s,t=1}^n \{\Phi_j(X_s)\Phi_k(X_t)\rho(|s-t|/l_n) - E[\Phi_j(X_s)\Phi_k(X_t)\rho(|s-t|/l_n)]\}\right| \\
\le\ & \left|\frac{1}{n}\sum_{s,t=1}^n \{\Phi_j(X_s)\Phi_k(X_t) - Y_{j,s}Y_{k,t}\}\,\rho(|s-t|/l_n)\right| \\
&+ \left|\frac{1}{n}\sum_{s,t=1}^n \left\{Y_{j,s}Y_{k,t} - \bar{Y}_{j,s}\bar{Y}_{k,t}\right\}\rho(|s-t|/l_n)\right| \\
&+ \left|\frac{1}{n}\sum_{s,t=1}^n \left\{\bar{Y}_{j,s}\bar{Y}_{k,t} - E[\bar{Y}_{j,s}\bar{Y}_{k,t}]\right\}\rho(|s-t|/l_n)\right| \\
&+ \left|\frac{1}{n}\sum_{s,t=1}^n \left\{E[\bar{Y}_{j,s}\bar{Y}_{k,t}] - E[\Phi_j(X_s)\Phi_k(X_t)]\right\}\rho(|s-t|/l_n)\right| \\
=:\ & R_{n,1} + R_{n,2} + R_{n,3} + R_{n,4},
\end{aligned}$$

say.

It follows from the choice of $(M_n)_{n\in\mathbb{N}}$ that

$$\begin{aligned}
&P\left(R_{n,1} \ne 0\right) \\
&\le\ P\left(\Phi_j(X_t) \ne Y_{j,t} \text{ or } \Phi_k(X_t) \ne Y_{k,t} \quad \text{for some } t \in \{1,\ldots,n\}\right) \underset{n\to\infty}{\longrightarrow} 0. \qquad (6.30)
\end{aligned}$$

Furthermore, we obtain from

$$\begin{aligned}
|Y_{l,t} - \bar{Y}_{l,t}| &= |EY_{l,t}| = |E[Y_{l,t} - \Phi_l(X_t)]| \\
&\le E[|\Phi_l(X_t)|\mathbb{1}(|\Phi_l(X_t)| > M_n)] \\
&\le M_n^{-1}\,E[\Phi_l^2(X_t)\mathbb{1}(|\Phi_k(l_t)| > M_n)] \\
&= o(M_n^{-1})
\end{aligned}$$

that

$$
\begin{aligned}
R_{n,2} & \\
\leq \;& \frac{1}{n}\sum_{s,t=1}^{n} |Y_{j,s}-\bar{Y}_{j,s}|\,|Y_{k,t}-\bar{Y}_{k,t}|\,|\rho(|s-t|/l_n)| \\
& + \left|\frac{1}{n}\sum_{s=1}^{n}\bar{Y}_{j,s}\left\{\sum_{t=1}^{n}(Y_{k,t}-\bar{Y}_{k,t})\rho(|s-t|/l_n)\right\}\right| + \left|\frac{1}{n}\sum_{t=1}^{n}\bar{Y}_{k,t}\left\{\sum_{s=1}^{n}(Y_{j,s}-\bar{Y}_{j,s})\rho(|s-t|/l_n)\right\}\right| \\
= \;& o\left(\frac{l_n}{M_n^2}\right) + O_P\left(\frac{l_n}{\sqrt{n}\,M_n}\right). \qquad (6.31)
\end{aligned}
$$

Let A_n denote the $n\times n$-matrix with entries $(A_n)_{s,t}=\rho(|s-t|/l_n)$. We have that

$$
\begin{aligned}
& ER_{n,3}^2 \\
&= \frac{1}{n^2}\sum_{s,t,u,v=1}^{n}(A_n)_{s,t}(A_n)_{u,v}\left\{E[\bar{Y}_{j,s}\bar{Y}_{k,t}\bar{Y}_{j,u}\bar{Y}_{k,v}] - E[\bar{Y}_{j,s}\bar{Y}_{k,t}]E[\bar{Y}_{j,u}\bar{Y}_{k,v}]\right\} \\
&= \frac{1}{n^2}\sum_{s,t,u,v=1}^{n}(A_n)_{s,t}(A_n)_{u,v}\,\mathrm{cum}(\bar{Y}_{j,s},\bar{Y}_{k,t},\bar{Y}_{j,u},\bar{Y}_{k,v}) \\
&\quad + \frac{1}{n^2}\sum_{s,t,u,v=1}^{n}(A_n)_{s,t}(A_n)_{u,v}E[\bar{Y}_{j,s}\bar{Y}_{j,u}]E[\bar{Y}_{k,t}\bar{Y}_{k,v}] \\
&\quad + \frac{1}{n^2}\sum_{s,t,u,v=1}^{n}(A_n)_{s,t}(A_n)_{u,v}E[\bar{Y}_{j,s}\bar{Y}_{k,v}]E[\bar{Y}_{k,t}\bar{Y}_{j,u}], \qquad (6.32)
\end{aligned}
$$

where $\mathrm{cum}(\bar{Y}_{j,s},\bar{Y}_{k,t},\bar{Y}_{j,u},\bar{Y}_{k,v}) = \{E[\bar{Y}_{j,s}\bar{Y}_{k,t}\bar{Y}_{j,u}\bar{Y}_{k,v}] - E[\bar{Y}_{j,s}\bar{Y}_{k,t}]E[\bar{Y}_{j,u}\bar{Y}_{k,v}] - E[\bar{Y}_{j,s}\bar{Y}_{j,u}]E[\bar{Y}_{k,t}\bar{Y}_{k,v}] - E[\bar{Y}_{j,s}\bar{Y}_{k,v}]E[\bar{Y}_{k,t}\bar{Y}_{j,u}]\}$ denotes the joint cumulant of $\bar{Y}_{j,s},\bar{Y}_{k,t},\bar{Y}_{j,u},\bar{Y}_{k,v}$. It follows from Lemma 6.2 below that the first term on the right-hand side is of order $O(M_n^2 n^{-1})$. Furthermore, we have that

$$
\mathrm{cov}(\bar{Y}_{j,s},\bar{Y}_{k,t}) \;=\; O(\sqrt{\tau(|s-t|}).
$$

This implies that the second and the third term on the right-hand side of (6.32) are of order $O(l_n/n)$. Hence, we obtain

$$
ER_{n,3}^2 \;=\; O\left(M_n^2 n^{-1} + l_n n^{-1}\right) \;=\; o(1). \qquad (6.33)
$$

Let $D_{n,l} := \sqrt{E(\bar{Y}_{l,0}-\Phi_l(X_0))^2}$. We have, for $s\leq t$,

$$
\begin{aligned}
& \left|E[\bar{Y}_{j,s}\bar{Y}_{k,t}] - E[\Phi_j(X_s)\Phi_k(X_t)]\right| \\
&\leq \left|E[(\bar{Y}_{j,s}-\Phi_j(X_s))\bar{Y}_{k,t}]\right| + \left|E[\Phi_j(X_s)(\bar{Y}_{k,t}-\Phi_k(X_t))]\right| \\
&\leq D_{n,j}\sqrt{\tau(t-s)}\sqrt{\mathrm{Lip}(h)}/\sqrt{\lambda_k} + \min\{D_{n,j},\sqrt{\tau(t-s)}\sqrt{\mathrm{Lip}(h)}/\sqrt{\lambda_k}\}
\end{aligned}
$$

and for $s > t$ that

$$\begin{aligned} & \left|E[\bar{Y}_{j,s}\bar{Y}_{k,t}] \,-\, E[\Phi_j(X_s)\Phi_k(X_t)]\right| \\ & \quad \leq \quad D_{n,k}\sqrt{\tau(s-t)}\sqrt{\mathrm{Lip}(h)}/\sqrt{\lambda_j} \,+\, \min\{D_{n,k}, \sqrt{\tau(s-t)}\sqrt{\mathrm{Lip}(h)}/\sqrt{\lambda_j}\}. \end{aligned}$$

Therefore, we obtain by majorized convergence that

$$R_{n,4} \underset{n\to\infty}{\longrightarrow} 0. \tag{6.34}$$

The assertion follows now from (6.30), (6.32), (6.33) and (6.34). □

Lemma 6.2. *Suppose that the assumptions (A2) and (B1) hold. Then, for $s \leq t \leq u \leq v$,*

$$\left|\mathrm{cum}(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t}, \bar{Y}_{j_u,u}, \bar{Y}_{j_v,v})\right| \,\leq\, 3\; C\; (M_n^2 \,+\, 1)\; \sqrt{\tau(\max\{t-s, u-t, v-u\})},$$

where $C = \sqrt{\mathrm{Lip}(h)/\min\{\lambda_{j_s}, \lambda_{j_t}, \lambda_{j_u}, \lambda_{j_v}\}}$.

Proof. Let $r := \max\{t-s, u-t, v-u\}$. Recall that $\mathrm{cum}(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t}, \bar{Y}_{j_u,u}, \bar{Y}_{j_v,v}) = E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}\bar{Y}_{j_v,v}] - E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}]\; E[\bar{Y}_{j_u,u}\bar{Y}_{j_v,v}] - E[\bar{Y}_{j_s,s}\bar{Y}_{j_u,u}]\; E[\bar{Y}_{j_t,t}\bar{Y}_{j_v,v}] - E[\bar{Y}_{j_s,s}\bar{Y}_{j_v,v}]\; E[\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}]$. We distinguish between three cases, $t-s=r$, $u-t=r$ and $v-u=r$.

Case a: $t - s = r$

According to (B1), there exist random variables $\widetilde{Y}_{j_t,t}$, $\widetilde{Y}_{j_u,u}$ and $\widetilde{Y}_{j_v,v}$ such that $(\widetilde{Y}_{j_t,t}, \widetilde{Y}_{j_u,u}, \widetilde{Y}_{j_v,v})'$ has the same distribution as $(\bar{Y}_{j_t,t}, \bar{Y}_{j_u,u}, \bar{Y}_{j_v,v})'$ and is independent of $\bar{Y}_{j_s,s}$, and $\sqrt{E(\bar{Y}_{j_w,w} - \widetilde{Y}_{j_w,w})^2} \leq C\; \sqrt{\tau(r)}$, for $w \in \{t,u,v\}$.

Now we have

$$\begin{aligned} & |E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}\bar{Y}_{j_v,v}]| \\ & \quad = \quad |E[\bar{Y}_{j_s,s}(\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}\bar{Y}_{j_v,v} \,-\, \widetilde{Y}_{j_t,t}\widetilde{Y}_{j_u,u}\widetilde{Y}_{j_v,v})]| \\ & \quad \leq \quad \sqrt{E\bar{Y}_{j_s,s}^2}\; \left\{\sqrt{E(\bar{Y}_{j_t,t} - \widetilde{Y}_{j_t,t})^2} \,+\, \sqrt{E(\bar{Y}_{j_u,u} - \widetilde{Y}_{j_u,u})^2} \,+\, \sqrt{E(\bar{Y}_{j_v,v} - \widetilde{Y}_{j_v,v})^2}\right\}\; M_n^2 \\ & \quad \leq \quad 3\; C\; \sqrt{\tau(r)}\; M_n^2. \end{aligned} \tag{6.35}$$

Futhermore, we have for $w \in \{t,u,v\}$ that

$$|E\bar{Y}_{j_s,s}\bar{Y}_{j_w,w}| \,\leq\, |E[\bar{Y}_{j_s,s}(\bar{Y}_{j_w,w} - \widetilde{Y}_{j_w,w})]| \,\leq\, \sqrt{E\bar{Y}_{j_s,s}^2}\sqrt{E(\bar{Y}_{j_w,w} - \widetilde{Y}_{j_w,w})^2} \,\leq\, C\; \sqrt{\tau(r)}, \tag{6.36}$$

which implies that

$$|\,\mathrm{cum}(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t}, \bar{Y}_{j_u,u}, \bar{Y}_{j_v,v})| \,\leq\, 3\; C\; \sqrt{\tau(r)}\; (M_n^2 \,+\, 1). \tag{6.37}$$

Case b: $u - t = r$

Here we choose random variables $\widetilde{Y}_{j_u,u}$ and $\widetilde{Y}_{j_v,v}$ such that $(\widetilde{Y}_{j_u,u}, \widetilde{Y}_{j_v,v})'$ has the same distribution as $(\bar{Y}_{j_u,u}, \bar{Y}_{j_v,v})'$, is independent of $(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t})'$, and $\sqrt{E(\bar{Y}_{j_w,w} - \widetilde{Y}_{j_w,w})^2} \leq C\sqrt{\tau(r)}$ for $w \in \{u, v\}$.

Then

$$\begin{aligned}
&|E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}\bar{Y}_{j_v,v}] - E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}]\, E[\bar{Y}_{j_u,u}\bar{Y}_{j_v,v}]| \\
&\quad = |E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}(\bar{Y}_{j_u,u}\bar{Y}_{j_v,v} - \widetilde{Y}_{j_u,u}\widetilde{Y}_{j_v,v})]| \\
&\quad \leq |E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}(\bar{Y}_{j_u,u} - \widetilde{Y}_{j_u,u})\widetilde{Y}_{j_v,v}]| + |E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}(\bar{Y}_{j_v,v} - \widetilde{Y}_{j_v,v})]| \\
&\quad \leq \sqrt{E\bar{Y}_{j_s,s}^2}\left\{\sqrt{E(\bar{Y}_{j_u,u} - \widetilde{Y}_{j_u,u})^2} + \sqrt{E(\bar{Y}_{j_v,v} - \widetilde{Y}_{j_v,v})^2}\right\} M_n^2 \\
&\quad \leq 2\,C\,\sqrt{\tau(r)}\,M_n^2.
\end{aligned}$$

Furthermore, we obtain analogously to (6.36)

$$|E\bar{Y}_{j_s,s}\bar{Y}_{j_u,u}|, |E\bar{Y}_{j_s,s}\bar{Y}_{j_v,v}|, |E\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}|, |E\bar{Y}_{j_t,t}\bar{Y}_{j_v,v}| \leq C\,\sqrt{\tau(r)},$$

which yields that

$$|\operatorname{cum}(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t}, \bar{Y}_{j_u,u}, \bar{Y}_{j_v,v})| \leq 2\,C\,\sqrt{\tau(r)}\,(M_n^2 + 1). \tag{6.38}$$

Case c: $v - u = r$

In this case we choose a random variable $\widetilde{Y}_{j_v,v}$ with the same distribution as $\bar{Y}_{j_v,v}$ and independent of $(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t}, \bar{Y}_{j_u,u})'$ such that $\sqrt{E(\bar{Y}_{j_v,v} - \widetilde{Y}_{j_v,v})^2} \leq C\sqrt{\tau(r)}$. We have

$$\begin{aligned}
|E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}\bar{Y}_{j_v,v}]| &= |E[\bar{Y}_{j_s,s}\bar{Y}_{j_t,t}\bar{Y}_{j_u,u}(\bar{Y}_{j_v,v} - \widetilde{Y}_{j_v,v})]| \\
&= C\,\sqrt{\tau(r)}\,\sqrt{E\bar{Y}_{j_s,s}^2}\,M_n^2.
\end{aligned} \tag{6.39}$$

Furthermore, we have for $w \in \{s, t, u\}$ that $|E\bar{Y}_{j_w,w}\bar{Y}_{j_v,v}| \leq C\sqrt{\tau(r)}$, which yields that

$$|\operatorname{cum}(\bar{Y}_{j_s,s}, \bar{Y}_{j_t,t}, \bar{Y}_{j_u,u}, \bar{Y}_{j_v,v})| \leq C\,\sqrt{\tau(r)}\,(M_n^2 + 3). \tag{6.40}$$

The assertion of the lemma follows from (6.37), (6.38) and (6.40). □

Lemma 6.3. *Suppose that the prerequisites of Theorem 3.1 are satisfied. Then condition (6.28) of Corollary 6.1 holds true.*

Proof. Let $\widetilde{X}^*_{n,t_1}$ be a copy of X^*_{n,t_1} that is independent of $(W^*_k)_{k\leq L_n s_u}$. Then

$$P\Big(\big|cov^*\left(g(X^*_{n,s_1},\ldots,X^*_{n,s_u})X^*_{n,s_u,j},X^*_{n,t_1,k}\right)\big| \leq \Big(E^*\left[X^*_{n,s_u,j}\right]^2 + E^*\left[Y^*_{n,t_1,k}\right]^2 + L_n/n\Big)\,\theta_r,$$
$$\forall\, u\in\mathbb{N}, 1\leq s_1<\cdots<s_u<s_u+r=t_1\leq n/L_n\Big)$$
$$\geq P\Big(\sqrt{E^*\left|X^*_{n,s,j}\right|^2 E^*\left[X^*_{n,t,k}-\widetilde{X}^*_{n,t,k}\right]^2} \leq \theta_r\, L_n/n, \quad \forall\, 1\leq s<s+r=t\leq n/L_n\Big)$$

The latter probability tends to one for a suitable choice of the sequence $(\theta_r)_r$ if

$$E^*\left[X^*_{n,s,k}\right]^2 = EE^*\left[X^*_{n,s,k}\right]^2 + R_{n,s,k} \quad \text{with} \sum_{s=1}^{k_n}|R_{n,s,k}| = o_P(1) \quad \forall\, k \tag{6.41}$$

and

$$P\Big(\sqrt{E^*\left[X^*_{n,t,k}-\widetilde{X}^*_{n,t,k}\right]^2} \leq C\theta_r\, L_n/n \quad \forall\, r<t\leq n/L_n\Big) \underset{n\to\infty}{\longrightarrow} 1 \tag{6.42}$$

hold for some $C<\infty$.

The verification of (6.41) follows the lines of the proof of Lemma 6.1. We decompose

$$\Big|E^*\left[X^*_{n,s,k}\right]^2 - EE^*\left[X^*_{n,s,k}\right]^2\Big| = \frac{1}{n}\sum_{u,v\in I_{n,s}}\left[\Phi_k(X_u)\Phi_k(X_v) - E\Phi_k(X_u)\Phi_k(X_v)\right]\rho(|u-v|/l_n)$$

into summands $R^{(s)}_{n,1}$ to $R^{(s)}_{n,4}$, the counterparts of $R_{n,1}$ to $R_{n,4}$ with $\sum_{u,v\in I_{n,s}}$ instead of $\sum_{s,t=1}^n$, and show that $\sum_{s=1}^{k_n} R^{(s)}_{n,i} = o_P(1),\ i=1,\ldots,4$. We obtain

$$\sum_{s=1}^{k_n} R^{(s)}_{n,1} = o_P(1) \tag{6.43}$$

from (6.30). In analogy to (6.31),

$$\sum_{s=1}^{k_n} R^{(s)}_{n,2} \;=\; o\left(\frac{l_n}{M_n^2}\right) \;+\; O_P\left(\frac{l_n}{\sqrt{L_n}M_n}\right) \;=\; o_P(1) \tag{6.44}$$

if we choose $L_n = o(n)$ such that the latter terms vanish asymptotically as $n\to\infty$. In view of Lemma 6.2, we have $E\left[R^{(s)}_{n,3}\right]^2 = O(M_n^2 L_n/n^2 + l_n L_n/n^2)$ for all s, which in turn implies

$$\sum_{s=1}^{k_n} R^{(s)}_{n,3} \;=\; O_P\left(\frac{M_n}{\sqrt{L_n}} + \sqrt{\frac{l_n}{L_n}}\right) \;=\; o_P(1) \tag{6.45}$$

if L_n tends to infinity sufficiently fast. Finally, we get $\sum_{s=1}^{k_n} R^{(s)}_{n,4} = o_P(1)$ in a similar manner as in the proof of Lemma 6.1. In conjunction with (6.43) to (6.45), we obtain (a).

To show (6.42), we prove that, on the one hand,

$$EE^*\left[X^*_{n,t,k}-\widetilde{X}^*_{n,t,k}\right]^2 \leq C\zeta^{\delta r}\,\frac{L_n}{n} \tag{6.46}$$

for some $C < \infty$ uniformly in t and, on the other hand,

$$\sum_{t=r+1}^{k_n} \left| E^* \left[X^*_{n,t,k} - \widetilde{X}^*_{n,t,k} \right]^2 - EE^* \left[X^*_{n,t,k} - \widetilde{X}^*_{n,t,k} \right]^2 \right| \stackrel{P}{\longrightarrow} 0. \tag{6.47}$$

We specialize $\widetilde{X}_{n,t,k} = \frac{1}{n} \sum_{u \in I_{n,t}} \Phi_k(X_u) \widetilde{W}^*_u$, where $(\widetilde{W}^*_u)_{u \in I_{n,t}}$ is an copy of $(W^*_u)_{u \in I_{n,t}}$ that is independent of $(W^*_k)_{k \leq t-r}$ and of $(X_k)_{k=1}^n$ and such that $E^* |\widetilde{W}^*_u - W^*_u| \leq \zeta^r$ for all $u \in I_{n,t}$. (This procedure might require an enlargement of the underlying probability space.) Then

$$EE^* \left[X^*_{n,t,k} - \widetilde{X}^*_{n,t,k} \right]^2 \leq \frac{1}{L_n} \sum_{u,v=1}^{L_n} \left| E(\Phi_k(X_u)\Phi_k(X_v)) E^*([W^*_u - \widetilde{W}^*_u][W^*_v - \widetilde{W}^*_v]) \right|$$

W.l.o.g. we only take the case $u < v$ into further consideration and obtain (6.46) from

$$\begin{aligned}
&\sum_{t=r+1}^{k_n} \frac{1}{n} \sum_{\substack{u,v \in I_{n,t} \\ u<v}} \left| E(\Phi_k(X_u)\Phi_k(X_v)) E^*([W^*_u - \widetilde{W}^*_u][W^*_v - \widetilde{W}^*_v]) \right| \\
&= \sum_{t=r+1}^{k_n} \frac{1}{n} \sum_{\substack{u,v \in I_{n,t} \\ u<v}} \left| E(\Phi_k(X_u)[\Phi_k(X_v) - \Phi_k(\widetilde{X}_v)]) E^*([W^*_u - \widetilde{W}^*_u][W^*_v - \widetilde{W}^*_v]) \right| \\
&\leq \sqrt{\frac{Lip(h)}{\lambda_k}} \sum_{t=r+1}^{k_n} \frac{1}{n} \sum_{\substack{u,v \in I_{n,t} \\ u<v}} \tau(v-u) \left(E^*[W^*_u - \widetilde{W}^*_u]^{1/(1-\delta)} [W^*_v - \widetilde{W}^*_v] \right)^{1-\delta} \zeta^{\delta r L_n / l_n} \\
&\leq C\, \zeta^{\delta r} \frac{L_n}{n}.
\end{aligned}$$

Once again, we follow the lines of the proof of Lemma 6.1 in order to verify (6.47). For each t, we decompose the term within the absolute value sign into $\bar{R}^{(t)}_{n,1}, \ldots, \bar{R}^{(t)}_{n,4}$, where we substitute $\rho(|u-v|/l_n)$ by $R_{n,u,v} = E^*([W^*_u - \widetilde{W}^*_u][W^*_v - \widetilde{W}^*_v])$ in the definition of $R^{(t)}_{n,1}, \ldots, R^{(t)}_{n,4}$. We obtain

$$\sum_{t=1}^{k_n} \bar{R}^{(t)}_{n,1} = o_P(1) \tag{6.48}$$

from (6.30). Moreover,

$$\sum_{t=r+1}^{k_n} \bar{R}^{(t)}_{n,2} = o(1) \frac{L_n}{M_n^2} \zeta^{\delta r L_n / l_n} + O_P \left(\frac{\sqrt{L_n}}{M_n} \zeta^{\delta r L_n / l_n} \right) = o_P(1). \tag{6.49}$$

For the estimation of $\sum_{t=r+1}^{k_n} \bar{R}^{(t)}_{n,3}$ we substitute $(A_n)_{u,v}$ by $R_{n,u,v}$ and n by L_n in part (ii) of the proof of Lemma 6.1. This leads to

$$\sum_{t=r+1}^{k_n} \bar{R}^{(t)}_{n,3} = O_P \left(\frac{M_n}{\sqrt{L_n}} + \frac{n}{L_n} \zeta^{\delta r L_n / l_n} \right) = o_P(1). \tag{6.50}$$

Finally, we get $\sum_{t=r+1}^{k_n} \bar{R}^{(t)}_{n,4} = o_P(1)$ and we can deduce (6.47) from the latter result in conjunction with (6.48), (6.49), and (6.50). □

Acknowledgment . This research was funded by the German Research Foundation DFG, project NE 606/2-2.

References

Alba-Fernández, V., Jiménez-Gamero, M. D., and Muñoz-García, J. (2008). A test for the two-sample problem based on the empirical characteristic function. *Computational Statistics and Data Analysis* **52**, 3730–3748.

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *J. Amer. Statist. Assoc.* **49**, 765–769.

Anderson, N. H., Hall, P., and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Mult. Analysis* **50**, 41–54.

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

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

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Borisov, I. S. and Volodko, N. V. (2008). Orthogonal series and limit theorems for canonical U- and V-statistics of stationary connected observations. *Siberian Advances in Mathematics* **18**, 242–257.

Carlstein, E. (1988). Degenerate U-statistics based on non-independent observations. *Calcutta Statist. Assoc. Bull.* **37**, 55–65.

Dehling, H., Olivier, D., and Volny, D. (2009). New techniques for empirical processes of dependent data. *Stochastic Process. Appl.* **119**, 3699–3718.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.* **51**, 392–413.

Dedecker, J. and Prieur, C. (2004). Couplage pour la distance minimale. *C. R. Acad. Sci. Paris, Ser. I* **338**, 805–808.

Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Relat. Fields* **132**, 203–236.

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.

Eagleson, G. K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.* **21**, 221–237.

Fan, Y. and Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *J. Nonparametr. Statist.* **11**, 337–360.

Ghosh, B. K. and Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Ann. Statist.* **19**, 999–1009.

Gleser, L. J. and Moore D. S. (1983). The effect of dependence on chi-squared and empiric distribution tests of fit. *Ann. Statist.* **11**, 1100–1108.

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5**, 110–123.

Huang, W. and Zhang, L.-X. (2006). Asymptotic normality for U-statistics of negatively associated random variables. *Statist. Probab. Lett.* **76**, 1125–1131.

Ignaccolo, R. (2004). Goodness-of-fit tests for dependent data. *J. Nonparametr. Statist.* **16**, 19–38.

Leucht, A. (2010). Characteristic function-based goodness-of-fit tests under weak dependence. *Reports of the Dept. Mathematics and Computer Science, Friedrich Schiller University Jena* **10-06**.

Leucht, A. (2011). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli*, forthcoming.

Leucht, A. and Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate U- and V-type statistics. *J. Multivariate Anal.* **100**, 1622–1633.

Leucht, A. and Neumann, M. H. (2011). Degenerate U- and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Reports of the Dept. Mathematics and Computer Science, Friedrich Schiller University Jena* **11-01**.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews* **3**, 261 -274.

Lütkepohl, H. (1996). *Handbook of Matrices.* Chichester: John Wiley & Sons.

Munk, A., Stockis, J.-P., Valeinis J., and Giese, G. (2011). Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Ann. Inst. Stat. Math.* **63**, 639–659.

Neumann, M. H. (2011). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, doi: 10.1051/ps/2011144, Published online by Cambridge University Press: 25 November 2011.

Neumann, M. H. and Paparoditis, E. (2000). On bootstrapping L_2-type statistics in density testing. *Statist. Probab. Lett.* **50**, 137–147.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap, *J. Amer. Statist. Assoc.* **89**, 1303–1313.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3**, 1–14.

Shao, X. (2010). The dependent wild bootstrap. *J. Amer. Statist. Assoc.* **105**, 218–235.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35**, 1773-1801.

Sun, H. (2005). Mercer theorem for RKHS on noncompact sets. *J. Complexity* **214**, 337–349.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261–1350.