Dependent wild bootstrap for degenerate $U$- and $V$-statistics

Anne Leucht
Universität Mannheim
Abteilung Volkswirtschaftslehre
L7, 3-5
D-68131 Mannheim
Germany
E-mail: leucht@uni-mannheim.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 determine critical values for these 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 was reasonably close to the nominal one, we obtained 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 $U$- and $V$-statistics.
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 $\chi^2$-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 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 (2013) 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 usual degeneracy and under easily verifiable conditions imposed directly on the kernel function.

For bootstrapping degenerate 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 (2013). 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 $H(X_n) - H(E X_1)$ by that of $H\left(n^{-1} \sum_{t=1}^{n} X_t \right) - H\left(\bar{X}_n \right)$, where $\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t$. The bootstrap approach for independent random variables is easily explained: While a nondegenerate distribution of $W_{t,n}^*$ with $E^* W_{t,n}^* = 0$ introduces the necessary randomness, the condition of $\text{cov}^*(W_{t,n}^*, W_{t,n}^*) \to_{n \to \infty} 1$ 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 $V_n = n^{-1} \sum_{t=1}^{n} h(X_s, X_t)$ 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 $V_n = \sum_k \lambda_k \left(n^{-1/2} \sum_{t=1}^{n} \Phi_k(X_t)^2 \right)$, where $(\lambda_k)_k$ are the nonzero eigenvalues and $(\Phi_k)_k$ the corresponding eigenfunctions of a certain integral equation. 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 $V_n$ results 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 $V_n^* = \sum_k \lambda_k \left(n^{-1/2} \sum_{t=1}^{n} \Phi_k(X_t) W_{t,n}^* \right)^2 = n^{-1} \sum_{s,t=1}^{n} h(X_s, X_t) W_{s,n}^* W_{t,n}^*$. A similar method, based on independent auxiliary variables $(W_t^*_{t=1,\ldots,n})$, was proposed in Dehling and Mikosch (1994) for degenerate U-statistics of independent random variables. We prove that the distribution of...
$V^*_n$ consistently approximates that of $V_n$. 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. Besides the fact that our resampling method is very easy to implement, it is also computationally less intensive than ordinary block bootstrap algorithms which cannot be applied naïvely to $U$-statistics but require a modification of the kernel on the bootstrap side; see also Remark 4 in Section 3 for details.

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 distributions of two matched 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 too often if the quantities of the i.i.d. setting are used as an approximation for the corresponding quantities under positive dependence. On the other hand, 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. We additionally included a real-data application of the two-sample test introduced in Section 4.2.

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 2013] 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 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

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

where $h: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a symmetric function that is degenerate, i.e. $E h(X_0, y) = 0 \ \forall y \in \text{supp}(P^{X_0}) := \{ x \in \mathbb{R}^d \mid \text{open } O: x \in O \Rightarrow P^{X_0}(O) > 0 \}$. To impose a restriction on the dependence of the underlying process $(X_t)_{t \in \mathbb{Z}}$, we invoke the concept of $\tau$-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 $\tau$-(weakly) dependent if

$$\tau(r) = \sup_{t \in \mathbb{N}} \frac{1}{\tau \leq t_1 < \cdots < t_\ell} \{ \tau(\sigma(X_{t_1}, t \leq 0), (X_{t_1}, \ldots, X_{t_\ell})) \} \to 0,$$
where  
\[ \tau(\mathcal{M}, X) = E \left( \sup_{f \in \Lambda_1(\mathbb{R}^p)} \left| \int_{\mathbb{R}^p} f(x) dp^{X|M}(x) - \int_{\mathbb{R}^p} f(x) dp^X(x) \right| \right). \]

Here, $\mathcal{M}$ is a sub-$\sigma$-algebra of $\mathcal{A}$, $p^{X|M}$ 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 a process $(X_t)_{t \in \mathbb{Z}}$ on $(\Omega, \mathcal{A}, P)$ is $\tau$-dependent and if $\mathcal{A}$ is rich enough, then there exists, for all $t, t_1, \ldots, t_l \in \mathbb{Z}$ with $t < t_1 < \cdots < t_l$, $l \in \mathbb{N}$, a random vector $(\tilde{X}_{t_1}, \ldots, \tilde{X}_{t_l})'$ that is independent of $(X_s)_{s \leq t}$, has the same distribution as $(X_{t_1}, \ldots, X_{t_l})'$ and satisfies

\[ \frac{1}{l} \sum_{j=1}^l E\|\tilde{X}_{t_j} - X_{t_j}\|_1 \leq \tau(t_1 - t). \] (2.1)

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

We assume

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

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

(A2) (i) $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is symmetric and degenerate, i.e. $h(x, y) = h(y, x) \forall x, y \in \mathbb{R}^d$ and $E h(X_0, y) = 0$ for all $y \in \text{supp}(p^{X_0})$.

(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 $m \in \mathbb{N}$, $\sum_{j=1}^m c_j h(x_i, x_j) \geq 0$.

(iii) $E h(X_0, X_0) < \infty$.

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

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

\[ V_n \overset{d}{\to} Z := \sum_{k} \lambda_k Z_k^2 \quad \text{and} \quad U_n \overset{d}{\to} Z - E h(X_0, X_0), \]

where $(Z_k)_k$ is a sequence of centered, jointly normal random variables with $\text{cov}(Z_j, Z_k) = \sum_{r=-\infty}^{\infty} \text{cov}(\Phi_j(X_0), \Phi_k(X_r))$, and $(\lambda_k)_k$ and $(\Phi_k)_k$ are the sequences of non-zero eigenvalues and the corresponding orthonormal eigenfunctions of $E[h(x, X_0) \Phi(X_0)] = \lambda \Phi(x)$. It holds that $EZ = \sum_{r \in \mathbb{Z}} E h(X_0, X_r) < \infty$, i.e., the infinite series defining $Z$ converges in $L_1$ and, since all summands are non-negative, also almost surely.

There are already several papers on the asymptotic distribution 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. Moreover, the representation of the limit variables was more complicated than in the present paper. Especially for cases where results on the weak convergence of a weighted empirical process are already at hand, Beutner and Zähle (2012) devised a unified approach based on the continuous mapping theorem to derive the asymptotics of both degenerate and non-degenerate \( U \) - and \( V \)-statistics. In the degenerate case and for real-valued random variables \( X_t \), they obtained a representation of the limit as a double stochastic integral. A spectral decomposition of the kernel was also employed by Leucht and Neumann (2013) 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 \( f: \mathbb{R}^{4d} \to \mathbb{R} \), 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

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

for some \( \delta \in (0,1) \) with \( \sum_{r=1}^{\infty} |\tau(r)|^{\delta/2} < \infty \). Here the supremum is taken over all random variables \( Y_1,\ldots,Y_5 \) with arbitrary joint distribution such that the marginal distribution is \( P_{X_0} \).

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. Some knowledge of this limit is necessary when critical values have to be determined. It depends on the eigenvalues of the equation \( E[h(x,X_0)\Phi(X_0)] = \lambda \Phi(x) \) that are typically unknown in applications and, more seriously, it depends on the whole dependence structure of the underlying process via sums of covariances. 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 appropriate 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 (2013), we 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 was to construct the pseudo-observations as

\[
X^*_t = \bar{X}_n + (X_t - \bar{X}_n)W^*_t, \quad t = 1,\ldots,n.
\]

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 \( X_1,\ldots,X_n \). Shao (2010) verified that under certain regularity conditions

\[
\sup_{x \in \mathbb{R}} \left| P \left( \sqrt{n} \left[ H(\bar{X}_n) - H(Ex_1) \right] \leq x \right) - P^* \left( \sqrt{n} \left[ H \left( \frac{1}{n} \sum_{t=1}^{n} X^*_t \right) - H(\bar{X}_n) \right] \leq x \right) \right| \xrightarrow{P} 0,
\]

where \( H \) is a smooth function. Moreover, if additionally \( EX_1 = 0 \), then \( X^*_t = X_tW^*_t \) would also lead to a consistent bootstrap approximation. Motivated by this and by the representation
$V_n = \sum_k \lambda_k (n^{-1/2} \sum_{i=1}^n \Phi_k(X_i))^2$ of the $V$-statistic, a natural bootstrap counterpart is given by 
$\sum_k \lambda_k (n^{-1/2} \sum_{i=1}^n \Phi_k(X_i)W^*_s)^2 = n^{-1} \sum_{s,t=1}^n W^*_s h(X_s, X_t) W^*_t$. More precisely, our first proposal is

$$U^*_{n,1} = \frac{1}{n} \sum_{s,t=1, s \neq t}^n W^*_s h(X_s, X_t) W^*_t \quad \text{and} \quad V^*_{n,1} = \frac{1}{n} \sum_{s,t=1}^n W^*_s h(X_s, X_t) W^*_t.$$ 

Note that in contrast to Shao (2010), we do not generate a bootstrap counterpart of the observations themselves, but of \( h(X_s, X_t), s, t = 1, \ldots, n \). We show that the statistics above 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 \( n \) converge to some positive constant while its bootstrap version, also divided by \( n \), is of order \( o_P(1) \). This implies consistency of the test. However, we have learned from our simulations reported in Section 4 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 \( W^*_n \) has zero mean. On the other hand, this effect is mitigated by the fact that condition (B2) implies \( \text{cov}^*(W^*_s, W^*_t) \rightarrow_{n \rightarrow \infty} 1 \). Therefore, we propose the following versions that are modified by an empirical degeneration:

$$U^*_{n,2} = \frac{1}{n} \sum_{s,t=1, s \neq t}^n W^*_s \bar{h}(X_s, X_t) W^*_t \quad \text{and} \quad V^*_{n,2} = \frac{1}{n} \sum_{s,t=1}^n W^*_s \bar{h}(X_s, X_t) W^*_t,$$

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 \( \bar{W}_n = n^{-1} \sum_{t=1}^n W^*_t \).

Remark 1. As a referee of this paper has remarked, the idea of an external randomization is anything else than new in the statistics literature. The classical example are randomized tests where an independent random source is used to obtain an exact size \( \alpha \) test. Another application of an external randomization is the construction of distribution-free tests of composite hypotheses which are again exact size \( \alpha \) tests. Durbin (1961, Section 7) considered the problem of testing normality of a sample \( X_1, \ldots, X_n \) where ignorance of the common mean \( \mu \) and the common variance \( \sigma^2 \) leads to a composite null hypothesis. Using the fact that the conditional distribution of \( X = (X_1, \ldots, X_n)' \) given the sufficient statistic \( T = (\bar{X}_n, S_n)' \), where \( S_n^2 = (n - 1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \), does not depend on \( \theta = (\mu, \sigma^2)' \), he proposed to draw first independently of \( X \) an auxiliary random variable \( T^* = (\bar{X}_n^*, S_n^2)' \) possessing the same distribution as \( T \) in the case of \( X_1, \ldots, X_n \sim \mathcal{N}(0, 1) \), and to define then new random quantities \( X_i^* = X_i + S_n^*(\bar{X}_n - X_i)/S_n \), \( i = 1, \ldots, n \). It turns out that \( X_1^*, \ldots, X_n^* \) are independent \( \mathcal{N}(0, 1) \) variables under the (composite) null hypothesis of normality. Therefore, we can test composite hypotheses concerning \( X_1, \ldots, X_n \) as simple hypotheses concerning \( X_1^*, \ldots, X_n^* \) and any goodness-of-fit statistic such as the Cramér-von Mises or the Kolmogorov-Smirnov statistic is now distribution-free and can therefore be used for an exact size \( \alpha \) test. A further discussion of this approach and additional examples and references are given in Pesarin (1984). A point of criticism on the above-mentioned procedures is that the outcome of these tests depends directly on the external randomization. That is, if these procedures are repeated twice using the same data but with independent randomizing variables, it might happen that the null hypothesis is rejected first but accepted in the second experiment and vice versa.

In our context, however, the purpose of the randomization is a different one. Although Theorem 23 gives a complete picture of the asymptotic behavior of degenerate $U$- and $V$-statistics, the limit distribution is nevertheless not completely known since the involved eigenvalues \( (\lambda_k)_k \) as well as the dependence structure of the underlying process are unknown in almost all applications. In
contrast to the papers mentioned above, we do not attempt to transform our test statistic with the aid of an auxiliary randomization to a new statistic with a known distribution. Rather, we intend to mimic the behavior of the statistic under the null by our dependent wild bootstrap. Representing the V-statistic as $V_n = \sum_k \lambda_k (n^{-1/2} \sum_{t=1}^n \Phi_k(X_t))^2$ we see that joint asymptotic normality of the quantities $\gamma_{n,k} = n^{-1/2} \sum_{t=1}^n \Phi_k(X_t)$ is actually the cause for the limit distribution of $V_n$. The external randomization with $W_1^*, \ldots, W_n^*$ in our DWB version $Y_{n,k}^* = n^{-1/2} \sum_{t=1}^n \Phi_k(X_t)W_t^*$ of $Y_{n,k}$ provides just some randomness that makes, in conjunction with $\text{cov}(W_t^*, W_t^*) = \rho(|s-t|/l_n)$, $l_n/n \to n \to \infty$ 0, the conditional distribution of $Y_{n,k}^*$ given $X_1, \ldots, X_n$ asymptotically normal. Moreover, the condition $l_n \to n \to \infty$ ensures that the covariance structure of $\Phi_k(X_1), \ldots, \Phi_k(X_n)$ is asymptotically correctly replicated by that of $\Phi_k(X_1)W_1^*, \ldots, \Phi_k(X_n)W_n^*$. This is the basic reason why $V_n$ can be bootstrapped by $V_{n,1} = \sum_k \lambda_k (n^{-1/2} \sum_{t=1}^n \Phi_k(X_t)W_t^*)^2$. Finally, note that in sharp contrast to the statistical procedures mentioned above, the outcome of a bootstrap-based test does not depend on the random outcome of the external randomization; it only depends on the original random variables $X_1, \ldots, X_n$. This is because we can generate, at least theoretically, an arbitrary number of bootstrap loops which means that we can approximate the bootstrap distribution as good as we wish.

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, $\tau$-dependent process with $\sum_{r=1}^{\infty} r^2 \sqrt{\tau_r} < \infty$.

Regarding the variables $(W_t^*)$, 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 $\text{cov}(W_t^*, W_t^*) = \rho(|s-t|/l_n)$, where $\rho(u) \to u \to 0 1$, $\sum_{r=1}^{n-1} \rho(|r|/l_n) = O(l_n)$ with $l_n \to n \to \infty$ and $l_n = o(n)$. The variables $(W_{t,n}^*)_{t=1}^n$ are $\tau$-weakly dependent with coefficients $\tau^*(r) \leq K \zeta r/l_n$ for $r = 1, \ldots, n$, some $\zeta \in (0, 1)$ and a $K < \infty$.

Remark 2.

(i) 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, we could use an Ornstein-Uhlenbeck process, i.e., a Gaussian process with continuous sample paths, $EU_t = 0$ and $\text{cov}(U_s, U_t) = \exp(-|s-t|)$. Then the practical implementation becomes easy since a discrete sample of an Ornstein-Uhlenbeck process forms an AR(1) process, i.e., $W_{t,n}^* = e^{-l_n/2} W_{t-1,n}^* + \sqrt{1 - e^{-2/l_n}} \varepsilon_t^*$, where $W_{0,n}^*, \varepsilon_1^*, \ldots, \varepsilon_n^*$ are independent standard normal variables. This in turn implies that $W_{1,n}^*, \ldots, W_{n,n}^*$ are standard normally distributed, too. Further admissible choices of the process $(W_{t,n}^*)_{t=1}^n$ can be found in Section 2.2 of Shao (2010). There, these processes are chosen to be $l_n$-dependent.

(ii) The role of the parameter $l_n$ is similar to that of the block length in blockwise bootstrap methods. While Shao (2010) Theorem 3.1) imposed the condition $l_n^{-1} + l_n^{-\delta/(2+2\delta)} \to n \to \infty 0$, for some $\delta > 2$, we relaxed this to the obviously necessary condition $l_n^{-1} + l_n^{-1} \to n \to \infty 0$. This suggests that the statistician has much freedom of choosing $l_n$. Simulations presented in Shao (2010) indicate that the choice of $l_n$ is not very critical, at least if the MSE of a bootstrap variance estimator or the coverage probability of a confidence interval for the mean are concerned. A more detailed investigation of this issue is certainly of interest, however, we think that this is well beyond the scope of this paper. In our simulations reported in Section 5 below we also experienced that the choice of $l_n$ does not influence too much the performance of the tests.
Lemma 3.1. Suppose that (A1) and (A2) hold true. In the following lemma, the first variant of our bootstrap proposal is at least computationally less intensive. 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.

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

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

If additionally the limiting distribution function is continuous, then

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

Remark 3. 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. \( \sum_{i=1}^{\infty} r^2 (\tau(r))^{\delta/2} < \infty \) for some \( \delta \in (0, 1) \).

A necessary and sufficient condition for 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 \( EZ = \sum_{r \in \mathbb{Z}} \text{Eh}(X_0, X_r) \neq 0 \).

The latter condition in this lemma simplifies to \( \text{Eh}(X_0, X_0) > 0 \) in all the applications considered by Leucht and Neumann (2013).

Remark 4. 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) 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 consider the test problem

\[ H_0: \quad F = F_0 \quad \text{vs.} \quad H_1: \quad F \neq F_0. \]

Below we discuss a test of generalized Cramér-von Mises type based on the test statistic

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

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 $\chi^2$-test and Neyman’s smooth test with a fixed number of components for $\alpha$-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 especially all alternatives. It is clear that our 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 $\beta$-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 for detecting local alternatives characterized by densities with sharp peaks, they suffer from a loss of power against classical 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 $h(x, y) = \int [\mathbb{I}_{x \leq z} - F_0(z)] [\mathbb{I}_{y \leq z} - F_0(z)] w(z) \lambda^d(dz)$ ($x \leq y$ 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 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,

$$T_{n,1}^* = \frac{1}{n} \sum_{s,t=1}^{n} h(X_s, X_t) W^*_s W^*_t$$

resp.

$$T_{n,2}^* = \frac{1}{n} \sum_{s,t=1}^{n} h(X_s, X_t) (W^*_s - \overline{W_n^*}) (W^*_t - \overline{W_n^*}).$$

3. Independently, 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 $H_0$ if $T_n > t^*_{\alpha, i}$.

In view of Theorem 3.1

$$P(T_n > t^*_{\alpha, i}) \rightarrow \alpha$$

under $H_0$ as the sample size increases if (B1) is satisfied and if $\sum_{r=-\infty}^{\infty} \mathbb{E} h(X_0, X_r) \neq 0$. The next proposition characterizes consistency of the test under the alternative.

**Proposition 4.1.**

(i) If (A1) is fulfilled, then

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

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

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

$$n^{-1} T_{n,1}^* \xrightarrow{P^*} 0 \quad \text{in probability}$$
and
\[ P\left(P^* \left(T_{n,2}^* > K(\epsilon) \right) < \epsilon \right) \xrightarrow{n \to \infty} 1 \]
for all \(\epsilon > 0\) and appropriate \(K(\epsilon) < \infty\).

(iii) If (B1) and (B2) are fulfilled, \(F \neq F_0\), and if \(w\) is positive almost everywhere w.r.t. Lebesgue measure, then
\[ P\left(T_n > t_{n,i}^* \right) \xrightarrow{n \to \infty} 1 \quad (i = 1, 2). \]

Remark 5. For test problems with composite null hypotheses depending on a finite-dimensional parameter, asymptotic theory for \(U\)- and \(V\)-statistics with estimated parameters is required. Typically, the effect of estimating some parameter on a statistic of Cramér-von Mises-type is asymptotically not negligible; see, however, Subsections 3.4 and 3.5 in [de Wet and Randles, 1987] for some exceptions. Under certain smoothness conditions on \(F_\theta\) and the assumption that the estimator \(\hat{\theta}_n\) of the true parameter \(\theta_0\) satisfies \(\hat{\theta}_n - \theta_0 = n^{-1} \sum_{i=1}^{n} l_i + o_P(n^{-1/2})\), with \(l_1, \ldots, l_n\) satisfying certain conditions, the Cramér-von Mises test statistic with estimated parameters has to be approximated by a \(V\)-statistic with kernel
\[ \hat{h}((x', t'), (y', t')') = \int_{\mathbb{R}^d} \left[ \mathbb{I}_{X_{1} \leq z} - F_0(z) - \hat{F}_{\theta_0}(z) l_s \right] \left[ \mathbb{I}_{X_{2} \leq w} - F_0(z) - \hat{F}_{\theta_0}(z) l_t \right] w(z) \lambda^d(dz), \]
where \(\hat{F}_{\theta_0}(z)\) denotes the row vector of partial derivatives w.r.t. \(\theta = (\theta_1, \ldots, \theta_p)'\) of \(F_\theta(z)\) at \(\theta = \theta_0\). With such an approximation, limit distributions can be derived in complete analogy to the approach described in this paper; see Leucht and Neumann (2013) who also verified consistency of model-based bootstrap methods in this context. The application of the dependent wild bootstrap, however, requires some care. If the effect of estimating the unknown parameter \(\theta_0\) cannot be neglected, then we have to apply the DWB procedure to the kernel \(\hat{h}\) or, if not available, to an appropriate approximation thereof. In the case of testing (non-)linear (auto-)regression models, we can try to adapt the hybrid method described in Section 3 in [Escanciano, 2007] by replacing the independent wild bootstrap employed there by our dependent wild bootstrap method. A rigorous study of these extensions is beyond the scope of the present paper and should be carried out elsewhere.

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 \(((X'_i, Y'_i)')_i\) 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.
\[ \tilde{H}_0: \quad F_X = F_Y \quad \text{vs.} \quad \tilde{H}_1: \quad F_X \neq F_Y. \]
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,
\[ \tilde{T}_n = n \int_{\mathbb{R}^d} \left( F_X^{(n)}(z) - F_Y^{(n)}(z) \right)^2 w(z) \lambda^d(dz) = \frac{1}{n} \sum_{s,t=1}^{n} \tilde{h}((X'_s, Y'_s)', (X'_t, Y'_t)'), \]
where \(\tilde{h}((x'_1, y'_1)', (x'_2, y'_2)') = \int_{\mathbb{R}^d} \left( \mathbb{I}_{x_1 \leq z} - \mathbb{I}_{y_1 \leq z} \right) \left( \mathbb{I}_{x_2 \leq z} - \mathbb{I}_{y_2 \leq z} \right) w(z) \lambda^d(dz)\) and \(F_X^{(n)}\) and \(F_Y^{(n)}\) denote the empirical distribution functions based on \(X_1, \ldots, X_n\) and \(Y_1, \ldots, Y_n\), respectively. \(\tilde{T}_n\) is a \(V\)-statistic in \(((X'_i, Y'_i)')_i\) that is degenerate under \(\tilde{H}_0\). 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 w.r.t. Lebesgue measure.
4.3. Further applications. Assume that real-valued random variables $X_1, \ldots, X_n$ from a strictly stationary process $(X_t)_{t \in \mathbb{Z}}$ are observed. Let $F_n^X$ and $F_n^{-X}$ be the empirical distribution functions based on $X_1, \ldots, X_n$ and $-X_1, \ldots, -X_n$, respectively. In the case of i.i.d. random variables, Leucht (2012) proposed the following statistic for testing symmetry about zero:

$$T_n^{symm} = n \int_{-\infty}^{0} (F_n^X(u) - F_n^{-X}(u))^2 \, du$$

$$= \frac{1}{n} \sum_{s,t=1}^{n} \int_{-\infty}^{0} \left( \mathbb{I}(X_s \leq u) - \mathbb{I}(X_s \geq -u) \right) \left( \mathbb{I}(X_t \leq u) - \mathbb{I}(X_t \geq -u) \right) \, du.$$ 

Evaluating the integral on the right-hand side we obtain that

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

where $h(x, y) = (|x| \wedge |y|)(\mathbb{I}_{x > 0} - \mathbb{I}_{x < 0})$. The kernel $h$ is obviously symmetric, positive definite and degenerate if $P(X_0 \leq -u) = P(X_0 \geq u)$ for all $u \geq 0$. Furthermore, it is also Lipschitz continuous and it holds $Eh(X_0, X_0) = E|X_0|$. Hence, if $E|X_0| < \infty$ and if the underlying process satisfies (A1), then the limit distribution of $T_n$ follows from Theorem 2.1. Moreover, if additionally (B1) and (B2) are fulfilled, we obtain consistency of the dependent wild bootstrap from Theorem 5.1.

Model-based bootstrap-aided $L_2$-type tests for symmetry and the parametric class of the marginal distribution of a time series based on the empirical characteristic function were proposed by Arcones and Giné (1992) and by Neumann and Paparoditis (2000) based on fixed kernel estimates can also be constructed along these lines.

5. Numerical examples

We illustrate the finite sample behavior of the dependent wild bootstrap for $V$-statistics by two numerical examples related to our applications in Section 4.

**Example 5.1.** We revisit the Cramér-von Mises type test of Section 4.1. The statistic $T_n$ is applied to test

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

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 $T_n = n \int_{-\infty}^{\infty} [F_n(z) - F_0(z)]^2 f_0(z) \, dz$. In order to simulate the performance of our test under the null, we draw samples of size $n = 200, 300$ and $400$ from the stationary version of an AR(1) process

$$X_t = a X_{t-1} + \varepsilon_t,$$ \hspace{1cm} (5.1)

where $a = 0.5$ and the innovations $(\varepsilon_t)_t$ are i.i.d. $\mathcal{N}(0,0.75)$ random variables and thus the sample has standard normal marginals. Note that AR(1) processes with $|a| < 1$ and integrable innovations are $\tau$-dependent, cf. Dedecker and Prieur (2004).

Under the null hypothesis, the variables $X_1, \ldots, X_n$ are positive dependent in the sense of Gleser and Moore (1983). Recall that these authors defined a stochastic process, whose bivariate distributions are exchangeable, to be positive dependent if

$$Eh(X_i)h(X_j) \geq 0 \quad \forall \ h \quad \text{with} \quad E|h(X_i)h(X_j)| < \infty \quad \forall \ i,j.$$ \hspace{1cm} (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. Here, using quantiles from the i.i.d. case as critical values we obtained, for nominal significance levels of $\alpha = 0.05$ and $\alpha = 0.1$,
empirical sizes of 0.290 and 0.355, respectively. It can be seen from Figure 1 and Table 1 that both our bootstrap approximations are much more accurate.

For power investigations of the test we draw samples according to (5.1) with \( N(0, 2) \), \( N(0, 3, 1) \), and \( N(0.5, 1) \) marginals. The wild bootstrap variables \( (W^*_t)_{t=1}^\infty \) are generated via the procedure described in Remark 2 with \( l_n = 7, 10 \) and 15. 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).

![Figure 1](image)

**Figure 1.** Simulated cumulative distribution functions of \( T_{200} \) (black solid), \( T_{200}^{(\ast)} \) (blue dotted), \( T_{200, 2}^{(\ast)} \) (green dash-dotted), and \( T_{200}^{(ind)} \) (red dashed) under \( H_0 \), l.h.s. \( n=200 \), r.h.s. \( n=500 \)

The rejection frequencies of our test for nominal significance levels \( \alpha = 0.05 \) and \( \alpha = 0.1 \) are summarized in the Tables 1 to 3. Under the null scenario, the distribution of \( T_{n, 1}^{\ast} \) mimics the one of \( T_n \) better than that of the artificially degenerated version \( T_{n, 2}^{\ast} \). In contrast, the power properties of the test based on the second bootstrap statistic outperforms the first one for samples of moderate size. This is however not surprising, since \( T_{n, 1}^{\ast} \) tends to infinity with increasing sample size while \( T_{n, 2}^{\ast} \) remains bounded in the sense of \( P \left( P^* \left( T_{n, 2}^{\ast} > K(\epsilon) \right) < \epsilon \right) \rightarrow_{n \to \infty} 1 \). Moreover, Tables 1 to 3 indicate that our method is quite robust with respect to changes of the tuning parameter \( l_n \).

**Table 1.** Rejection frequencies \( (l_n = 7) \)

<table>
<thead>
<tr>
<th></th>
<th>( \mathcal{N}(0, 1) )</th>
<th>( \mathcal{N}(0, 2) )</th>
<th>( \mathcal{N}(0.3, 1) )</th>
<th>( \mathcal{N}(0.5, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{200, 1} )</td>
<td>0.055 0.095</td>
<td>0.250 0.520</td>
<td>0.650 0.805</td>
<td>0.960 0.985</td>
</tr>
<tr>
<td>( T_{200, 2} )</td>
<td>0.075 0.155</td>
<td>0.355 0.630</td>
<td>0.755 0.850</td>
<td>0.980 0.985</td>
</tr>
<tr>
<td>( T_{300, 1}^{\ast} )</td>
<td>0.075 0.150</td>
<td>0.620 0.845</td>
<td>0.885 0.910</td>
<td>1.000 1.000</td>
</tr>
<tr>
<td>( T_{300, 2}^{\ast} )</td>
<td>0.100 0.185</td>
<td>0.695 0.880</td>
<td>0.900 0.930</td>
<td>1.000 1.000</td>
</tr>
<tr>
<td>( T_{400, 1}^{\ast} )</td>
<td>0.085 0.125</td>
<td>0.815 0.965</td>
<td>0.930 0.960</td>
<td>1.000 1.000</td>
</tr>
<tr>
<td>( T_{400, 2}^{\ast} )</td>
<td>0.100 0.145</td>
<td>0.850 0.965</td>
<td>0.945 0.965</td>
<td>1.000 1.000</td>
</tr>
</tbody>
</table>

**Example 5.2.** As an illustration, we apply the two-sample test presented in Section 4.2 to the Nile data set that reports the annual flows of the river Nile from 1871 until 1970 and is contained in the R-package `strucchange`; see Figure 2. These data have been frequently used to study the
finite sample behavior of tests for structural changes since the construction of the Aswan dam in 1898 regularized the natural behavior of the river; cf. [Kirch and Tadjuidje (2011)] and references therein. To answer the question whether the dam has influenced the annual flows of the Nile, we invoke our test based on the statistic
\[ T_n = n \int_{\mathcal{R}} d \left( F_n^X(z) - F_n^Y(z) \right)^2 w(z) \lambda(dz). \]
Here \( w \) is chosen to be the density of \( \mathcal{N}(900, 22500) \), which puts sufficient mass on the interval that is spanned by all observations. The wild bootstrap variables are generated in the same manner as in the previous example. Let \( X_1, \ldots, X_{28} \) denote annual flows from 1871 until 1898 and \( Y_1, \ldots, Y_{28} \) those of the years 1899 up to 1926. Then the value of the test statistic is equal to \( T_{28} = 7.4 \) while the bootstrap approximation based on 500 resamplings of the upper 5%-quantile is 0.63. Thus, the null hypothesis of an equivalent stochastic behavior (in the sense of an identical marginal distribution) of the river before and after the construction of the Aswan dam is clearly rejected. We also apply our test to compare the flows from 1899 until 1926 to those of 1927 up to 1954, i.e. both time series describe flows of the Nile after the installation of the dam. Here, the value of the test statistic is \( T_{28} = 0.08 \) which lies below the estimated upper 5%-quantile 0.34, that is, our test supports the hypothesis of identical marginal distributions in this case.

**Table 2. Rejection frequencies \((l_n = 10)\)**

<table>
<thead>
<tr>
<th></th>
<th>( \mathcal{N}(0, 1) )</th>
<th>( \mathcal{N}(0, 2) )</th>
<th>( \mathcal{N}(0.3, 1) )</th>
<th>( \mathcal{N}(0.5, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.05 )</td>
<td>0.050</td>
<td>0.080</td>
<td>0.205</td>
<td>0.445</td>
</tr>
<tr>
<td>( \alpha = 0.1 )</td>
<td>0.080</td>
<td>0.155</td>
<td>0.360</td>
<td>0.625</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>0.070</td>
<td>0.130</td>
<td>0.525</td>
<td>0.810</td>
</tr>
<tr>
<td>( \alpha = 0.1 )</td>
<td>0.095</td>
<td>0.175</td>
<td>0.680</td>
<td>0.865</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>0.075</td>
<td>0.125</td>
<td>0.765</td>
<td>0.945</td>
</tr>
<tr>
<td>( \alpha = 0.1 )</td>
<td>0.100</td>
<td>0.145</td>
<td>0.820</td>
<td>0.960</td>
</tr>
</tbody>
</table>

**Table 3. Rejection frequencies \((l_n = 15)\)**

<table>
<thead>
<tr>
<th></th>
<th>( \mathcal{N}(0, 1) )</th>
<th>( \mathcal{N}(0, 2) )</th>
<th>( \mathcal{N}(0.3, 1) )</th>
<th>( \mathcal{N}(0.5, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.05 )</td>
<td>0.020</td>
<td>0.065</td>
<td>0.130</td>
<td>0.380</td>
</tr>
<tr>
<td>( \alpha = 0.1 )</td>
<td>0.090</td>
<td>0.155</td>
<td>0.380</td>
<td>0.605</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>0.050</td>
<td>0.115</td>
<td>0.425</td>
<td>0.760</td>
</tr>
<tr>
<td>( \alpha = 0.1 )</td>
<td>0.095</td>
<td>0.195</td>
<td>0.660</td>
<td>0.870</td>
</tr>
<tr>
<td>( \alpha = 0.05 )</td>
<td>0.060</td>
<td>0.115</td>
<td>0.695</td>
<td>0.900</td>
</tr>
<tr>
<td>( \alpha = 0.1 )</td>
<td>0.100</td>
<td>0.155</td>
<td>0.825</td>
<td>0.895</td>
</tr>
</tbody>
</table>

**Figure 2. Annual flows of the river Nile**
6. Proofs

6.1. Proofs of the main theorems.

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

We denote by \( (\lambda_k) \) 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 \( (\Phi_k) \). 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 \( \text{Sun (2005)} \), with \( X = \text{supp}(P_{X_0}) \)) that

\[
\begin{align*}
    h^{(K)}(x, y) = \sum_{k=1}^{K} \lambda_k \Phi_k(x) \Phi_k(y) & \rightarrow h(x, y) \quad \forall x, y \in \text{supp}(P_{X_0}). \\
\end{align*}
\]

The convergence of the series in (6.1) is absolute and uniform on compact subsets of \( \text{supp}(P_{X_0}) \). The prerequisites of this result can be checked fairly easily here: Clearly, \( P_{X_0} \) is non-degenerate on \( \text{supp}(P_{X_0}) \) and there are compact sets \( A_1 \subseteq A_2 \subseteq \cdots \) such that \( \text{supp}(P_{X_0}) = \bigcup_{n=1}^{\infty} A_n \). Moreover, \( h \) is a Mercer kernel (i.e. continuous, symmetric, positive semidefinite), \( \int h^2(x, y)P_{X_0}(dy) \leq h(x, x)Eh(X_0, X_0) < \infty \quad \forall x \in \text{supp}(P_{X_0}) \) and \( \int \int h^2(x, y)P_{X_0}(dx)P_{X_0}(dy) \leq (Eh(X_0, X_0))^2 < \infty. \)

Thus, assumptions 1, 2 and 3 of \( \text{Sun (2005)} \) are fulfilled; see also his Proposition 1.

We approximate \( V_n \) by a \( V \)-statistic with a kernel having a finite spectral decomposition,

\[
    V_n^{(K)} = \frac{1}{n} \sum_{k=1}^{n} h^{(K)}(X_s, X_t).
\]

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

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

\[
\sum_{r=1}^{n-1} 2 (1 - r/n) E \left[ h(X_0, X_r) - h^{(K)}(X_0, X_r) \right]
\]

\[
\leq 2 \sum_{r=1}^{\infty} E \left[ \sum_{k=K+1}^{\infty} \lambda_k \Phi_k(X_0) \Phi_k(X_r) \right]
\]

\[
= 2 \sum_{r=1}^{\infty} \left[ E \left[ \sum_{k=K+1}^{\infty} \lambda_k \Phi_k(X_0) \left( \Phi_k(X_r) - \Phi_k(\bar{X}_r) \right) \right] \right]
\]

\[
\leq 2 \sum_{r=1}^{\infty} \sqrt{\sum_{k=K+1}^{\infty} \lambda_k \sum_{r=1}^{\infty} \left| \sum_{k=1}^{\infty} \lambda_k \left( \Phi_k(X_r) - \Phi_k(\bar{X}_r) \right)^2 \right|}
\]

\[
\leq 2 \sum_{k=K+1}^{\infty} \lambda_k \sum_{r=1}^{\infty} \sqrt{E \left[ h(X_r, X_r) - h(X_r, \bar{X}_r) - h(\bar{X}_r, X_r) + h(\bar{X}_r, \bar{X}_r) \right]}
\]

\[
\leq 2 \sum_{k=K+1}^{\infty} \lambda_k \sum_{r=1}^{\infty} \sqrt{2 \text{Lip}(h) \sqrt{\tau(r)}},
\]

where $\bar{X}_r$ denotes a copy of $X_r$ that is independent of $X_0$ and satisfies $E\|X_r - \bar{X}_r\|_1 \leq \tau(r)$. Since $\sum_{k=1}^{\infty} \lambda_k = E h(X_0, X_0) < \infty$ and thus $\sum_{k=K+1}^{\infty} \lambda_k \to 0$ as $K \to \infty$, we obtain

\[
\sup_n E \left[ V_n - V_n^{(K)} \right] \xrightarrow{K \to \infty} 0. \tag{6.2}
\]

(ii) CLT for partial sums

We are going to show that for $K \leq L$

\[
V_n^{(K)} = \sum_{k=1}^{K} \lambda_k \left( n^{-1/2} \sum_{t=1}^{n} \Phi_k(X_t) \right)^2 \xrightarrow{d} \sum_{k=1}^{K} \lambda_k Z_k^2. \tag{6.3}
\]

By the continuous mapping theorem, this follows from

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t \xrightarrow{d} (Z_1, \ldots, Z_K)',
\]

with $Y_t = (\Phi_1(X_t), \ldots, \Phi_K(X_t))'$. To prove this, we apply a multivariate version of the central limit theorem of Neumann (2013) to the random variables $n^{-1/2} Y_t$; see Theorem 6.1 below. We 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 (6.27) is fulfilled.

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

\[
(\text{cov}(\Phi_j(X_s), \Phi_k(X_{s+t})))^2 = \left( E \left[ \Phi_j(X_s) \left( \Phi_k(X_{s+t}) - \Phi_k(\bar{X}_{s+t}) \right) \right] \right)^2 \\
\leq E \left( \Phi_k(X_{s+t}) - \Phi_k(\bar{X}_{s+t}) \right)^2 \\
\leq \frac{1}{\lambda_k} E \left[ h(\bar{X}_{s+t}, X_{s+t}) - h(X_{s+t}, \bar{X}_{s+t}) \right] \\
\leq \frac{2 \text{Lip}(h)}{\lambda_k} \tau(r).
\]

(6.5)

Hence, we have that $\sum_{r=-\infty}^{\infty} |\text{cov}(\Phi_j(X_0), \Phi_k(X_r))| < \infty$ and we obtain by majorized convergence

\[
\left| \frac{1}{n} \sum_{r=1}^{n} \text{cov}(\Phi_j(X_s), \Phi_k(X_t)) - \sum_{r=-\infty}^{\infty} \text{cov}(\Phi_j(X_0), \Phi_k(X_r)) \right| \\
\leq \sum_{r \in \mathbb{Z}} \min\{|r|/n, 1\} |\text{cov}(\Phi_j(X_0), \Phi_k(X_r))| \xrightarrow{n \to \infty} 0,
\]

i.e., condition (6.25) is fulfilled.

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

\[
E \left( \Phi_k(X_{s+t}) - \Phi_k(\bar{X}_{s+t}) \right)^2 = O(\tau(r)).
\]

(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 $g: \mathbb{R}^{k_u} \to \mathbb{R}$ with $\|g\|_\infty = \sup_{x \in \mathbb{R}^{k_u}} |g(x)| \leq 1$ that

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

and

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

i.e., conditions (6.27) and (6.28) are also fulfilled. Hence, (6.4) follows from Theorem 6.1 below.

(iii) Convergence of $V_n$ and $U_n$

The convergence

\[
V_n \xrightarrow{d} Z := \sum_k \lambda_k Z_k^2
\]

follows from (6.2), (6.3), and Theorem 2 of Dehling, Durieu, and Volya (2009), which is an improved variant of Theorem 4.2 of Billingsley (1968) for complete spaces. Moreover, note that $U_n = \sum_{t=1}^{n-1} h(X_t, X_t)$, 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 $g: \mathbb{R} \to \mathbb{R}$ is locally Lipschitz-continuous function and $(X_t)_t$ is a stationary sequence of $\tau$-dependent random variables with $E|g(X_1)| < \infty$, can be applied here.

(iv) Expectation of $Z$

It follows from (6.5) that $\sum_{r=-\infty}^{\infty} |E[\Phi_k(X_0) \Phi_k(X_r)]| < \infty$. Therefore, we obtain by majorized convergence that

\[
E[V_n^{(K)}] = \sum_{k=1}^{K} \lambda_k \sum_{r=-\infty}^{(n-1)} \left( 1 - \frac{|r|}{n} \right) E[\Phi_k(X_0) \Phi_k(X_r)] \\
\xrightarrow{n \to \infty} \sum_{k=1}^{K} \lambda_k Z_k^2 = E \left[ \sum_{k=1}^{K} \lambda_k Z_k^2 \right],
\]

for $1 \leq j, k \leq K$, that

\[
(\text{cov}(\Phi_j(X_s), \Phi_k(X_{s+t})))^2 = \left( E \left[ \Phi_j(X_s) \left( \Phi_k(X_{s+t}) - \Phi_k(\bar{X}_{s+t}) \right) \right] \right)^2 \\
\leq E \left( \Phi_k(X_{s+t}) - \Phi_k(\bar{X}_{s+t}) \right)^2 \\
\leq \frac{1}{\lambda_k} E \left[ h(\bar{X}_{s+t}, X_{s+t}) - h(X_{s+t}, \bar{X}_{s+t}) \right] \\
\leq \frac{2 \text{Lip}(h)}{\lambda_k} \tau(r).
\]
which implies in conjunction with (6.2) that
\[ EV_n \xrightarrow{n \to \infty} EZ. \] (6.7)

On the other hand, it follows from the Lipschitz continuity of \( h \) that \( \sum_{r=-(n-1)}^{n-1} |Eh(X_0, X_r)| < \infty \). This yields, again by majorized convergence,
\[ E[V_n^{(K)}] = \sum_{r=-(n-1)}^{n-1} \left( 1 - \frac{|r|}{n} \right) Eh(X_0, X_r) \xrightarrow{n \to \infty} \sum_{r=-(n-1)}^{n-1} Eh(X_0, X_r). \] (6.8)

From (6.7) and (6.8) we deduce that
\[ EZ = \sum_{r=-(n-1)}^{n-1} Eh(X_0, X_r). \]

**Proof of Corollary 2.7** In the proof of Theorem 2.1, Lipschitz continuity of \( h \) is invoked to obtain the inequality
\[ \sum_{r \in \mathbb{N}} \sqrt{E[h(X_r, X_r) - h(X_r, \tilde{X}_r) - h(\tilde{X}_r, X_r) + h(\tilde{X}_r, \tilde{X}_r)]} < \infty, \]
where \( \tilde{X}_r \) denotes a copy of \( X_r \) that is independent of \( X_0, X_1, \ldots \) and satisfies \( E\|X_r - \tilde{X}_r\|_1 \leq \tau(r) \). Under (A3) one obtains a similar result by Hölder’s inequality,
\[ \sum_{r \in \mathbb{N}} \sqrt{E[h(X_r, X_r) - h(X_r, \tilde{X}_r) - h(\tilde{X}_r, X_r) + h(\tilde{X}_r, \tilde{X}_r)]} \]
\[ \leq \sum_{r \in \mathbb{N}} \sqrt{E\left( |f(X_r, X_r, X_r, \tilde{X}_r) + f(\tilde{X}_r, \tilde{X}_r, X_r, \tilde{X}_r)| \right) \|X_r - \tilde{X}_r\|_1} \]
\[ \leq \sum_{r \in \mathbb{N}} \sqrt{E[f(X_r, X_r, X_r, \tilde{X}_r) + f(\tilde{X}_r, \tilde{X}_r, X_r, \tilde{X}_r)]^{1/(1-\delta)} \left( \|X_r\|_1 + \|\tilde{X}_r\|_1 \right)^{1-\delta}} \left( \|X_r - \tilde{X}_r\|_1 \right)^{\delta} \]
\[ \leq \sum_{r \in \mathbb{N}} \sqrt{E[f(X_r, X_r, X_r, \tilde{X}_r) + f(\tilde{X}_r, \tilde{X}_r, X_r, \tilde{X}_r)]^{1/(1-\delta)} \left( \|X_r\|_1 + \|\tilde{X}_r\|_1 \right)^{1-\delta}} \left( \tau(r) \right)^{\delta/2} \]
\[ < \infty. \]

Therefore, the limit distribution of \( V_n \) is as given in Theorem 2.1. 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. \( \square \)

**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
\[ V_{n,2}^* - V_{n,1}^* = -\frac{2}{n^2} \sum_{s,t,k=1}^{n} h(X_s, X_k)W_s^*W_t^* + \frac{1}{n^3} \sum_{k,l,s,t=1}^{n} h(X_k, X_l)W_s^*W_t^* \]
\[ = -\frac{1}{2n^2} \sum_{s,t=1}^{n} \{(W_s^* - 1)h(X_s, X_t)(W_t^* - 1) - (W_s^* + 1)h(X_s, X_t)(W_t^* + 1)\} \sum_{k=1}^{n} W_k^* \]
\[ + V_n \left( \frac{1}{n} \sum_{s=1}^{n} W_s^* \right)^2. \]
\[ V_{n,-} := n^{-1} \sum_{s,t=1}^n (W^*_{s,t} - 1) h(X_s, X_t)(W^*_{s,t} - 1) \text{ and } V_{n,+} := n^{-1} \sum_{s,t=1}^n (W^*_{s,t} + 1) h(X_s, X_t)(W^*_{s,t} + 1) \]

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

\[ P \left( P^* \left( |V^*_{n,2} - V^*_{n,1}| > \epsilon \right) > \delta \right) \to 0 \quad \forall \epsilon > 0, \delta > 0, \]
i.e., the difference between \( V^*_{n,1} \) and \( V^*_{n,2} \) is asymptotically negligible. Therefore, we only consider \( V^*_{n,1} \) in the sequel.

(b) Approximation of the \( V \)-statistic

We define

\[ V^{(K)*}_{n,1} := \frac{1}{n} \sum_{s,t=1}^n h^{(K)}(X_s, X_t) W^*_s W^*_t = \sum_{k=1}^K \lambda_k \left( \frac{1}{\sqrt{n}} \sum_{s=1}^n \Phi_k(X_s) W^*_s \right)^2. \]

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

\[ \limsup_{n \to \infty} P \left( P^* \left( |V^*_{n,1} - V^{(K)*}_{n,1}| > \epsilon \right) > \delta \right) \to 0 \quad \forall \delta, \epsilon > 0. \]

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

\[ EE^*(V^*_{n,1} - V^{(K)*}_{n,1}) \]

\[ \begin{align*}
&= \frac{1}{n} \sum_{k=K+1}^\infty \lambda_k \sum_{s,t=1}^n E[\Phi_k(X_s)\Phi_k(X_t)] \rho(|s-t|/l_n) \\
&\leq \sum_{k=K+1}^\infty \lambda_k \left\{ 1 + \sum_{r=1}^{n-1} \frac{2(n-r)}{n} \left| E \left[ \Phi_k(X_0) \left( \Phi_k(X_r) - \Phi_k(\bar{X}_r) \right) \right] \right| \rho(r/l_n) \right\} \\
&\leq \sum_{k=K+1}^\infty \lambda_k + 2 \sum_{r=1}^{n-1} \lambda_k \sqrt{2} \text{Lip}(h) \sum_{r=1}^\infty \sqrt{r},
\end{align*} \]

where \( \bar{X}_r \) denotes a copy of \( X_r \) that is independent of \( X_0 \) and such that \( E\|X_r - \bar{X}_r\|_1 \leq \tau(r) \).

(c) CLT for the partial sums

We define the vectors \( Y^*_t = (\Phi_1(X_t), \ldots, \Phi_K(X_t))/W^*_t \). We show that

\[ \frac{1}{\sqrt{n}} \sum_{t=1}^n Y^*_t \xrightarrow{d} (Z_1, \ldots, Z_K) \sim N(0_K, \Sigma_K) \quad \text{in probability}, \]

where \( (\Sigma_K)_{j,k} = \sum_{r=-\infty}^\infty \text{cov}(\Phi_j(X_0), \Phi_k(X_r)) \). To this end, we decompose the index set \( \{1, \ldots, n\} \) into disjoint blocks of length \( L_n = o(n) \), \( I_{n,s} = \{ t \in \mathbb{N} : (s-1)L_n < t \leq sL_n \} \), for \( s = 1, \ldots, k_n - 1 \) with \( k_n - 1 = [n/L_n] \), and a block of the remaining indices, \( I_{n,k_n} = \{ t \in \mathbb{N} : (k_n - 1)L_n < t \leq n \} \).

We define

\[ X^*_{n,s} = (X^*_{n,s,1}, \ldots, X^*_{n,s,K})' := n^{-1/2} \sum_{t \in I_{n,s}} Y^*_t \]

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
\[ P \left( \sum_{s=1}^{k_n} E^*(X_{n,s,j}^*)^2 \leq v_0 \right) \xrightarrow{n \to \infty} 1. \]

According to the definition of the blocks, we obtain
\[ \sum_{s=1}^{k_n} E^*(X_{n,s,j}^*)^2 = \sum_{s=1}^{k_n-1} \frac{1}{n} \sum_{u,v \in I_{n,s}} \Phi_j(X_u)\Phi_j(X_v)\rho(|u-v|/l_n) + o_P(1) \]
\[ \leq \frac{1}{L_n} \sum_{u,v \in I_{n,1}} E[\Phi_j(X_u)\Phi_j(X_v)] \rho(|u-v|/l_n) + o_P(1). \]

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 \( \tau \)-dependence condition (B1) on \( (X_t)_t \).

Next we show that
\[ \text{Cov}^* \left( X_{n,1}^* + \cdots + X_{n,k_n}^* \right) \xrightarrow{P} \Sigma_K. \quad (6.12) \]

We have that
\[ \left( \text{Cov}^* \left( X_{n,1}^* + \cdots + X_{n,k_n}^* \right) \right)_{j,k} = \text{cov}^* \left( n^{-1/2} \sum_{s=1}^{n} \Phi_j(X_s)W_{s}^*, n^{-1/2} \sum_{t=1}^{n} \Phi_k(X_t)W_{t}^* \right) \]
\[ = \frac{1}{n} \sum_{s,t=1}^{n} \Phi_j(X_s)\Phi_k(X_t)\rho(|s-t|/l_n) \]
\[ = \frac{1}{n} \sum_{s,t=1}^{n} \{ \Phi_j(X_s)\Phi_k(X_t) - E[\Phi_j(X_s)\Phi_k(X_t)] \} \rho(|s-t|/l_n) \]
\[ + \sum_{r=-\infty}^{\infty} E[\Phi_j(X_0)\Phi_k(X_r)] \rho(|r|/l_n) \max\{1-|r|/n,0\}. \]

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, (6.12) holds true.

\( (c.2) \) **Lindeberg condition**

We want to show that, for arbitrary \( \epsilon > 0 \),
\[ L_{n,j}^*(\epsilon) := \frac{1}{n} \sum_{s=1}^{k_n} E^* \left[ (X_{n,s,j}^*)^2 1(|X_{n,s,j}^*| > \epsilon \sqrt{n}) \right] \xrightarrow{P} 0. \quad (6.13) \]

For this, it suffices to show that
\[ E[L_{n,j}^*(\epsilon)] \xrightarrow{n \to \infty} 0. \quad (6.14) \]

We obtain from the stationarity of the involved processes that
\[ E[L_{n,j}^*(\epsilon)] \leq \frac{k_n - 1}{n} \ E E^* \left[ (X_{n,1,j}^*)^2 1(|X_{n,1,j}^*| > \epsilon \sqrt{n}) \right] + \frac{1}{n} \ E E^* \left[ (X_{n,k_n,j}^*)^2 \right]. \]

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
\[ n^{1/2}L_n^{-1/2}X_{n,1,j}^* = L_n^{-1/2} \sum_{t \in I_{n,1}} \Phi_j(X_t)W_t^* \xrightarrow{d} \mathcal{N}(0, v_j) \quad (6.15) \]

with some \( v_j \in [0, \infty) \). On the other hand, we have that
\[ E E^* \left[ n/L_n (X_{n,1,j}^*)^2 \right] \xrightarrow{n \to \infty} v_j. \quad (6.16) \]
Now (6.15) and (6.16) imply uniform integrability of \((n/L_n (X_{n,1,j}^*)^2)_{n \in \mathbb{N}}\), which also yields that
\[
E E^* \left[ n/L_n (X_{n,1,j}^*)^2 \mathbb{1}(|\sqrt{n/L_n} X_{n,1,j}^*| > c) \right] \xrightarrow{n \to \infty} E[Y_j^2 \mathbb{1}(|Y_j| > c)],
\]
where \(Y_j \sim \mathcal{N}(0, v_j)\). Since \(E[Y_j^2 \mathbb{1}(|Y_j| > c)] \to 0,\) we obtain from (6.17) that (6.14) holds true.

(c.3) Weak dependence
The conditions (6.31) and (6.32) can be proved along the same lines. Therefore, we only state the proof of (6.31) in Lemma 6.3 below.

(d) Conclusion
Now we can apply Corollary 6.1 and we obtain (6.10), which in turn implies by the continuous mapping theorem that
\[
V_n^{(K)^*} \xrightarrow{d} \sum_{k=1}^{K} \lambda_k Z_k^2 \quad \text{in probability.}
\]
Invoking Theorem 2 of Dehling, Durieu, and Volny (2009), we obtain from (6.9) and (6.18) that
\[
V_n^* \xrightarrow{d} Z \quad \text{in probability.}
\]

(ii) U-statistics
In order to deduce the corresponding result for \(U_{n,1}^*\), we have to show that
\[
P^* \left( \frac{1}{n} \sum_{t=1}^{n} h(X_t, X_t)(W_t^*)^2 - E h(X_0, X_0) \right) > \epsilon \right) \xrightarrow{P} 0 \quad \forall \epsilon > 0.
\]
With \(M > 0\), we decompose:
\[
\frac{1}{n} \sum_{t=1}^{n} h(X_t, X_t)(W_t^*)^2 - E h(X_0, X_0) \leq \frac{1}{n} \sum_{t=1}^{n} [h(X_t, X_t) - h(X_t, X_t) \wedge M] (W_t^*)^2 \]
\[
+ \frac{1}{n} \sum_{t=1}^{n} [h(X_t, X_t) \wedge M] [(W_t^*)^2 - 1] \]
\[
+ \frac{1}{n} \sum_{t=1}^{n} h(X_t, X_t) \wedge M - E[h(X_0, X_0) \wedge M] \]
\[
+ |E[h(X_0, X_0) \wedge M] - E[h(X_0, X_0)]|
\]
\[
= R_{n,1} + \cdots + R_{n,4}.
\]
We obtain from monotone convergence that
\[
R_{n,4} \xrightarrow{M \to \infty} 0
\]
and
\[
E E^*[R_{n,1}] \xrightarrow{M \to \infty} 0.
\]
Moreover, it follows from the weak law of large numbers of Leucht (2011, Lemma 5.1) that
\[
R_{n,3} \xrightarrow{P} 0.
\]
Thus, it remains to prove that
\[
P^* \left( \frac{1}{n} \sum_{t=1}^{n} [h(X_t, X_t) \wedge M] [(W_t^*)^2 - 1] \right) > \epsilon \right) \xrightarrow{P} 0 \quad \forall \epsilon > 0.
\]
To this end, we truncate the bootstrap variables and obtain

\[ P^* \left( \left| \frac{1}{n} \sum_{t=1}^{n} [h(X_t, X_t) \wedge M]([W_t^*]^2 - 1) \right| > \epsilon \right) \]

\[ \leq P^* \left( \left| \frac{1}{n} \sum_{t=1}^{n} [h(X_t, X_t) \wedge M]([W_t^*]^2 - C - E^*([W_t^*]^2 \wedge K)] \right| > \epsilon/3 \right) \]

\[ + \frac{3M}{\epsilon} (1 - E^*([W_t^*]^2 \wedge K)) \]

\[ + \mathbb{1} \left( M(1 - E^*([W_t^*]^2 \wedge K]) > \epsilon/3 \right). \]

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 \( \tau \)-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 (6.20) and thus (6.19).

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

\[ P^* \left( \left| \frac{1}{n} \sum_{t=1}^{n} \bar{h}(X_t, X_t)(W_t^* - Eh(X_0, X_0)) \right| > \epsilon \right) \xrightarrow{P} 0 \quad \forall \epsilon > 0. \quad (6.21) \]

We split up

\[ \frac{1}{n} \sum_{t=1}^{n} \bar{h}(X_t, X_t)(W_t^* - Eh(X_0, X_0)) ^2 \]

\[ = \frac{1}{n} \sum_{t=1}^{n} h(X_t, X_t)(W_t^* - Eh(X_0, X_0))^2 - \frac{2}{n} \sum_{t,k=1}^{n} h(X_t, X_k)(W_t^*)^2 + \frac{1}{n} \sum_{t=1}^{n} (W_t^*). \]

According to (6.19), 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

\[ EE^* \left( \frac{1}{n^2} \sum_{t,k=1}^{n} h(X_t, X_k) \left( (W_t^*)^2 - (W_t^*)^2 \wedge K \right) \right) \xrightarrow{K \to \infty} 0. \]

Since the kernel \( h \) is positive semidefinite, the \((n \times n)\)-matrix \( H \) with \((s, t)\)th entry \( H_{s,t} = h(X_s, X_t) \) is also positive semidefinite. Therefore, we obtain from inequality 12.1.b(i) on page 258 in Seber (2008) that

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

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 (6.21).

\subsection{(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; see e.g. van der Vaart (1998, Lemma 2.11).

\begin{proof}[Proof of Lemma 3.7]
If \( EZ = 0 \), then \( Z = 0 \) almost surely.

Suppose now that \( EZ > 0 \). We deduce continuity of the limit distribution from continuity properties of \( Z^{(K)} = \sum_{k=1}^{K} \lambda_k Z_k^2 \). Since \( EZ < \infty \) we obtain

\[ Z^{(K)} \xrightarrow{d} Z, \]
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 $\varepsilon > 0$ that

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

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

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

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 $\Lambda_K = \text{Diag}(\lambda_1, \ldots, \lambda_K)$, $\Sigma_K = \text{Cov}((Z_1, \ldots, Z_K)'$ and $N_K \sim \mathcal{N}(0, \mathbf{I}_K)$. For any symmetric matrix $M$, denote by $\alpha_i(M)$ the $i$th 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 $\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}$, we obtain by the inclusion principle for principle submatrices (see Lütkepohl (1996, Section 9.13.4, page 160)) that $\alpha_i(\Sigma_{K}^{1/2} \Lambda_K \Sigma_{K}^{1/2}) \leq \alpha_k(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2})$. 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 $\alpha_1(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}) \geq c_0 > 0, \forall K > K_0$ and for some $c_0 > 0$. As convolution preserves the continuity properties of the smoother function, the latter inequality implies continuity of the distribution functions of $Z^{(K)} = \sum_{k=1}^{K} \alpha_k(\Lambda_K^{1/2} \Sigma_K \Lambda_K^{1/2}) Y_k$, uniformly for all $K > K_0$, where $Y_1, \ldots, Y_K$ are i.i.d. standard normal random variables.

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

$$E[n^{-1}T_n] = E[h(X_0, \tilde{X}_0)] = \frac{1}{n} \sum_{r=-(n-1)}^{n-1} (1 - |r|/n) E[h(X_0, X_r) - h(X_0, \tilde{X}_0)] = O\left(\frac{1}{n}\right)$$

and

$$\text{var}(n^{-1}T_n) = \frac{1}{n^4} \sum_{s,t,u,v=1}^{n} \{E[h(X_s, X_t) h(X_u, X_v)] - E[h(X_s, X_t)] E[h(X_u, X_v)]\} \to 0 \text{ as } n \to \infty.$$

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 $\varepsilon, \delta > 0$,

$$P\left(P^*\left(n^{-1}T_{n,1}^* > \varepsilon\right) \leq \delta\right) \to \infty, \text{ as } n \to \infty,$$

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 $h_{\text{deg}}$ as

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

It follows that

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

i.e., under empirical degeneration, we can replace $h$ by $h_{\text{deg}}$. This means that $T_{n,2}^*$ can be written as an empirically degenerated $V$-statistic with the kernel $h_{\text{deg}}$ 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 $Z$, see what could happen, assume that there is a sequence of asymptotic normality for a sequence of random vectors has to be proved and only a univariate CLT

$$\epsilon' Z_n^* \overset{d}{\to} \epsilon' Z$$

in probability, \hspace{\textwidth}

(6.23)

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

$$Z_n^* \overset{d}{\to} Z$$

in probability. \hspace{\textwidth}

(6.24)

To see why, note that [6.23] can be reformulated in such a way that there exist “bad sets” $\Omega_1(c), \Omega_2(c), \ldots$ such that

$$P ((X_1, \ldots, X_n)' \in \Omega_n(c)) \to_n 0$$

and, for any sequence $(\omega_n)_{n \in \mathbb{N}}$ with $\omega_n \not\in\Omega_n(c)$,

$$P'Z_n^*((X_1,\ldots,X_n)'=\omega_n) \rightarrow P'Z.$$

If the universal sets $\Omega_n := \bigcup_{c \in \mathbb{R}^d} \Omega_n(c)$ were measurable and satisfy $P((X_1, \ldots, X_n)' \in \Omega_n) \to_n 0$, then we could actually conclude from [6.23] that [6.24] 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 $(X_{n,k})_{k=1,\ldots,n}$, $n \in \mathbb{N}$, is a triangular scheme of $\mathbb{R}^d$-valued random vectors with $EX_{n,k} = 0_d$ for all $n, k$ and $\sum_{k=1}^{k_n} E X_{n,k,j}^2 \leq v_0$, for all $n \in \mathbb{N}, j \in \{1, \ldots, d\}$ and some $v_0 < \infty$. We assume that

$$\Sigma_n := \text{Cov}(X_{n,1} + \cdots + X_{n,k_n}) \to_{n \to \infty} \Sigma,$$

for some positive semidefinite matrix $\Sigma$, and that

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

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 $1 \leq s_1 < s_2 < \cdots < s_u < s_u + r = t_1 \leq t_2 \leq k_n$ and all $j_1, j_2 \in \{1, \ldots, d\}$, the following upper bounds for covariances hold true: for all measurable functions $g: \mathbb{R}^{du} \to \mathbb{R}$ with $\|g\|_\infty = \sup_{x \in \mathbb{R}^{du}} |g(x)| \leq 1$,

$$|\text{cov}(g(X_{n,s_1}, \ldots, X_{n,s_u}), X_{n,s_{u+1}}, \ldots, X_{n,k_n})| \leq (EX_{n,s_1,j_1}^2 + EX_{n,t_1,j_2}^2 + k_n^{-1}) \theta_r$$

and

$$|\text{cov}(g(X_{n,s_1}, \ldots, X_{n,s_u}), X_{n,t_1,j_1}, X_{n,t_2,j_2})| \leq (EX_{n,t_1,j_1}^2 + EX_{n,t_2,j_2}^2 + k_n^{-1}) \theta_r.$$

Then

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

**Proof.** Let $c \in \mathbb{R}^d$ be arbitrary. It can be easily seen that the triangular scheme $(c'X_{n,k})_{k=1,\ldots,n}$, $n \in \mathbb{N}$, satisfies the conditions of the univariate CLT of Neumann [2013]. (This theorem remains true if we have $k_n$ rather than $n$ summands in the $n$-th row of the triangular scheme. Note in
Suppose that assumptions (A2), (B1), and (B2) are fulfilled. Then, for any fixed

\[ c'(X_{n,1} + \cdots + X_{n,k_n}) \overset{d}{\rightarrow} \mathcal{N}(0, c'\Sigma), \]

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 \( X_{n,1}, \ldots, X_{n,n} \) defined on a probability space \((\Omega, \mathcal{A}, P)\), \( \mathbb{R}^d \)-valued bootstrap variables \( X_{n,1}^*, \ldots, X_{n,k_n}^* \) are available with \( E^* X_{n,k}^* = 0 \) for all \( n, k \),

\[
P\left( \sum_{k=1}^{k_n} E^*(X_{n,k,j}^*)^2 \leq v_0 \right) \overset{n \to \infty}{\to} 1
\]

for all \( j \in \{1, \ldots, d\} \) and some \( v_0 < \infty \). We assume that

\[
\Sigma_n^* := \text{Cov}^*(X_{n,1}^* + \cdots + X_{n,k_n}^*) \overset{P}{\rightarrow} \Sigma,
\]

for some positive semidefinite matrix \( \Sigma \), and that

\[
\sum_{k=1}^{k_n} E^*[\{(X_{n,k,j}^*)^2 \mathbb{1}(|X_{n,k,j}^*| > \epsilon)\}] \overset{P}{\rightarrow} 0
\]

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 \( u \in \mathbb{N} \), the following upper bounds for covariances hold true: for all measurable functions \( g: \mathbb{R}^{du} \to \mathbb{R} \) with \( \|g\|_{\infty} = \sup_{x \in \mathbb{R}^{du}} |g(x)| \leq 1 \),

\[
P\left( |\text{cov}^* (g(X_{n,s_1}, \ldots, X_{n,s_u})) X_{n,s_1},j_1 \cdots X_{n,s_2},j_2) \right) \leq (E^*(X_{n,s_1,j_1})^2 + E^*(X_{n,s_2,j_2})^2 + k_n^{-1}) \theta_r
\]

\[
\forall u \in \mathbb{N}, 1 \leq s_1 < s_2 < \ldots < s_u < s + r = t_1 \leq k_n, j_1, j_2 \in \{1, \ldots, d\} \overset{n \to \infty}{\to} 1 \tag{6.31}
\]

and

\[
P\left( |\text{cov}^* (g(X_{n,s_1}, \ldots, X_{n,s_u}), X_{n,t_1,j_1} \cdots X_{n,t_2,j_2}) \right) \leq (E^*(X_{n,t_1,j_1})^2 + E^*(X_{n,t_2,j_2})^2 + k_n^{-1}) \theta_r
\]

\[
\forall u \in \mathbb{N}, 1 \leq s_1 < s_2 < \ldots < s_u < s + r = t_1 \leq t_2 \leq k_n, j_1, j_2 \in \{1, \ldots, d\} \overset{n \to \infty}{\to} 1. \tag{6.32}
\]

Then

\[
X_{n,1}^* + \cdots + X_{n,k_n}^* \overset{d}{\rightarrow} \mathcal{N}(0, \Sigma) \quad \text{in probability.}
\]

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

6.3. Some auxiliary lemmas.

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

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

**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

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

\[
\frac{1}{n} \sum_{s,t=1}^{n} \{ \Phi_j(X_s) \Phi_k(X_t) \rho(|s-t|/l_n) - E[\Phi_j(X_s) \Phi_k(X_t) \rho(|s-t|/l_n)] \} \overset{P}{\rightarrow} 0.
\]
for $l = j, k$ are fulfilled. (Since $P(|\Phi_l(X_0)| > \sqrt{n}/\delta) \leq n^{-1}E[\Phi_l^2(X_0) \mathbb{1}(|\Phi_l(X_0)| > \sqrt{n}/\delta)]\delta^2 = o(n^{-1})$ 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:

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

say.

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

\[
P( R_{n,1} \neq 0 ) \\
\leq P( \Phi_j(X_t) \neq Y_{j,t} \text{ or } \Phi_k(X_t) \neq Y_{k,t} \text{ for some } t \in \{1, \ldots, n\} ) \rightarrow 0. \quad (6.33)
\]

Furthermore, we obtain from

\[
|Y_{l,t} - \bar{Y}_{l,t}| = |EY_{l,t}| = |E[Y_{l,t} - \Phi_l(X_t)]| \\
\leq E[|\Phi_l(X_t)| \mathbb{1}(|\Phi_l(X_t)| > M_n)] \\
\leq M_n^{-1} E[\Phi_l^2(X_t) \mathbb{1}(|\Phi_l(X_t)| > M_n)] \\
= o(M_n^{-1})
\]

that

\[
R_{n,2} \\
\leq \frac{1}{n} \sum_{s,t=1}^{n} |Y_{j,s} - \bar{Y}_{j,s}| \cdot |Y_{k,t} - \bar{Y}_{k,t}| \cdot \rho(|s-t|/l_n) \\
+ \frac{1}{n} \sum_{s=1}^{n} \bar{Y}_{j,s} \sum_{t=1}^{n} (Y_{k,t} - \bar{Y}_{k,t}) \rho(|s-t|/l_n) \\
+ \frac{1}{n} \sum_{t=1}^{n} \bar{Y}_{k,t} \sum_{s=1}^{n} (Y_{j,s} - \bar{Y}_{j,s}) \rho(|s-t|/l_n) \\
= o \left( \frac{l_n}{M_n^2} \right) + O_p \left( \frac{l_n}{\sqrt{n} M_n} \right). \quad (6.34)
\]
Let $A_n$ denote the $n \times n$-matrix with entries $(A_n)_{s,t} = \rho(|s-t|/l_n)$. We have that

\[ ER_{n,3}^2 = \frac{1}{n^2} \sum_{s,t,u,v=1}^{n} (A_n)_{s,t}(A_n)_{u,v} \left\{ E[\bar{Y}_{j,s}\bar{Y}_{k,t}, \bar{Y}_{j,u}\bar{Y}_{k,v}] - E[\bar{Y}_{j,s}\bar{Y}_{k,t}]E[\bar{Y}_{j,u}\bar{Y}_{k,v}] \right\} \]

\[ = \frac{1}{n^2} \sum_{s,t,u,v=1}^{n} (A_n)_{s,t}(A_n)_{u,v} \text{cum}(\bar{Y}_{j,s}\bar{Y}_{k,t}, \bar{Y}_{j,u}\bar{Y}_{k,v}) \]

\[ + \frac{1}{n^2} \sum_{s,t,u,v=1}^{n} (A_n)_{s,t}(A_n)_{u,v}E[\bar{Y}_{j,s}\bar{Y}_{j,u}]E[\bar{Y}_{k,t}\bar{Y}_{k,v}] \]

\[ + \frac{1}{n^2} \sum_{s,t,u,v=1}^{n} (A_n)_{s,t}(A_n)_{u,v}E[\bar{Y}_{j,s}\bar{Y}_{k,t}]E[\bar{Y}_{j,u}\bar{Y}_{k,v}], \tag{6.35} \]

where \text{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,t}]E[\bar{Y}_{j,u}\bar{Y}_{k,v}]\] 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 $O(M_n^2n^{-1})$. Furthermore, we have that

\[ \text{cov}(\bar{Y}_{j,s}, \bar{Y}_{k,t}) = O(\sqrt{\tau(|s-t|)}). \]

This implies that the second and the third term on the right-hand side of (6.35) are of order $O(l_n/n)$. Hence, we obtain

\[ ER_{n,3}^2 = O(M_n^2n^{-1} + l_n n^{-1}) = o(1). \tag{6.36} \]

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

\[ |E[\bar{Y}_{j,s}\bar{Y}_{k,t}] - E[\Phi_j(X_s)\Phi_k(X_t)]| \]

\[ \leq \left| E[\bar{Y}_{j,s} - \Phi_j(X_s)Y_{k,t}] \right| + \left| E[\Phi_j(X_s)(\bar{Y}_{k,t} - \Phi_k(X_t))] \right| \]

\[ \leq D_{n,j}\sqrt{\tau(t-s)}\sqrt{\text{Lip}(h)/\sqrt{\lambda_k}} + \min\{D_{n,j}, \sqrt{\tau(t-s)}\sqrt{\text{Lip}(h)/\sqrt{\lambda_k}}\} \]

and for $s > t$ that

\[ |E[\bar{Y}_{j,s}\bar{Y}_{k,t}] - E[\Phi_j(X_s)\Phi_k(X_t)]| \]

\[ \leq D_{n,k}\sqrt{\tau(s-t)}\sqrt{\text{Lip}(h)/\sqrt{\lambda_j}} + \min\{D_{n,k}, \sqrt{\tau(s-t)}\sqrt{\text{Lip}(h)/\sqrt{\lambda_j}}\}. \]

Therefore, we obtain by majorized convergence that

\[ R_{n,4} \xrightarrow{n \to \infty} 0. \tag{6.37} \]

The assertion follows now from (6.33), (6.34), (6.36) and (6.37).

\[ \square \]

**Lemma 6.2.** Suppose that the assumptions (A2) and (B1) hold and let $\bar{Y}_{l,t}$ be defined as in the proof of Lemma 6.1 above. Then, for $s \leq t \leq u \leq v$,

\[ |\text{cum}(\bar{Y}_{j,s}\bar{Y}_{j,t}, \bar{Y}_{j,u}\bar{Y}_{j,v})| \leq C (12M_n^2 + 3) \sqrt{\tau(\max\{t-s, u-t, v-u\})}, \]

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

\[ \text{Proof.} \] Let $r := \max\{t-s, u-t, v-u\}$. Recall that \text{cum}(\bar{Y}_{j,s}\bar{Y}_{j,t}, \bar{Y}_{j,u}\bar{Y}_{j,v}) = E[\bar{Y}_{j,s}\bar{Y}_{j,t}\bar{Y}_{j,u}\bar{Y}_{j,v}] - E[\bar{Y}_{j,s}\bar{Y}_{j,t}]E[\bar{Y}_{j,u}\bar{Y}_{j,v}] - E[\bar{Y}_{j,s}\bar{Y}_{j,u}]E[\bar{Y}_{j,t}\bar{Y}_{j,v}] - E[\bar{Y}_{j,s}\bar{Y}_{j,v}]E[\bar{Y}_{j,t}\bar{Y}_{j,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 \( \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v} \) and \( \bar{Y}_{j,v} \) such that \( \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v} \)' is independent of \( \bar{Y}_{j,v} \), and \( \sqrt{E(\bar{Y}_{j,v} - \bar{Y}_{j,v})^2} \leq C \sqrt{\tau(r)} \), for \( w \in \{ t, u, v \} \). Since \( |\bar{Y}_{j,v}| \leq 2M_n \) and \( E\bar{Y}_{j,v}^2 \leq E\Phi_{j,v}(X_{w}) = 1 \), we obtain

\[
\begin{align*}
|E[\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v}]| &= |E[\bar{Y}_{j,v}, (\bar{Y}_{j,t} - \bar{Y}_{j,u})\bar{Y}_{j,v}]| \\
&\leq \sqrt{E\bar{Y}_{j,v}^2} \sqrt{E(\bar{Y}_{j,v} - \bar{Y}_{j,v})^2} \leq C \sqrt{\tau(r)}.
\end{align*}
\]

Furthermore, we have for \( w \in \{ t, u, v \} \) that

\[
|E\bar{Y}_{j,v} \bar{Y}_{j,w}| \leq |E[\bar{Y}_{j,v} (\bar{Y}_{j,w} - \bar{Y}_{j,w})]| \leq \sqrt{E\bar{Y}_{j,v}^2} \sqrt{E(\bar{Y}_{j,w} - \bar{Y}_{j,w})^2} \leq C \sqrt{\tau(r)},
\]

which implies that

\[
|\text{cum}(\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v})| \leq C \sqrt{\tau(r)} (12M_n^2 + 3).
\]

Case b: \( u - t = r \)

Here we choose random variables \( \bar{Y}_{j,u} \) and \( \bar{Y}_{j,v} \) such that \( \bar{Y}_{j,u}, \bar{Y}_{j,v} \)' is independent of \( \bar{Y}_{j,v} \), and \( \sqrt{E(\bar{Y}_{j,v} - \bar{Y}_{j,v})^2} \leq C \sqrt{\tau(r)} \) for \( w \in \{ u, v \} \).

Then

\[
\begin{align*}
|E[\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v}]| &= |E[\bar{Y}_{j,v}, \bar{Y}_{j,t}]| E[\bar{Y}_{j,u}, \bar{Y}_{j,v}]| \\
&\leq \sqrt{E\bar{Y}_{j,v}^2} \sqrt{E(\bar{Y}_{j,v} - \bar{Y}_{j,u})^2} \leq C \sqrt{\tau(r)}.
\end{align*}
\]

Furthermore, we obtain analogously to (6.39)

\[
|E\bar{Y}_{j,v} \bar{Y}_{j,w}| \leq C \sqrt{\tau(r)} (8M_n^2 + 2),
\]

which yields that

\[
|\text{cum}(\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v})| \leq C \sqrt{\tau(r)} (8M_n^2 + 2).
\]

Case c: \( v - u = r \)

In this case we choose a random variable \( \bar{Y}_{j,v} \) with the same distribution as \( \bar{Y}_{j,v} \) and independent of \( \bar{Y}_{j,v} \) such that \( \sqrt{E(\bar{Y}_{j,v} - \bar{Y}_{j,v})^2} \leq C \sqrt{\tau(r)} \). We have

\[
\begin{align*}
|E[\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v}]| &= |E[\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v}]| \\
&= C \sqrt{\tau(r)} \sqrt{E\bar{Y}_{j,v}^2} 4M_n^2.
\end{align*}
\]

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

\[
|\text{cum}(\bar{Y}_{j,v}, \bar{Y}_{j,t}, \bar{Y}_{j,u}, \bar{Y}_{j,v})| \leq C \sqrt{\tau(r)} (4M_n^2 + 3).
\]

The assertion of the lemma follows from (6.40), (6.41) and (6.43).

\[\square\]

**Lemma 6.3.** Suppose that the prerequisites of Theorem 3.1 are satisfied. Then condition 3.1 of Corollary 6.1 holds true for \( (X_{n,s}) \) defined in (6.11) of the proof of Theorem 3.1.
Proof. Let $\tilde{X}_{n,t_1}$ be a copy of $X_{n,t_1}$ that is independent of $(W^*_u)_{s \leq L_n(t_1 - r)}$. Then, because of $\|g\|_\infty \leq 1$,

$$P\left(\left|\text{cov}^*(g(X^*_{n,s_1}, \ldots, X^*_{n,s_u}), X^*_{n,s_j}, X^*_{n,t_1})\right| \leq \left( E^*\left[ X^*_{n,s_j}\right]^2 + E^*\left[ X^*_{n,t_1}\right]^2 + L_n/n \right) \theta_r, \quad \forall u \in \mathbb{N}, 1 \leq s_1 < \cdots < s_u < s_1 + r = t_1 \leq n/L_n \right)$$

\[
\geq P\left( \sqrt{\frac{E^*\left[ X^*_{n,t_1,k} - \tilde{X}^*_{n,t_1,k}\right]^2}{\frac{L_n}{n}}} \leq \frac{\theta_r}{2} \sqrt{\frac{L_n}{n}}, \quad t_1 = r + 1, \ldots, n/L_n \right).
\]

To show that the latter probability tends to one as $n \to \infty$, we prove that, on the one hand,

$$E E^*\left[ X^*_{n,t_1,k} - \tilde{X}^*_{n,t_1,k}\right]^2 \leq C \zeta^r \frac{L_n}{n} \tag{6.44}$$

for some $C < \infty$, $\zeta \in (0, 1)$, and $\delta > 0$ uniformly in $t_1$ and, on the other hand,

$$\sum_{t_1=r+1}^{k_n} \left| E^*\left[ X^*_{n,t_1,k} - \tilde{X}^*_{n,t_1,k}\right]^2 - E E^*\left[ X^*_{n,t_1,k} - \tilde{X}^*_{n,t_1,k}\right]^2 \right| \to 0. \tag{6.45}$$

We specialize $\tilde{X}^*_{n,t_1,k} = n^{-1} \sum_{u \in I_{n,t_1}} \Phi_k(X_u)\tilde{W}_u^*$, where $(\tilde{W}_u^*)_{u \in I_{n,t_1}}$ is a copy of $(W^*_u)_{u \in I_{n,t_1}}$ that is independent of $(W^*_u)_{s \leq (t-r)L_n}$ and of $(X_u)_{u=1}^n$ and such that $E^*[\tilde{W}_u^* - W_u^*] \leq K \zeta^r L_n/n$ for all $u \in I_{n,t_1}$. (This procedure might require an enlargement of the underlying probability space.) Then

$$E E^*\left[ X^*_{n,t_1,k} - \tilde{X}^*_{n,t_1,k}\right]^2 \leq \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \left| E(\Phi_k(X_u)\Phi_k(X_v)) E^*([W_u^* - \tilde{W}_u^*][W_v^* - \tilde{W}_v^*]) \right|.$$

W.l.o.g. we only take the case $u < v$ into further consideration and obtain (6.44) from

$$\frac{1}{n} \sum_{u,v \in I_{n,t_1}} \left| E(\Phi_k(X_u)\Phi_k(X_v)) E^*([W_u^* - \tilde{W}_u^*][W_v^* - \tilde{W}_v^*]) \right|$$

$$= \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \left| E(\Phi_k(X_u)\Phi_k(X_v) - \tilde{X}_v) E^*([W_u^* - \tilde{W}_u^*][W_v^* - \tilde{W}_v^*]) \right|$$

$$\leq \sqrt{\frac{\text{Lip}(h) K}{\lambda_k}} \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \sqrt{r(u-v)} \left( E^*[W_u^* - \tilde{W}_u^*]^{1/(1-\delta)}[W_v^* - \tilde{W}_v^*] \right)^{1-\delta} \zeta^r L_n/n$$

$$\leq C \zeta^r \frac{L_n}{n},$$
where $\tilde{X}_v$ denotes a copy of $X_v$ that is independent of $X_u$. For the verification of (6.45), we define $R_{n,u,v} := E^*[|W^*_u - \tilde{W}^*_u||W^*_v - \tilde{W}^*_v|]$ and decompose

\[
\left| E^* \left[ X^*_{n,t,k} - \tilde{X}^*_{n,t,k} \right]^2 - EE^* \left[ X^*_{n,t,k} - \tilde{X}^*_{n,t,k} \right]^2 \right| \\
\leq \left| \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \{ \Phi_j(X_u) \Phi_k(X_v) - Y_{j,u} Y_{k,v} \} R_{n,u,v} \right| \\
+ \left| \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \{ Y_{j,u} Y_{k,v} - \tilde{Y}_{j,u} \tilde{Y}_{k,v} \} R_{n,u,v} \right| \\
+ \left| \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \{ \tilde{Y}_{j,u} \tilde{Y}_{k,v} - E[\tilde{Y}_{j,u} \tilde{Y}_{k,v}] \} R_{n,u,v} \right| \\
+ \left| \frac{1}{n} \sum_{u,v \in I_{n,t_1}} \{ E[\tilde{Y}_{j,u} \tilde{Y}_{k,v}] - E[\Phi_j(X_u) \Phi_k(X_v)] \} R_{n,u,v} \right| =: S_{n,t_1,1} + S_{n,t_1,2} + S_{n,t_1,3} + S_{n,t_1,4},
\]

using the notion of Lemma 6.1. We obtain

\[
\sum_{t_1=r+1}^{k_n} S_{n,t_1,1} = o_P(1) \tag{6.46}
\]

from (6.33). Moreover, a suitable choice of $M^2_n = o(L_n)$ yields

\[
\sum_{t=r+1}^{k_n} S_{n,t_1,2} = o(1) \frac{L_n}{M^2_n} \epsilon^{\alpha \epsilon L_n/L_n} + O_P \left( \frac{\sqrt{L_n}}{M^2_n} \epsilon^{\alpha \epsilon L_n/L_n} \right) = o_P(1). \tag{6.47}
\]

Choosing $(L_n)_n$ appropriately, we get

\[
\sum_{t=r+1}^{k_n} S_{n,t_1,3} = O_P \left( \frac{M_n}{L_n} + \frac{n}{L_n} \epsilon^{\alpha \epsilon L_n/L_n} \right) = o_P(1). \tag{6.48}
\]

Finally, we get $\sum_{t=r+1}^{k_n} S_{n,t_1,4} = o_P(1)$ and we can deduce (6.45) from the latter result in conjunction with (6.46), (6.47), and (6.48).

Acknowledgment. This research was funded by the German Research Foundation DFG, project NE 606/2-2. The authors thank two anonymous referees for their helpful comments that led to a significant improvement of the paper.

REFERENCES


