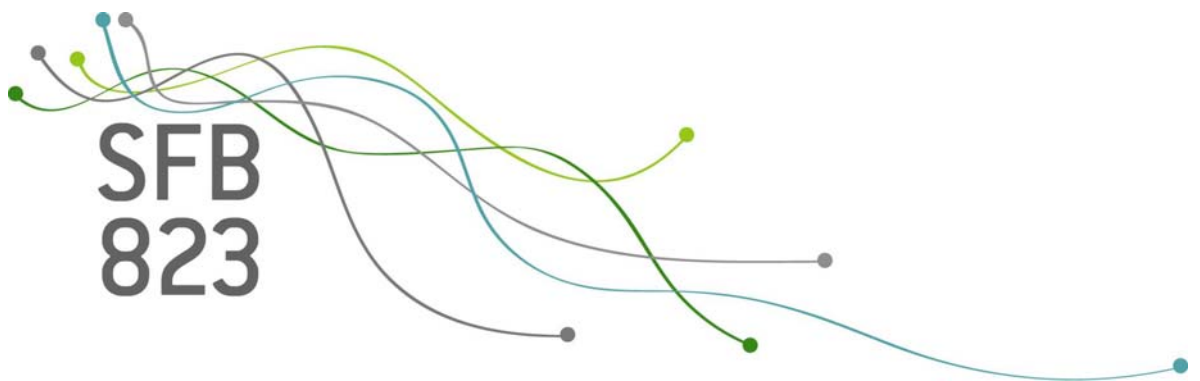


SFB  
823

# Bootstrap for $U$ -Statistics: A new approach

Olimjon Sh. Sharipov, Johannes Tewes,  
Martin Wendler

Nr. 15/2015



Discussion Paper



# BOOTSTRAP FOR $U$ -STATISTICS: A NEW APPROACH

OLIMJON SH. SHARIPOV, JOHANNES TEWES, MARTIN WENDLER

ABSTRACT. Bootstrap for nonlinear statistics like  $U$ -statistics of dependent data has been studied by several authors. This is typically done by producing a bootstrap version of the sample and plugging it into the statistic. We suggest an alternative approach of getting a bootstrap version of  $U$ -statistics, which can be described as a compromise between bootstrap and subsampling. We will show the consistency of the new method and compare its finite sample properties in a simulation study.

## 1. INTRODUCTION

In many statistical applications, the asymptotic limit distribution cannot be used to construct tests or confidence intervals, as the limit might be unknown or dependent on unknown parameters, which are hard to estimate. Bootstrap versions of the statistical estimator provide a nonparametric alternative, so establishing the consistency of such bootstrap procedure is important. In the case of independent and identically distributed (iid) random variables, the main idea of the bootstrap consists of replacing the original sample  $X_1, \dots, X_n$  of observations with unknown marginal distribution function  $F(x)$  by a new iid sample  $X_1^*, \dots, X_n^*$  with the marginal distribution function  $F_n(x)$  which is the empirical distribution function constructed from the original sample (see Efron [8]). As a bootstrap version of a statistic  $T(X_1, \dots, X_n)$  one can take the plug-in version  $T(X_1^*, \dots, X_n^*)$ . The validity of bootstrap for the sample mean of iid observations was first established by Bickel and Freedman [2] and Singh [21]. It is well-known that in some cases the bootstrap method provides a better approximation to the distribution of the statistic than, for instance, normal approximation (see Hall [10]). Especially when sample size is relatively small, bootstrap has some preferences.

In the case of dependent observations this idea of getting a new sample does not work since a new iid sample does not capture the dependence structure, see Singh [21]. Therefore several so-called blockwise bootstrap methods of getting a new sample were introduced (for overview of blockwise bootstrap methods see Lahiri [12]). In all these methods, we resample from blocks of  $l$  consecutive observations, so that inside the blocks the dependence structure will be kept. We will consider two of them: The circular block bootstrap method introduced by Politis and Romano [15] and nonoverlapping block bootstrap method by Carlstein [3]. In the case of circular block bootstrap method we extend the original sample to  $X_1, \dots, X_n, X_1, \dots, X_{l-1}$  where  $l = l(n)$  is a block length. To get a new sample  $X_{1,1}^*, \dots, X_{1,m}^*$  we choose randomly and independently  $l$  consecutive observations of the sample  $m = \lfloor \frac{n}{l} \rfloor$  times

---

*Date:* May 29, 2015.

*Key words and phrases.* primary 62G09, secondary 60G10, Mixing Processes,  $U$ -Statistics, Bootstrap, Subsampling.

with

$$P^*(X_{1,kl+1}^* = X_j, \dots, X_{1,(k+1)l}^* = X_{j+l-1}) := \\ P\left(X_{1,kl+1}^* = X_j, \dots, X_{1,(k+1)l}^* = X_{j+l-1} \mid X_1, \dots, X_n\right) = \frac{1}{n}$$

for  $j = 1, \dots, n, k = 0, \dots, m-1$ . Here and in what follows we denote by  $P^*, E^*, \text{Var}^*$  the conditional probability, conditional expectation and conditional variance respectively. For instance the bootstrap version of sample mean  $\bar{X}_n$  will be

$$\bar{X}_n^{1*} = \frac{1}{ml} \sum_{i=1}^{ml} X_{1,i}^*$$

In the case of nonoverlapping block bootstrap method we construct a new sample  $X_{2,1}^*, \dots, X_{2,ml}^*$  by choosing randomly and independently blocks of  $l$  consecutive observations of the sample  $X_1, \dots, X_n$  as above  $m = \lfloor \frac{n}{l} \rfloor$  times with

$$P^*(X_{2,kl+1}^* = X_{(j-1)l+1}, \dots, X_{2,(k+1)l}^* = X_{jl}) = \frac{1}{m}$$

for  $j = 1, \dots, m, k = 0, \dots, m-1$ . In this case the bootstrap version of the sample mean is

$$\bar{X}_n^{2*} = \frac{1}{ml} \sum_{i=1}^{ml} X_{2,i}^*$$

For weakly dependent observations strong consistency (almost surely convergence) of the  $\bar{X}_n^{1*}$  and  $\bar{X}_n^{2*}$  were proved in Shao and Yu [19] and Peligrad [18]. Gonçalves and White [9] and Dehling, Sharipov and Wendler [5] extended these results to functionals of mixing processes. In this paper we will concentrate on bootstrap for  $U$ -statistics of weakly dependent observations. Bootstrap for  $U$ -statistics of independent observations were studied by Bickel and Freedman [2], Arcones and Gine [1], Dehling and Mikosch [4], Leucht and Neumann [13]. Recently, Dehling and Wendler [6], Sharipov and Wendler [20] and Leucht and Neumann [14] established the consistency of bootstrap estimators for  $U$ -statistics of weakly dependent observations. In the aforementioned papers the same idea of getting bootstrap versions of  $U$ -statistics have been used, which is based on the plug-in principle. We will now introduce  $U$ -statistics. Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence of random variables with a common distribution function  $F(x)$ . For simplicity reasons we consider  $U$ -statistics of degree two i.e.:

$$U_n(h) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j)$$

where  $h : R \times R \rightarrow R$  is the symmetric and measurable kernel. The kernel  $h(x, y)$  is called degenerate, if

$$Eh(X_1, x) = 0 \quad \text{for all } x \in R.$$

If this does not hold we call the kernel  $h(x, y)$  and the corresponding  $U$ -statistics nondegenerate. In what follows we will only consider  $U$ -statistics with a nondegenerate kernel. In the case of iid observations  $X_1, \dots, X_n$  Bickel and Freedman [2] used the bootstrap sample  $X_1^*, \dots, X_n^*$  of conditional independent observations with common distribution function  $F_n(x)$ , which is the empirical distribution function

constructed by original sample. In order to get bootstrap versions of  $U$ -statistics they plugged the bootstrapped observations in  $U$ -statistics i.e.

$$U_n^*(h) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i^*, X_j^*)$$

In the case of dependent observations, the same idea was explored by Dehling, Wendler [6] using the circular block bootstrap method, the corresponding bootstrap version is

$$U_{n,1}^*(h) = \frac{2}{ml(ml-1)} \sum_{1 \leq i < j \leq ml} h(X_{1,i}^*, X_{1,j}^*),$$

while in the case of nonoverlapping block bootstrap method the corresponding bootstrap version of  $U$ -statistics is

$$U_{n,2}^*(h) = \frac{2}{ml(ml-1)} \sum_{1 \leq i < j \leq ml} h(X_{2,i}^*, X_{2,j}^*).$$

Consistency of  $U_n^*(h)$  (and  $U_{n,k}^*(h), k = 1, 2$ ) can be proved using the Hoeffding decomposition (see [11]):

$$U_n^*(h) = \theta + \frac{2}{n} \sum_{i=1}^n h_1(X_i^*) + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h_2(X_i^*, X_j^*)$$

where

$$\begin{aligned} \theta &:= Eh(X_1, X_2), \\ h_1(x) &:= Eh(x, X_2) - \theta, \\ h_2(x, y) &:= h(x, y) - h_1(x) - h_1(y) - \theta. \end{aligned}$$

In order to prove consistency of  $U_n^*(h)$  it is enough to show the convergence of bootstrapped distribution of the sample mean for the second summand of this decomposition and convergence to zero (in probability or almost surely) of the third summand. The goal of this paper is to suggest a new resampling method for  $U$ -statistics. The main idea is the following: We suggest to draw with replacement from the  $U$ -statistics calculated on subsamples instead of drawing with replacement from blocks of observations and then plug them in  $U$ -statistics. We introduce subsets  $\{1, \dots, n\}^2$  by

$$\begin{aligned} B'_1(i) &= \{(a, b) : i \leq a, b \leq i + l - 1, a < b\} \quad i = 1, \dots, n, \\ B'_2(i) &= \{(a, b) : l(i-1) + 1 \leq a, b \leq il, a < b\} \quad i = 1, \dots, m = \lfloor \frac{n}{l} \rfloor \end{aligned}$$

where  $l = l(n) \rightarrow \infty$  us called block length. Furthermore, let  $\{T_1(j), j \geq 1\}$  and  $\{T_2(j), j \geq 1\}$  be two sequences of iid random variables with distributions:

$$\begin{aligned} P(T_1(1) = i) &= \frac{1}{n}, \quad i = 1, \dots, n, \\ P(T_2(1) = i) &= \frac{1}{m}, \quad i = 1, \dots, m. \end{aligned}$$

A fixed realisations of the sample  $X_1, \dots, X_n$  we denote by  $x_1, \dots, x_n$ . We set

$$\begin{aligned}\bar{x}_1(i) &= \frac{2}{l(l-1)} \sum_{(a,b) \in B'_1(i)} h(x_a, x_b), \quad i = 1, \dots, n, \\ \bar{x}_2(i) &= \frac{2}{l(l-1)} \sum_{(a,b) \in B'_2(i)} h(x_a, x_b), \quad i = 1, \dots, m, \\ u_j^{1*} &= \sum_{i=1}^n I(T_1(j) = i) \bar{x}_1(i) = \bar{x}_1(T_1(j)), \\ u_j^{2*} &= \sum_{i=1}^m I(T_2(j) = i) \bar{x}_2(i) = \bar{x}_2(T_2(j)),\end{aligned}$$

where  $I(\cdot)$  is the indicator function. So  $\bar{x}_1(i)$  respectively  $\bar{x}_2(i)$  are the  $U$ -statistics calculated from the  $i$ -th block and  $u_j^{1*}$  and  $u_j^{2*}$  are the results of the drawing with replacement from these  $U$ -statistics. In the case of  $\bar{x}_1(i)$  we assume that  $x_k = x_{k-n}$  if  $k > n$ . As the bootstrap versions of  $U$ -statistics  $U_n(h)$  we take the following:

$$\begin{aligned}U_n^{1*}(h) &= \frac{1}{m} \sum_{i=1}^m u_j^{1*}, \\ U_n^{2*}(h) &= \frac{1}{m} \sum_{i=1}^m u_j^{2*}.\end{aligned}$$

Note that

$$\begin{aligned}E^* U_n^{1*}(h) &= \frac{1}{nl(l-1)} \sum_{i=1}^n \sum_{(a,b) \in B'_1(i)} h(x_a, x_b) \\ E^* U_n^{2*}(h) &= \frac{1}{ml(l-1)} \sum_{i=1}^m \sum_{(a,b) \in B'_2(i)} h(x_a, x_b).\end{aligned}$$

In the next sections we will state our main result: the strong consistency of the bootstrap based on  $U_n^{k*}(h)$ ,  $k = 1, 2$ . The new approach reduces the computational burden of the Monte Carlo method usually used for the bootstrap. After calculating the values  $u_j^{k*}$ , for every run of the Monte Carlo evaluation one need only  $O(\frac{n}{l})$  calculation steps, while for the plug-in method, in every run the  $U$ -statistic has to be calculated again in  $O(n^2)$  steps. As the linear parts of the plug-in bootstrap version and the new bootstrap version of a  $U$ -statistic are the same, we expect a similar behaviour.

We will give an upper bound for the mean square error (MSE) of the variances of  $U_n^{k*}(h)$ ,  $k = 1, 2$ , which suggest a choice of the block length of order  $O(\sqrt{n})$ . In section 3, we will present our simulation results. We consider the sample variance as a  $U$ -statistic and we will compare the new bootstrap approach with plug-in bootstrap and subsampling. The proofs of the main results will be given in section 4.

## 2. MAIN RESULTS

Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence of random variables with values in a separable linear space. We will assume that this sequence satisfies some form of short

range dependence. Namely we will consider strong mixing and absolute regularity conditions. Recall that strong mixing coefficients  $(\alpha(k))_{k \in \mathbb{N}}$  and absolute regularity coefficients  $(\beta(k))_{k \in \mathbb{N}}$  are defined as

$$\begin{aligned}\alpha(n) &= \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+n}^\infty, k \in \mathbb{N}\}, \\ \beta(n) &= \sup\{E(\sup |P(B/\mathcal{F}_1^k) - P(B)| : B \in \mathcal{F}_{k+n}^\infty, k \in \mathbb{N})\}\end{aligned}$$

where  $\mathcal{F}_a^b$  is the  $\sigma$ -field generated by  $X_a, \dots, X_b$ . Now we can formulate our results.

**Theorem 2.1.** *Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence of absolutely regular random variables. Assume that the following conditions hold (for some  $\delta > 0$ )*

- $\iint |h(x_1, x_2)|^{2+\delta} dF(x_1) dF(x_2) \leq M$  for some  $M < \infty$ ,
- $E|h(X_1, X_{1+k})|^{2+\delta} \leq M$  for all  $k \in \mathbb{N}$ ,
- $\beta(n) = O(n^{-\rho})$  for some  $\rho > \frac{2+\delta}{\delta}$ ,
- $l(n) \leq Cn^\epsilon$  for some  $\epsilon \in (0, 1)$ .

Then as  $n \rightarrow \infty$  for  $i = 1, 2$

$$\begin{aligned}\sup_{x \in \mathbb{R}} \left| P^* \left[ \sqrt{ml} (U_n^{i*}(h) - E^* [U_n^{i*}]) \leq x \right] - P \left[ \sqrt{n} (U_n(h) - \theta) \leq x \right] \right| &\rightarrow 0, \\ \left| \text{Var}^* \left[ \sqrt{ml} U_n^{i*}(h) \right] - \text{Var} \left[ \sqrt{n} U_n(h) \right] \right| &\rightarrow 0,\end{aligned}$$

in probability.

In the next theorem we consider strongly mixing random variables. This is a weaker assumption on the dependence, but in this case we assume that the kernel satisfies the following condition: We say that a kernel is  $\mathcal{P}$ -Lipschitz-continuous, if there exists a constant  $C > 0$  such that

$$E [|h(X, Y) - h(X', Y)| I(|X - X'| \leq \epsilon)] \leq C\epsilon$$

for every  $\epsilon > 0$ , every pair  $X$  and  $Y$  with the common distribution  $\mathcal{P}_{X_1, X_k}$  for some  $k \in \mathbb{N}$  or  $\mathcal{P}_{X_1} \times \mathcal{P}_{X_1}$  and  $X'$  and  $Y$  also with one of these common distributions. For the examples of kernels which satisfy above condition see [6], [20].

**Theorem 2.2.** *Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence of strongly mixing random variables. Assume that the following conditions hold*

- $\iint |h(x_1, x_2)|^{2+\delta} dF(x_1) dF(x_2) \leq M$ ,
- $E|h(X_1, X_{1+k})|^{2+\delta} \leq M$  for all  $k \in \mathbb{N}$ ,
- $l(n) \leq Cn^\epsilon$  for some  $\epsilon \in (0, 1)$ ,
- $h$  is  $\mathcal{P}$ -Lipschitz-continuous,
- $E|X_1|^\gamma > 0$  for some  $\gamma > 0$ ,
- $\alpha(n) = O(n^{-\rho})$  for some  $\rho > \frac{3\gamma\delta + \delta + 5\gamma + 2}{2\gamma\delta}$ .

Then the statements of Theorem 2.1 hold.

In the next theorem we give bounds for the mean squared error to give a deeper insight into the properties of bootstrap versions of  $U$ -statistics.

**Theorem 2.3.** *Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence of absolutely regular random variables. Assume that the following conditions hold:*

- $\iint |h(x_1, x_2)|^{12+\delta} dF(x_1) dF(x_2) \leq M$  for some  $\delta > 0$ ,
- $E|h(X_1, X_{1+m})|^{4+\frac{2}{3}\delta} \leq M$  for all  $m \in \mathbb{N}$ ,

- $\beta(n) = O\left(n^{-\frac{3(6+\delta')}{\delta'}}\right)$  for some  $0 < \delta' < \delta$ .

Then for  $k = 1, 2$

$$E\left(\text{Var}^*(\sqrt{ml}U_n^{k*}(h)) - \text{Var}(\sqrt{n}U_n(h))\right)^2 = O\left(\frac{1}{l} + \frac{l}{n}\right).$$

Choosing a block length of order  $1/\sqrt{n}$ , we can achieve that the mean squared error of  $\text{Var}^*(\sqrt{ml}U_n^{k*}(h))$  is of order  $O(1/\sqrt{n})$ .

### 3. SIMULATION

We consider the estimator  $\sigma_n^2$  for the variance  $\sigma^2 = \text{Var}(X_1)$  which is a  $U$ -statistic with the kernel  $h(x, y) = \frac{1}{2}(x - y)^2$ :

$$\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \frac{1}{2} (X_i - X_j)^2.$$

$(X_n)_{n \in \mathbb{N}}$  is a stationary, Gaussian, autoregressive process with  $X_n = \alpha X_{n-1} + \epsilon_n$ , where  $(\epsilon_n)_{n \in \mathbb{N}}$  is a sequence of iid standard normal random variables and  $\alpha \in \{0.2, 0.4, 0.6\}$ . We will compare three methods for constructing confidence intervals:

- The circular plug-in bootstrap (already used in Dehling, Wendler [6]).
- The new bootstrap version  $U_n^{1*}$ , the drawing with replacement from  $U$ -statistics calculated on subsamples.
- Subsampling: The estimator of the unknown distribution is given by the empirical distribution function of the  $U$ -statistics calculated on subsamples, see Politis and Romano [16].

For each combination of construction method, AR-coefficient, sample size ( $n = 50, 100, 200, 400$ ) and block length, we have simulated 10.000 samples and evaluated the empirical probability of the 95% confidence interval to cover the true parameter. The two bootstrap version are evaluated by the Monte Carlo method with 1000 times drawing with replacement. The results are summarized in the table below.

In general, we draw the following conclusions from the simulation study: All three methods lead to confidence intervals with a coverage probability lower than the nominal confidence level of 95%. The performance of the subsampling is worse than the performance of two bootstrap methods in all situations. If the dependence in the AR-process is weak ( $\alpha = 0.2$ ), both bootstrap methods lead to comparable results, while for stronger dependence ( $\alpha = 0.4, 0.6$ ), the plug-in bootstrap has a better performance.

For the plug-in bootstrap, the block length can be chosen smaller than for the new bootstrap method or for subsampling, especially in the case of stronger dependence ( $\alpha = 0.4, 0.6$ ). In Figur 1 below, we give a more detailed picture of the coverage probabilities for different block lengths in the case  $n = 200$ ,  $\alpha = 0.4$ .

## 4. PROOFS OF THE THEOREMS

**4.1. Preliminary results.** In this section we formulate some necessary results that will be used in the proofs of our theorems.



$n$	$\alpha$ $l$	0.2	0.4	0.6
50	3	0.876/0.895/0.827	0.863/0.791/0.693	0.794/0.600/0.569
	5	0.865/0.884/0.823	0.838/0.819/0.737	0.765/0.677/0.583
	7	0.862/0.871/0.814	0.823/0.818/0.753	0.762/0.718/0.614
	10	0.835/0.849/0.787	0.808/0.814/0.743	0.757/0.730/0.607
100	3	0.908/0.922/0.823	0.894/0.817/0.730	0.852/0.600/0.507
	5	0.911/0.910/0.852	0.896/0.838/0.770	0.851/0.682/0.617
	7	0.909/0.909/0.849	0.874/0.857/0.793	0.853/0.737/0.676
	10	0.890/0.902/0.851	0.880/0.859/0.815	0.849/0.767/0.716
200	5	0.926/0.930/0.873	0.908/0.859/0.813	0.883/0.684/0.620
	7	0.925/0.926/0.881	0.905/0.870/0.831	0.891/0.738/0.704
	10	0.924/0.920/0.889	0.911/0.885/0.857	0.887/0.805/0.761
	15	0.910/0.910/0.884	0.902/0.893/0.849	0.885/0.835/0.800
400	7	0.927/0.932/0.896	0.927/0.887/0.855	0.920/0.748/0.711
	10	0.931/0.927/0.906	0.924/0.901/0.873	0.920/0.814/0.777
	15	0.923/0.929/0.901	0.923/0.904/0.882	0.902/0.847/0.828
	20	0.924/0.918/0.906	0.902/0.909/0.887	0.904/0.866/0.849

TABLE 1. Simulated coverage probabilities for plug-in bootstrap/new bootstrap/ subsampling, sample size  $n = 50, 100, 200, 400$ , AR-coefficient  $\alpha = 0.2, 0.4, 0.6$ , block length  $l = 3, 5, 7, 10, 15, 20$

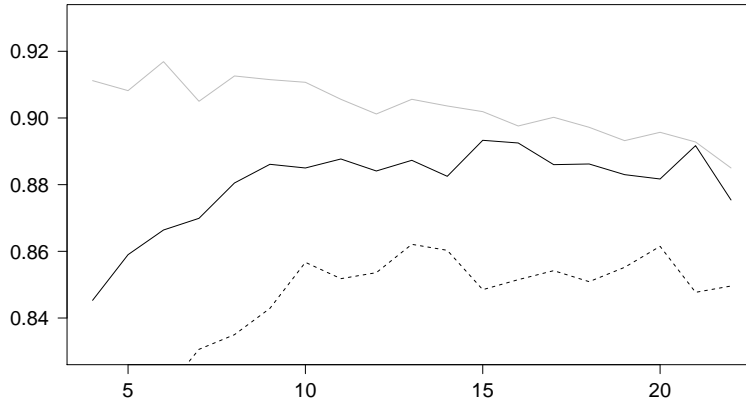


FIGURE 1. Simulated coverage probabilities for plug-in bootstrap (grey), new bootstrap (black), subsampling (dashed) as a function of the block length  $l$ , sample size  $n = 200$ , AR-coefficient  $\alpha = 0.4$

**Lemma 4.1.** (Dehling, Wendler [7]) Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence and  $h(x, y)$  a degenerate kernel satisfying  $\iint |h(x_1, x_2)|^{12+\delta} dF(x_1) dF(x_2) \leq M$  for some  $\delta > 0$  and  $E |h(X_1, X_{1+m})|^{4+\frac{2}{3}\delta} \leq M$  for all  $m \in \mathbb{N}$ . Let be  $\tau \geq 0$  such that one of the following two conditions holds:

- (1)  $(X_n)_{n \in \mathbb{N}}$  is absolutely regular and  $\sum_{k=0}^n k \beta^{\frac{\delta}{2+\delta}}(k) = O(n^\tau)$ .

- (2)  $(X_n)_{n \in \mathbb{N}}$  is strongly mixing,  $E|X_1|^\gamma < \infty$  for a  $\gamma > 0$ ,  $h$  is  $\mathcal{P}$ -Lipschitz-continuous with constant  $C > 0$  and  $\sum_{k=0}^n k \alpha^{\frac{2\gamma\delta}{\gamma\delta + \delta + \delta\gamma + 2}}(k) = O(n^\tau)$ .

Then:

$$\sum_{i_1, i_2, i_3, i_4=1}^n |E[h_2(X_{i_1}, X_{i_2})h_2(X_{i_3}, X_{i_4})]| = O(n^{2+\tau}),$$

**Lemma 4.2.** (Yoshihara [23]) Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary sequence of absolutely regular random variables. Assume that the the following conditions hold (for some  $\delta > 0$ ) and

$$\begin{aligned} \iint |h(x_1, x_2)|^{4+\delta} dF(x_1) dF(x_2) &\leq M, \\ E|h(X_1, X_{1+k})|^{4+\delta} &\leq M \text{ for all } k \in \mathbb{N} \\ \beta(n) &= O\left(n^{-\frac{3(4+\delta')}{\delta'}}\right) \text{ for some } 0 < \delta' < \delta. \end{aligned}$$

Then

$$E(n^2 U_n^4(h_2)) = O(n^{-1-\eta}), \quad \eta = \min\left\{12 \frac{\delta - \delta'}{\delta'(4 + \delta)}, 1\right\}$$

**Lemma 4.3.** (Yokoyama [22]) Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary strongly mixing sequence of random variables with  $EX_1 = \mu$  and  $\|X_1\|_{2+\delta} < \infty$  for some  $0 < \delta \leq \infty$ . Suppose that  $2 \leq s < 2 + \delta$  and

$$\sum_{n=1}^{\infty} n^{\frac{s}{2}-1} (\alpha(n))^{(2+\delta-s)/(2+\delta)} < \infty.$$

Then there exists a constant  $C$  depending only on  $s, \delta$  and the mixing coefficients  $(\alpha(n))_{n \in \mathbb{N}}$  such that

$$E \left| \sum_{i=1}^n (X_i - \mu) \right|^s \leq Cn^{s/2}.$$

**4.2. Proof of Theorem 2.1.** Let us introduce blocks of indices

$$\begin{aligned} B_1(i) &= \{i, \dots, i+l-1\}, \quad i = 1, \dots, n, \\ B_2(i) &= \{(i-1)l+1, \dots, il\}, \quad i = 1, \dots, m. \end{aligned}$$

Note that the blocks  $\{B_1(i), i = 1, \dots, n\}$  and  $\{B_2(i), i = 1, \dots, m\}$  correspond to the circular and nonoverlapping blocking methods, respectively. In the circular blocking method instead of the sample  $X_1, \dots, X_n$  we consider the completed sample  $X_1, \dots, X_n, X_1, \dots, X_{l-1}$ . A fixed realization of the sample  $X_1, \dots, X_n$  we denote by  $x_1, \dots, x_n$  and set

$$\begin{aligned} \bar{x}_{1,1}(i) &= \frac{1}{l} \sum_{k \in B_1(i)} x_k, \quad i = 1, \dots, n, \\ \bar{x}_{2,1}(i) &= \frac{1}{l} \sum_{k \in B_2(i)} x_k, \quad i = 1, \dots, m, \\ X_j^{1*} &= \sum_{i=1}^n I(T_1(j) = i) \bar{x}_{1,1}(i), \\ X_j^{2*} &= \sum_{i=1}^m I(T_2(j) = i) \bar{x}_{2,1}(i), \end{aligned}$$

$$\begin{aligned}\bar{X}_{n,m}^{1*} &= \frac{1}{m} \sum_{i=1}^m X_i^{1*}, \\ \bar{X}_{n,m}^{2*} &= \frac{1}{m} \sum_{i=1}^m X_i^{2*}.\end{aligned}$$

First we will prove the statement of the theorem for  $i = 1$  (circular bootstrap). Using Hoeffding decomposition we have

$$\begin{aligned}(U_n^{1*}(h) - E^*[U_n^{1*}]) &= \frac{1}{m} \sum_{j=1}^m (u_j^{1*} - E^*U_n^{1*}(h)) \\ &= \frac{1}{m} \sum_{j=1}^m \left[ \theta + \frac{2}{l} \sum_{k \in B_1(T_1(j))} h_1(x_k) + \frac{2}{l(l-1)} \sum_{(a,b) \in B'_1(T_1(j))} h_2(x_a, x_b) \right. \\ &\quad \left. - \left( \theta + \frac{1}{n} \sum_{i=1}^n \frac{2}{l} \sum_{k \in B_1(i)} h_1(x_k) + \frac{1}{n} \sum_{i=1}^n \frac{2}{l(l-1)} \sum_{(a,b) \in B'_1(i)} h_2(x_a, x_b) \right) \right] \\ &= \frac{1}{m} \sum_{j=1}^m \left[ \frac{2}{l} \left( \sum_{k \in B_1(T_1(j))} h_1(x_k) - \frac{1}{n} \sum_{r=1}^n h_1(x_r) \right) \right. \\ &\quad \left. + \frac{2}{l(l-1)} \left( \sum_{(a,b) \in B'_1(T_1(j))} h_2(x_a, x_b) - \frac{1}{n} \sum_{r=1}^n \frac{2}{l(l-1)} \sum_{(a,b) \in B'_1(r)} h_2(x_a, x_b) \right) \right] \\ &= \frac{2}{ml} \sum_{j=1}^m \sum_{k \in B_1(T_1(j))} (h_1(x_k) - \frac{1}{n} \sum_{r=1}^n h_1(x_r)) \\ &\quad + \frac{2}{ml(l-1)} \sum_{j=1}^m \sum_{(a,b) \in B'_1(T_1(j))} (h_2(x_a, x_b) - \frac{1}{n} \sum_{r=1}^n \sum_{(a,b) \in B'_1(r)} h_2(x_a, x_b))\end{aligned}$$

Using the latter

$$\begin{aligned}\sqrt{ml} (U_n^{1*}(h) - E^*[U_n^{1*}]) &= \frac{2}{\sqrt{ml}} \sum_{j=1}^m \left[ \sum_{k \in B_1(T_1(j))} h_1(x_k) - \frac{1}{n} \sum_{i=1}^n h_1(x_i) \right] \\ &+ \frac{2}{\sqrt{ml}(l-1)} \sum_{j=1}^m \left[ \sum_{(a,b) \in B'_1(T_1(j))} (h_2(x_a, x_b) - \frac{1}{n} \sum_{r=1}^n \sum_{(a,b) \in B'_1(r)} h_2(x_a, x_b)) \right] \\ &= I_n + II_n.\end{aligned}$$

In order to prove the first statement of Theorem 2.1, we first show that  $II_n \rightarrow 0$  as  $n \rightarrow \infty$ . Note that

$$\begin{aligned}\text{Var}^* II_n &= \frac{4}{nl(l-1)^2} \sum_{i=1}^n \left( \sum_{(a,b) \in B'_1(i)} \left( h_2(x_a, x_b) - \frac{1}{n} \sum_{i=1}^n \sum_{(a,b) \in B'_1(i)} h_2(x_a, x_b) \right) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left( \frac{2}{\sqrt{l}(l-1)} \sum_{(a,b) \in B'_1(i)} h_2(x_a, x_b) \right)^2 - \left( \frac{1}{n} \sum_{i=1}^n \frac{2}{\sqrt{l}(l-1)} \sum_{(a,b) \in B'_1(i)} h_2(x_a, x_b) \right)^2.\end{aligned}$$

By Lemma 4.1 we have as  $n \rightarrow \infty$

$$E\left(\frac{2}{\sqrt{l}(l-1)} \sum_{(a,b) \in B'_1(i)} h_2(X_a, X_b)\right)^2 \rightarrow 0$$

and consequently  $\text{Var}^* I_n \rightarrow 0$  in probability as  $n \rightarrow \infty$ . With the Chebyshev inequality, it follows that  $I_n \rightarrow 0$  in conditional probability.

It remains to show the convergence of  $I_n$ . By Theorem 3.2 of Lahiri [12] we have as  $n \rightarrow \infty$

$$\sup_{x \in \mathbb{R}} \left| P^*(I_n \leq x) - P\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n h(X_i) \leq x\right) \right| \rightarrow 0.$$

in probability. Furthermore, we have  $\text{Var}(\sqrt{n}U_n(h_2)) \rightarrow 0$  by Lemma 4.1 and can conclude that

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n h(X_i) \leq x\right) - P\left(\sqrt{n}(U_n(h) - \theta) \leq x\right) \right| \rightarrow 0.$$

Using Slutsky's lemma and  $I_n \rightarrow 0$  in probability, we arrive at at the first statement of the Theorem:

$$\sup_{x \in \mathbb{R}} \left| P^* \left[ \sqrt{ml} (U_n^{i*}(h) - E_*[U_n^{i*}]) \leq x \right] - P \left[ \sqrt{n} (U_n(h) - \theta) \leq x \right] \right| \rightarrow 0.$$

For the second statement, we make use of Theorem 3.2 of Lahiri [12] again, which also states that

$$\left| \text{Var}^* I_n - \text{Var}\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n h(X_i)\right) \right| \rightarrow 0$$

This together with  $\text{Var}^* I_n \rightarrow 0$  in probability and  $\text{Var}(\sqrt{n}U_n(h_2)) \rightarrow 0$  (see Lemma 4.1) leads to

$$\left| \text{Var}^* \left[ \sqrt{ml} U_n^{i*}(h) \right] - \text{Var} \left[ \sqrt{n} U_n(h) \right] \right| \rightarrow 0$$

in probability as  $n \rightarrow \infty$ .

In the case of  $i = 2$  (nonoverlapping bootstrap) analogously to previous one we have

$$\begin{aligned} & \sqrt{ml} (U_n^{2*}(h) - E_*[U_n^{2*}]) \\ &= \frac{2}{\sqrt{ml}} \sum_{j=1}^m \sum_{i=1}^m I(T_2(j) = i) \sum_{k \in B_1(i)} (h_1(x_k) - \frac{1}{ml} \sum_{i=1}^{ml} h_1(x_i)) \\ &+ \frac{2}{\sqrt{ml}(l-1)} \sum_{j=1}^m \sum_{(a,b) \in B'_1(T_2(j))} (h_2(x_a, x_b) - \frac{1}{m} \sum_{i=1}^m \sum_{(a,b) \in B'_1(i)} h_2(x_a, x_b)) = \\ &= I'_n + II'_n. \end{aligned}$$

Theorem 3.2 of Lahiri [12] implies as  $n \rightarrow \infty$

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| P^*(I'_n \leq x) - P\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n h(X_i) \leq x\right) \right| \rightarrow 0 \\ & \left| \text{Var}^* I'_n - \text{Var}\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n h(X_i)\right) \right| \rightarrow 0 \end{aligned}$$

in probability as  $n \rightarrow \infty$ . It remains to show that  $\text{Var}^* II'_n \rightarrow 0$  in probability, the rest of the proof can be done along the lines of the case  $i = 1$  (circular bootstrap). We have

$$\begin{aligned} \text{Var}^* II'_n &= \frac{4}{ml(l-1)^2} \sum_{i=1}^m \left( \sum_{(a,b) \in B'_2(i)} (h_2(x_a, x_b) - \frac{1}{m} \sum_{i=1}^m \sum_{(a,b) \in B'_2(i)} h_2(x_a, x_b)) \right)^2 \\ &= \frac{1}{m} \sum_{i=1}^m \left( \frac{2}{\sqrt{l(l-1)}} \sum_{(a,b) \in B'_2(i)} h_2(x_a, x_b) \right)^2 - \left( \frac{1}{m} \sum_{i=1}^m \frac{2}{\sqrt{l(l-1)}} \sum_{(a,b) \in B'_2(i)} h_2(x_a, x_b) \right)^2. \end{aligned}$$

By Lemma 4.1 we have as  $n \rightarrow \infty$

$$E \left( \frac{2}{\sqrt{l(l-1)}} \sum_{(a,b) \in B'_2(i)} h_2(X_a, X_b) \right)^2 \rightarrow 0$$

so  $E(\text{Var}^* II'_n) \rightarrow 0$ .

**4.3. Proof of Theorem 2.2.** Theorem 2.2 can be proved in the same way as Theorem 2.1 and therefore the proof is omitted.

**4.4. Proof of Theorem 2.3.** To shorten the proof, we concentrate on the case  $k = 1$  (circular bootstrap). We split the mean squared error into three parts

$$\begin{aligned} &E \left( \text{Var}^* (\sqrt{ml} U_n^{k*}(h)) - \text{Var}(\sqrt{n} U_n(h)) \right)^2 \\ &\leq 3E \left( \text{Var}^* \left( \frac{1}{\sqrt{ml}} \sum_{i=1}^{ml} h_1(X_{1,i}^*) \right) - \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n h_1(X_i) \right) \right)^2 \\ &\quad + 3E \left( \text{Var}^* \left( \frac{1}{\sqrt{ml}} \sum_{i=1}^{ml} h_1(X_{1,i}^*) \right) - \text{Var}^* (\sqrt{ml} U_n^{k*}(h)) \right)^2 \\ &\quad + 3E \left( \text{Var}(\sqrt{n} U_n(h)) - \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n h_1(X_i) \right) \right)^2 = 3I_n + 3II_n + 3III_n \end{aligned}$$

For the first summand, we know by Theorem 3.1 of Politis and White [17] that

$$I_n = E \left( \text{Var}^* \left( \frac{1}{\sqrt{ml}} \sum_{i=1}^{ml} h_1(X_{1,i}^*) \right) - \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n h_1(X_i) \right) \right)^2 = O\left(\frac{1}{l^2} + \frac{l}{n}\right).$$

For the second summand, we make use of the equation  $a^2 - b^2 = (a+b)(a-b)$  and the Hölder-inequality (first for the bootstrap expectation and then for the

unconditional expectation) to obtain

$$\begin{aligned}
H_n &= E\left(\text{Var}^*\left(\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*)\right) - \text{Var}^*(\sqrt{ml}U_n^{k*}(h))\right)^2 \\
&= E\left(E^*\left(\left(\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) - E^*\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) - \sqrt{ml}U_n^{k*}(h) + E^*\sqrt{ml}U_n^{k*}(h)\right)\right.\right. \\
&\quad \left.\left.\left(\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) - E^*\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) + \sqrt{ml}U_n^{k*}(h) - E^*\sqrt{ml}U_n^{k*}(h)\right)\right)\right)^2 \\
&\leq E\left(E^*\left(\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) - E^*\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) - \sqrt{ml}U_n^{k*}(h) + E^*\sqrt{ml}U_n^{k*}(h)\right)^2\right. \\
&\quad \left.E^*\left(\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) - E^*\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) + \sqrt{ml}U_n^{k*}(h) - E^*\sqrt{ml}U_n^{k*}(h)\right)^2\right) \\
&= E\left(\text{Var}^*\left(\frac{1}{\sqrt{ml}(l-1)}\sum_{j=1}^m\sum_{i=1}^nI(T_1(j)=i)\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)\right. \\
&\quad \left.\text{Var}^*\left(2\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) + \frac{1}{\sqrt{ml}(l-1)}\sum_{j=1}^m\sum_{i=1}^nI(T_1(j)=i)\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)\right) \\
&\leq \left(E\left(\text{Var}^*\left(\frac{1}{\sqrt{ml}(l-1)}\sum_{j=1}^m\sum_{i=1}^nI(T_1(j)=i)\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)\right)^2\right)^{\frac{1}{2}} \\
&\quad \left(E\left(\text{Var}^*\left(2\frac{1}{\sqrt{ml}}\sum_{i=1}^{ml}h_1(X_{1,i}^*) + \frac{1}{\sqrt{ml}(l-1)}\sum_{j=1}^m\sum_{(a,b)\in B'_1(T_1(j))}h_2(X_a,X_b)\right)\right)^2\right)^{\frac{1}{2}} \\
&= A_n \times B_n.
\end{aligned}$$

We will treat these two factors separately, starting with  $A_n$ . By the definition of the bootstrap procedure and the stationarity of the sequence, we get the following:

$$\begin{aligned}
A_n &= \left(E\left(\text{Var}^*\left(\frac{1}{\sqrt{ml}(l-1)}\sum_{j=1}^m\sum_{i=1}^nI(T_1(j)=i)\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)\right)^2\right)^{\frac{1}{2}} \\
&\leq \left(EE^*\left(\frac{1}{\sqrt{ml}(l-1)}\sum_{j=1}^m\sum_{i=1}^nI(T_1(j)=i)\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)^4\right)^{\frac{1}{2}} \\
&= \left(E\frac{1}{l^2(l-1)^4}\frac{1}{n}\sum_{i=1}^n\left(\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)^4\right)^{\frac{1}{2}} \\
&= \left(\frac{1}{l^2(l-1)^4}E\left(\sum_{(a,b)\in B'_1(i)}h_2(X_a,X_b)\right)^4\right)^{\frac{1}{2}} = O\left(\frac{1}{l}\right),
\end{aligned}$$

as we know from Lemma 3 of Yoshihara [23] that

$$E\left(\sum_{1\leq i<j\leq k}h_2(X_i,X_j)\right)^4 = O(l^4).$$

With the help of the inequality  $(a + b)^2 \leq 2a^2 + 2b^2$ , we split the second summand into two parts:

$$\begin{aligned} B_n &= \left( E \left( \text{Var}^* \left( 2 \frac{1}{\sqrt{ml}} \sum_{i=1}^{ml} h_1(X_{1,i}^*) + \frac{1}{\sqrt{ml}(l-1)} \sum_{j=1}^m \sum_{(a,b) \in B'_1(T_1(j))} h_2(X_a, X_b) \right) \right)^2 \right)^{\frac{1}{2}} \\ &\leq 16 \left( E \left( \text{Var}^* \left( \frac{1}{\sqrt{ml}} \sum_{i=1}^{ml} h_1(X_{1,i}^*) \right) \right)^2 \right)^{\frac{1}{2}} + 4A_n = 16B'_n + 4A_n, \end{aligned}$$

so it suffices to study  $B'_n$ . We use again stationarity and the definition of the bootstrap:

$$\begin{aligned} B'_n &= \left( E \left( \text{Var}^* \left( \frac{1}{\sqrt{ml}} \sum_{i=1}^{ml} h_1(X_{1,i}^*) \right) \right)^2 \right)^{\frac{1}{2}} = \left( E \left( \text{Var}^* \left( \frac{1}{\sqrt{l}} \sum_{i=1}^l h_1(X_{1,i}^*) \right) \right)^2 \right)^{\frac{1}{2}} \\ &\leq \left( EE^* \left( \frac{1}{\sqrt{l}} \sum_{i=1}^l h_1(X_{1,i}^*) \right)^4 \right)^{\frac{1}{2}} = \left( E \frac{1}{n} \sum_{j=1}^n \left( \frac{1}{\sqrt{l}} \sum_{i=j}^{j+l-1} h_1(X_i) \right)^4 \right)^{\frac{1}{2}} \\ &= \left( E \left( \frac{1}{\sqrt{l}} \sum_{i=j}^{j+l-1} h_1(X_i) \right)^4 \right)^{\frac{1}{2}} \leq C. \end{aligned}$$

The last inequality follows from Lemma 4.3. This shows that  $II_n = O(\frac{1}{l})$ . In the same way one can show that  $III_n$  is of the same order, which completes the proof.

**Acknowledgments.** The work has been supported by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation (DFG). The authors would like to thank Herold Dehling for the interesting discussions on this topic, which inspired the new method.

#### REFERENCES

- [1] M.A. ARCONES, E. GINE, On the bootstrap for  $U$  and  $V$  statistics, *Ann. Stat.* **20** (1992) 655-674.
- [2] P.J. BICKEL, D.A. FREEDMAN, Some asymptotic theory for the bootstrap, *Ann. Stat.* **9** (1981) 1196-1217.
- [3] E. CARLSTEIN, The use of subseries values for estimating the variance of a general statistic from stationary sequence, *Ann. Stat.* **14** (1986) 1171-1179.
- [4] H. DEHLING, T. MIKOSCH, Random quadratic forms and the bootstrap for  $U$ -statistics, *J. Multivariate Anal.* **51** (1994) 392-413.
- [5] H. DEHLING, O.SH. SHARIPOV, M. WENDLER, Bootstrap for dependent Hilbert space-valued random variables with application to von Mises statistics, *J. Multivariate Anal.* **133** (2015) 200-215.
- [6] H. DEHLING, M. WENDLER, Central limit theorem and the bootstrap for  $U$ -statistics of strongly mixing data, *J. Multivariate Anal.* **101** (2010) 126-137.
- [7] H. DEHLING, M. WENDLER, Law of the iterated logarithm for  $U$ -statistics of weakly dependent observations, in: Berkes, Bradley, Dehling, Peligrad, Tichy (Eds): *Dependence in Probability, Analysis and Number Theory*, Kendrick Press, Heber City (2010).
- [8] B. EFRON, Bootstrap methods: another look at the jackknife, *Ann. Stat.* **7** (1979) 1-26.
- [9] S. GONÇALVES, H. WHITE, The bootstrap of the mean for dependent heterogeneous arrays, *Econometric Theory* **18** (2002) 1367-1384.
- [10] P. HALL, *The bootstrap and Edgeworth expansions*, Springer, New York (1992).

- [11] W. HOEFFDING, A class of statistics with asymptotically normal distribution, *Ann. Math. Stat.* **19** (1948) 293-325.
- [12] S.N. LAHIRI, *Resampling methods for dependent data*, Springer, New York (2003).
- [13] A. LEUCHT, M.H. NEUMANN, Consistency of general bootstrap methods for degenerate U- and V-type statistics, *J. Mult. Anal.* **100** (2009) 1622-1633.
- [14] A. LEUCHT, M.H. NEUMANN, Dependent wild bootstrap for degenerate U- and V-statistics. *J. Mult. Anal.*, **117** (2013) 257-280.
- [15] D.N. POLITIS, J.P. ROMANO, A circular block-resampling procedure for stationary data, *In: Exploring the Limits of the Bootstrap* (R. Lepage and L. Billard ,eds.) (1992) 263-270. Wiley, New York.
- [16] D.N. POLITIS, J.P. ROMANO, Large sample confidence regions based on subsamples under minimal assumptions, *Ann. Stat.* **22** (1994) 2031-2050.
- [17] D.N. POLITIS, H. WHITE, Automatic block-length selection for the dependent bootstrap, *Econometric Reviews* **23** (2004) 53-70.
- [18] M. PELIGRAD, On the blockwise bootstrap for empirical processes for stationary sequences, *Ann. Prob.* **2** (1998) 877-901.
- [19] Q. SHAO, H. YU, Bootstrapping the sample means for stationary mixing sequences, *Stochastic Process. Appl.* **48** (1993) 175-190.
- [20] O.SH. SHARIPOV, M. WENDLER, Bootstrap for the sample mean and for  $U$ -statistics of mixing and near-epoch dependent processes, *Journal of Nonparametric Statistics* **24** (2012) 317-342.
- [21] K. SINGH, On the asymptotic accuracy of Efron's bootstrap, *Ann. Stat.* **9** (1981) 1187-1195.
- [22] R. YOKOYAMA, Moment bounds for stationary mixing sequences, *Z. Wahrsch. verw. Gebiete* **52** (1980) 45-57.
- [23] K. YOSHIHARA, Limiting behavior of  $U$ -statistics for stationary, absolutely regular processes, *Z. Wahrsch. verw. Gebiete* **35** (1976) 237-252.





