The Annals of Statistics
1994, Vol. 22, No. 2, 995–1012

BLOCKWISE BOOTSTRAPPED EMPIRICAL PROCESS FOR STATIONARY SEQUENCES

BY PETER BÜHLMANN

ETH Zürich

We apply the bootstrap for general stationary observations, proposed by Künsch, to the empirical process for p-dimensional random vectors. It is known that the empirical process in the multivariate case converges weakly to a certain Gaussian process. We show that the bootstrapped empirical process converges weakly to the same Gaussian process almost surely, assuming that the block length l for constructing bootstrap replicates satisfies $l(n) = O(n^{1/2 - \varepsilon})$, $0 < \varepsilon < \frac{1}{2}$, and $l(n) \to \infty$.

An example where the multivariate setup arises are the robust GM-estimates in an autoregressive model. We prove the asymptotic validity of the bootstrap approximation by showing that the functional associated with the GM-estimates is Fréchet-differentiable.

1. Introduction. The bootstrap proposed by Efron (1979) has become a well-established nonparametric method for studying the sampling distribution of a given statistic. A large class of statistics can be written as statistical functionals of the empirical distribution function. This motivates one to study the bootstrap for the empirical process. In the i.i.d. setup Bickel and Freedman (1981) have shown the almost-sure weak convergence of the bootstrapped empirical process to the Brownian bridge, that is, the empirical process and its bootstrapped process have the same limiting distribution. This implies that the bootstrap works for smooth statistical functionals.

An extension of the bootstrap for general stationary sequences of observations has been given by Künsch (1989). Instead of selecting single observations X_i from the sample $\{X_1, \dots, X_n\}$ with replacement, this extended method selects k blocks of consecutive observations of length l. Here l is a function of n, tending to infinity, with $l(n) = o(n)$ and $n = kl$. Under some assumptions for the mixing coefficients of the process $\{X_i\}_{i \in \mathbb{Z}}$, Künsch shows that the law of the bootstrapped mean is an asymptotically valid approximation for the limiting law of the arithmetic mean.

In order to establish the validity of the bootstrap for the large class of statistics given by smooth functionals, one would like to show that the bootstrapped empirical process converges weakly to the right Gaussian process. This is done in this paper under a condition on the decay of the strong mixing coefficients and on the block length $l(n)$. Recently Naik-Nimbalkar and Rajarshi (1994) have given a similar result for the one-dimensional case $p = 1$. Our results are

Received May 1992; revised April 1993.

AMS 1991 *subject classifications.* Primary 62G09; secondary 62G20, 62M10.

Key words and phrases. Bootstrap, empirical process, Fréchet-differentiability, GM-estimates, resampling, stationary and strong-mixing sequences, weak convergence.

more general in that we have fewer restrictions on the block length and we consider also the case of vector-valued observations $\mathbf{X}_i \in \mathbb{R}^p$.

For almost-sure weak convergence, Naik-Nimbalkar and Rajarshi (1994) make the quite restrictive assumption that $l(n) = O(n^{1/2-\varepsilon})$, $\frac{1}{4} < \varepsilon < \frac{1}{2}$. From considerations of mean square error of the bootstrap variance, one would like to have an l of larger order than $n^{1/4}$. For $l(n) = O(n^{1/2-\varepsilon})$, $0 < \varepsilon < \frac{1}{4}$, they only get convergence in probability; this means that one has not very good bounds for the error in the bootstrap approximation, [cf. Gill (1989), Theorems 4 and 5].

The extension to the multivariate case is of considerable interest, because with time series data one often uses statistics which depend on some finite-dimensional marginal of the process. In order to bootstrap such a statistic, Künsch (1989) has proposed to consider a block of consecutive original observations as a new vector-valued observation and to apply the blockwise resampling technique to these vector-valued observations. In this way vector-valued observations arise naturally.

Künsch (1989) suggested various applications of his procedure. As an example we prove that, under some regularity conditions, the bootstrap works for the robust GM-estimates for the parameters of a stationary autoregressive model of order p. For this we show that under suitable conditions GM-estimates can be written as differentiable statistical functionals; this is sufficient for a valid bootstrap approximation. The finite-sample properties of the procedure are investigated by a simulation study.

2. Preliminaries and result. In the following we are working in the framework of strong-mixing (or, equivalently, α-mixing) sequences. The strong-mixing concept describes some kind of short-range dependence [cf. Ibragimov and Linnik (1971)]. Let $\{\mathbf{X}_i = (X_{i1}, \dots, X_{ip})'\}_{i \in \mathbb{Z}}$ be a stationary α-mixing sequence of stochastic vectors. Having observed realizations $\mathbf{X}_1, \dots, \mathbf{X}_n$, one is often interested in the empirical process based on these observations. However, instead of $\{\mathbf{X}_i\}_{i=1}^n$ we consider some transformed variables $\{\mathbf{Y}_i\}_{i=1}^n$ given below.

Denote the marginal distribution function of X_{ij} by $F^{(j)}$ and assume that $F^{(j)}$ is continuous. Let $Y_{ij} = F^{(j)}(X_{ij})$, $j = 1, \dots, p$, and let $Y_i = (Y_{i1}, \dots, Y_{ip})'$, $i \in \mathbb{Z}$. Denote the distribution function of $\mathbf{Y}_i$, by G, that is, $G(\mathbf{t}) = P[\mathbf{Y}_i \leq \mathbf{t}]$, $\mathbf{t} \in E^p$ where $E^p = \{\mathbf{t}; \mathbf{0} \leq \mathbf{t} \leq \mathbf{1}\}$ is the p-dimensional unit cube. Note that $G^{(j)}(t) = t$ and $G(\mathbf{t}) = 0$ if at least one coordinate of $\mathbf{t}$ is 0. By the continuous mapping theorem it suffices to study the empirical process of the random vectors $\mathbf{Y}$ (see the remark at the end of this section).

We set $\mathbf{0} = (0, \dots, 0)'$ and $\mathbf{1} = (1, \dots, 1)'$ and let $z^{(j)}$ denote the jth component of a vector $\mathbf{z}$. Moreover, $\mathbf{a} \leq \mathbf{b}$ means that $a^{(j)} \leq b^{(j)}$, $j = 1, \dots, p$, and we write $|\mathbf{t} - \mathbf{s}| = \sup\{|t^{(j)} - s^{(j)}|; j = 1, \dots, p\}$.

The empirical distribution function of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ is defined by

$$G_n(\mathbf{t}) = n^{-1} \sum_{i=1}^{n} 1_{[\mathbf{Y}_i \leq \mathbf{t}]}, \qquad \mathbf{t} \in E^p.$$

The empirical process $\{W_n(\mathbf{t})\}_{\mathbf{t} \in E^p}$ is then defined by

$$W_n(\mathbf{t}) = n^{1/2}\big[G_n(\mathbf{t}) - G(\mathbf{t})\big].$$

W_n belongs to the cadlag space $D^p[0,1]$. We study the weak convergence of empirical processes as random elements in the space $D^p[0,1]$ with respect to the (extended) Skorohod J_1-topology and denote it by $\Rightarrow$ [see Bickel and Wichura (1971) and Billingsley (1968)]. It is known that if

$$\alpha(n) = O\big(n^{-\max\{5/2+\delta, 3p/2+\delta\}}\big), \qquad \delta > 0,$$

then $W_n \Rightarrow W$, where W is a Gaussian process with

$$E\big[W(\mathbf{t})\big] = 0,$$

$$E\big[W(\mathbf{s})W(\mathbf{t})\big] = \sum_{k=-\infty}^{\infty} E\Big[\big(1_{[\mathbf{Y}_0 \le \mathbf{s}]} - G(\mathbf{s})\big)\big(1_{[\mathbf{Y}_k \le \mathbf{t}]} - G(\mathbf{t})\big)\Big].$$

[See, e.g., Yoshihara (1975).]

The bootstrapped empirical distribution function of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ is defined as

$$G_n^*(\mathbf{t}) = k^{-1}\sum_{i=1}^{k} l^{-1} \sum_{j=S_i+1}^{S_i+l} 1_{[\mathbf{Y}_j \le \mathbf{t}]}, \qquad \mathbf{t} \in E^p,$$

where $n = kl$, $l = l(n) = o(n)$ and $l(n) \to \infty$ $(n \to \infty)$; S_i i.i.d. $\sim$ Uniform($\{0, \dots, n-l\}$) [see Künsch (1989)]. In practice, if the sample size is not a multiple of l, we simply make the last block shorter. The bootstrapped empirical process is then defined as

$$W_n^*(\mathbf{t}) = n^{1/2}\Big[G_n^*(\mathbf{t}) - E^*\big[G_n^*(\mathbf{t})\big]\Big], \qquad \mathbf{t} \in E^p.$$

E^*, Var* and so on denote the moments under the conditional probability measure P^*, induced by the resampling mechanism. In the following we abbreviate $E^*[G_n^*(\mathbf{t})]$ by $\mu_n^*(\mathbf{t})$. We observe that $G_n^*(\mathbf{t})$ is the mean of k variables that are i.i.d. under P^*.

THEOREM 1. *Let $\{\mathbf{X}_i\}_{i \in \mathbb{Z}}$ be a stationary, strong-mixing sequence with $\sum_{i=0}^{\infty}(i+1)^{8p+7}\alpha^{1/2}(i) < \infty$. Assume that $\mathbf{X}_i$ has continuous marginal distributions and that $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$.*

Then $W_n^* \Rightarrow W$ almost surely in the Skorohod J_1-topology.

COROLLARY 1. *Under the conditions of Theorem 1 we have*

$$\sqrt{n}(G_n^* - G_n) \Rightarrow W \quad \textit{almost surely in the Skorohod } J_1\textit{-topology.}$$

PROOF. Since $\sqrt{n}(G_n - \mu_n^*) = O(ln^{-1/2})$ [see Künsch (1989), (3.14)] and using Slutsky's theorem the proof is obvious. □

REMARK. Theorem 1 and Corollary 1 remain true if the processes W_n^* and W are defined in terms of the untransformed $\mathbf{X}_i$'s with distribution F (instead of G), the processes being then elements of $D^p(\mathbb{R})$. The argument for this assertion is given in Billingsley [(1968), proof of Theorem 22.1]; the extension to the multidimensional case is straightforward.

Sections 3 and 4 deal with the proof of Theorem 1.

3. Convergence of the finite-dimensional distributions. We let $A_r(\alpha) = \sum_{i=0}^{\infty}(i+1)^{r/2-1}\alpha^{1/2}(i)$ and let $\|\cdot\|_q = E_P^{1/q}|\cdot|^q$ denote the $\mathcal{L}_q$-norm with respect to P.

LEMMA 1. *Let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary strong-mixing real valued sequence with mixing coefficients α. Let $\mathcal{F}_a^b$ denote the σ-algebra $\sigma(\{X_i; a \le i \le b\})$.*

(a) *If $\zeta \in \mathcal{F}_{-\infty}^0$, $\eta \in \mathcal{F}_n^{\infty}$, then*

$$|E[\zeta\eta] - E[\zeta]E[\eta]| \le 12\|\zeta\|_{q_1}\|\eta\|_{q_2}\alpha^{1/q_3}(n),$$

where $0 \le q_1, q_2, q_3 \le \infty$, $q_1^{-1} + q_2^{-1} + q_3^{-1} = 1$.

(b) *If $E[X_1] = 0$, $E|X_1|^{2m+\nu} < \infty$ and $\sum_{i=0}^{\infty}(i+1)^{m-1}\alpha(i)^{\nu/(2m+\nu)} < \infty$ for some $\nu > 0$, then*

$$E\left|\sum_{i=1}^{n} X_i\right|^{2m} \le K_\alpha \|X_1\|_{2m+\nu}^{2m} n^m, \quad \textit{for } K_\alpha \textit{ a constant depending on } \alpha.$$

(c) *If $E[X_1] = 0$ and $E|X_1|^8 < \infty$, then*

$$E\left|\sum_{i=1}^{n} X_i\right|^4 \le \text{const.}n^2\left\{\|X_1\|_8^4 A_4(\alpha) + \|X_1\|_4^4 \left(A_2(\alpha)\right)^2\right\}.$$

PROOF. (a) The proof is given in Deo (1973). Part (b) is Theorem 1 in Yokoyama (1980). Part (c) is seen from formula (4.4) in Yokoyama (1980). □

LEMMA 2. *If $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$, and $A_{16}(\alpha) < \infty$, then*

$$\mathrm{Cov}^*\left(W_n^*(\mathbf{s}), W_n^*(\mathbf{t})\right) = \mathrm{Cov}\left(W(\mathbf{s}), W(\mathbf{t})\right) + \Delta_{\mathbf{s},\mathbf{t}}(n),$$

where $\Delta_{\mathbf{s},\mathbf{t}}(n) = o(1)$ almost surely, that is, there exists a set $A(\mathbf{s},\mathbf{t})$ with $P[A(\mathbf{s},\mathbf{t})] = 1$ such that on the set $A(\mathbf{s},\mathbf{t})$ for all $\kappa > 0$ there exists an $n_0(\kappa,\omega)$ with $|\Delta_{\mathbf{s},\mathbf{t}}(n)| < \kappa$ for $n \ge n_0$, where ω is an element of the probability space with measure P.

PROOF. Let $G(\mathbf{s},\mathbf{t}) = G(\mathbf{t}) - G(\mathbf{s})$, $I_j(\mathbf{s},\mathbf{t}) = 1_{[\mathbf{Y}_j \le \mathbf{t}]} - 1_{[\mathbf{Y}_j \le \mathbf{s}]}$ and $\mu_n^*(\mathbf{s},\mathbf{t}) = \mu_n^*(\mathbf{t}) - \mu_n^*(\mathbf{s})$.

We consider

$$\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) = l(n-l+1)^{-1}\sum_{i=0}^{n-l}\left[l^{-1}\sum_{j=i+1}^{i+l}\left(I_j(\mathbf{s},\mathbf{t}) - \mu_n^*(\mathbf{s},\mathbf{t})\right)\right]^2.$$

Then

$$\begin{aligned}&\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) - E\left[\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right)\right]\\ &\quad = l(n-l+1)^{-1}\sum_{i=0}^{n-l} Z_i(\mathbf{s},\mathbf{t}) - lV(\mathbf{s},\mathbf{t}),\end{aligned} \tag{1}$$

where

$$Z_i(\mathbf{s},\mathbf{t}) = \left[l^{-1}\sum_{j=i+1}^{i+l}\left(I_j(\mathbf{s},\mathbf{t}) - G(\mathbf{s},\mathbf{t})\right)\right]^2 - E\left|l^{-1}\sum_{j=i+1}^{i+l}\left(I_j(\mathbf{s},\mathbf{t}) - G(\mathbf{s},\mathbf{t})\right)\right|^2,$$

$$V(\mathbf{s},\mathbf{t}) = \left[\mu_n^*(\mathbf{s},\mathbf{t}) - G(\mathbf{s},\mathbf{t})\right]^2 - E|\mu_n^*(\mathbf{s},\mathbf{t}) - G(\mathbf{s},\mathbf{t})|^2.$$

In the following we write Z_i and V:

$$\begin{aligned}&E\left|\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) - E\left[\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right)\right]\right|^4\\ &\quad \le \left[l(n-l+1)^{-1}E^{1/4}\left|\sum_{i=0}^{n-l} Z_i\right|^4 + lE^{1/4}|V|^4\right]^4.\end{aligned} \tag{2}$$

Note that $\{Z_i\}_{i\in\mathbb{Z}}$ are again stationary, strongly mixing with mixing coefficients

$$\alpha_Z(i) \le \begin{cases}\alpha_X(i-l+1), & i \ge l,\\ \alpha_X(0) \le 1, & i < l.\end{cases}$$

Using Lemma 1(c) and noting that $\|Z_1\|_p \le \|Z_1\|_q$ for $p \le q$, we get

$$E\left|\sum_{i=0}^{n-l} Z_i\right|^4 \le \text{const.}(n-l+1)^2 O(l^2)\|Z_1\|_8^4. \tag{3}$$

Using Lemma 1(b) we get $E|Z_1|^8 \le \text{const.}\, l^{-8}$. Therefore

$$E\left|\sum_{i=0}^{n-l} Z_i\right|^4 \le \text{const.}\, l^{-2}(n-l+1)^2. \tag{4}$$

On the other hand, we have $E|V|^4 = E|U - E[U]|^4$, where $U = [\mu_n^*(\mathbf{s},\mathbf{t}) - G(\mathbf{s},\mathbf{t})]^2$. Since $U \ge 0$ we have $E|V|^4 \le E|U|^4 + 6E|U|^2E^2[U]$. By the Minkowski inequality and using Lemma 1(b) we get $E|U|^r = O(n^{-r})$. Therefore

$$E|V|^4 = O(n^{-4}). \tag{5}$$

Putting (5) and (4) in (2) we conclude that

$$E\left|\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) - E\left[\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right)\right]\right|^4 = O\left(l^2(n-l+1)^{-2}\right). \tag{6}$$

So, by (6),

$$\sum_{n=1}^{\infty} E\left|\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) - E\left[\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right)\right]\right|^4 < \infty,$$

since $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$. Therefore

$$\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) = E\left[\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right)\right] + \xi_{\mathbf{s},\mathbf{t}}(n),$$

where $\xi_{\mathbf{s},\mathbf{t}}(n) = o(1)$ on a set $C(\mathbf{s},\mathbf{t})$ with $P[C(\mathbf{s},\mathbf{t})] = 1$.

From Lemma 1(a), it follows that

$$E\left[\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right)\right] = \operatorname{Var}\left(W(\mathbf{t}) - W(\mathbf{s})\right) + o(1) \quad \text{uniformly in } \mathbf{s}, \mathbf{t}$$

[see also Künsch (1989), Theorems 3.2 and 3.4]. This yields

$$\operatorname{Var}^*\left(W_n^*(\mathbf{t}) - W_n^*(\mathbf{s})\right) = \operatorname{Var}\left(W(\mathbf{t}) - W(\mathbf{s})\right) + \xi_{\mathbf{s},\mathbf{t}}(n) + o(1). \tag{7}$$

Finally, (7) yields

$$\operatorname{Cov}^*\left(W_n^*(\mathbf{t}), W_n^*(\mathbf{s})\right) = \operatorname{Cov}\left(W(\mathbf{t}), W(\mathbf{s})\right) + \Delta_{\mathbf{s},\mathbf{t}}(n),$$

where $\Delta_{\mathbf{s},\mathbf{t}}(n) = 1/2\left(\xi_{\mathbf{0},\mathbf{t}}(n) + \xi_{\mathbf{s},\mathbf{0}}(n) - \xi_{\mathbf{s},\mathbf{t}}(n)\right) + o(1)$. □

LEMMA 3. *If $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$, and $A_{16(p+1)}(\alpha) < \infty$, then*

$$\left(W_n^*(\mathbf{t}_1), \dots, W_n^*(\mathbf{t}_d)\right)' \to_{d^*} \left(W(\mathbf{t}_1), \dots, W(\mathbf{t}_d)\right)' \quad \textit{almost surely},$$

for all $(\mathbf{t}_1, \dots, \mathbf{t}_d)' \in (E^p)^d$, for all $d \in \mathbb{N}$, that is, there exists a set C, with $P[C] = 1$ such that on C the d^*-convergence holds for all $(\mathbf{t}_1, \dots, \mathbf{t}_d)' \in (E^p)^d$, for all $d \in \mathbb{N}$.

PROOF. We split the proof into two parts.

(i) We assume first that $(\mathbf{t}_1, \dots, \mathbf{t}_d) \in (\mathbb{Q}^p)^d \cap (E^p)^d$, where $\mathbb{Q}$ denotes the set of rational numbers.

Let $A := \cap_{\mathbf{t}_i, \mathbf{t}_j \in \mathbb{Q}^p} A(\mathbf{t}_i, \mathbf{t}_j)$, where $A(\mathbf{t}_i, \mathbf{t}_j)$ is defined via Lemma 2. Then A^C is still a nullset, that is, $P[A] = 1$. Let $Z_n^* = \sum_{i=1}^d c_i W_n^*(\mathbf{t}_i)$, $c_i \in \mathbb{R}$. From Lemma 2 we know that, on A, $\operatorname{Var}^*(Z_n^*) \to \operatorname{Var}(Z)$, where $Z = \sum_{i=1}^d c_i W(\mathbf{t}_i)$. So condition (B1) in Künsch (1989) is satisfied. Condition (B3) in Künsch (1989) holds since $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$, and $1_{[\mathbf{Y}_j \le \cdot]}$ is bounded by 1.

Then the Lindeberg condition holds for $Z_n^*/\sqrt{\operatorname{Var}(Z)}$ on the set A. So by the univariate central limit theorem we have $Z_n^* \to_{d^*} Z$ on A [see Künsch (1989)].

(ii) Assume that $(\mathbf{t}_1, \dots, \mathbf{t}_d) \in (E^p)^d$. In Lemma 11 we prove that W_n^* fulfills the tightness condition on a set $\widetilde{A}$ with $P[\widetilde{A}] = 1$, that is,

$$\lim_{\delta \searrow 0} \limsup_n P^*\left[w_\delta(W_n^*) > \kappa\right] = 0 \quad \forall \kappa > 0, \text{ on } \widetilde{A}, \tag{8}$$

where $w_\delta(x) = \sup\{|x(\mathbf{t}) - x(\mathbf{s})|;\ \mathbf{s}, \mathbf{t} \in E^p, |\mathbf{t} - \mathbf{s}| < \delta\}$, $x \in \mathcal{D}^p([0,1])$. Relation (8) implies that, for all $\kappa > 0$, for all $\eta > 0$, there exist $\mathbf{q}_i \in \mathbb{Q}^p \cap E^p$ in a neighborhood of $\mathbf{t}_i$ and n_0 such that

$$P^*\left[|W_n^*(\mathbf{t}_i) - W_n^*(\mathbf{q}_i)| > \kappa\right] < \eta \quad \text{for } n \geq n_0, \text{ on the set } \widetilde{A}. \tag{9}$$

Define $\mathbf{Y}_n^* = \sum_{i=1}^d c_i W_n^*(\mathbf{q}_i)$. In order to show $Z_n^* \to_{d^*} Z$ a.s., we proceed similarly to proving Slutsky's theorem. Let $a \in \mathbb{R}$ be arbitrary. Then, for any $\rho > 0$,

$$P^*\left[Z_n^* \leq a\right] \leq P^*\left[Y_n^* \leq a + \rho\right] + P^*\left[|Z_n^* - Y_n^*| > \rho\right].$$

Let $\sigma^2(\{c_i\}, \{\mathbf{s}_i\}) = \mathrm{Var}(\sum_{i=1}^d c_i W(\mathbf{s}_i))$. Then $\sigma^2(\{c_i\}, \{\mathbf{s}_i\})$ is continuous in both arguments since $W \in \mathcal{C}^p[0,1]$. By part (i) we get the following: for every $\gamma > 0$ there exists n_1 such that, for $n \geq n_1$ on the set A,

$$P^*\left[Y_n^* \leq a + \rho\right] \leq \Phi\left(\frac{a + \rho}{\sigma(\{c_i\}, \{\mathbf{q}_i\})}\right) + \gamma \leq \Phi\left(\frac{a + \rho}{\sigma(\{c_i\}, \{\mathbf{t}_i\})}\right) + 2\gamma \tag{10}$$

($\mathbf{q}_i$ can be chosen arbitrarily close to $\mathbf{t}_i$).

On the other hand, (9) implies

$$P^*\left[|Z_n^* - Y_n^*| > \rho\right] \leq \gamma \quad \text{for } n \text{ sufficiently large, on the set } \widetilde{A}. \tag{11}$$

Relations (10) and (11) yield

$$P^*[Z_n^* \leq a] \leq \Phi\left(\frac{a + \rho}{\sigma(\{c_i\}, \{\mathbf{t}_i\})}\right) + 3\gamma \quad \text{for } n \text{ sufficiently large, on the set } A \cap \widetilde{A}.$$

Analogously, we arrive at

$$P^*[Z_n^* \leq a] \geq \Phi\left(\frac{a - \rho}{\sigma(\{c_i\}, \{\mathbf{t}_i\})}\right) - 3\gamma \quad \text{for } n \text{ sufficiently large, on the set } A \cap \widetilde{A}.$$

Since γ and ρ are arbitrary, we have

$$P^*\left[Z_n^* \leq a\right] \to \Phi\left(\frac{a}{\sigma(\{c_i\}, \{\mathbf{t}_i\})}\right) \qquad (n \to \infty) \text{ on the set } A \cap \widetilde{A}.$$

Therefore $Z_n^* \to_{d^*} Z$ on $A \cap \widetilde{A}$. Set $C = A \cap \widetilde{A}$, and note that $P[A \cap \widetilde{A}] = 1$.

Finally, the lemma follows by the Cramér–Wold device. □

4. Tightness of the bootstrapped empirical process. In order to show weak convergence of W_n^* to W we have to prove tightness for the W_n^*-process. In the following we denote by ω an element of the probability space with measure P. It suffices to show that on a set C with $P[C] = 1$, we have the following: $\forall\, \kappa > 0, \forall\, \eta > 0, \exists\, \delta > 0$, and $\exists\, n_0(\omega) \in \mathbb{N}$ such that, $\forall\, n \geq n_0(\omega)$,

$$P^*\left[\omega_\delta(W_n^*) > \kappa\right] < \eta. \tag{12}$$

where $w_\delta(W_n^*) = \sup\{|W_n^*(\mathbf{t}) - W_n^*(\mathbf{s}); \mathbf{s}, \mathbf{t} \in E^p, |\mathbf{t} - \mathbf{s}| < \delta\}$.

Let $B = B(\mathbf{b}_0, \delta) = \{\mathbf{t};\ \mathbf{b}_0 \leq \mathbf{t} \leq \mathbf{b}_0 + \delta\mathbf{1}\}$. Instead of proving (12), it suffices to show that for all $\mathbf{b}_0$, on a set $C = C(\mathbf{b}_0)$ with $P[C] = 1$, we have the following: $\forall\, \kappa > 0, \forall\, \eta > 0, \exists\, \delta, 0 < \delta < 1$, and $\exists\, n_0(\omega) \in \mathbb{N}$ such that, $\forall\, n \geq n_0(\omega)$,

$$P^*\left[\sup_{\mathbf{t}\in B} |W_n^*(\mathbf{t}) - W_n^*(\mathbf{b}_0)| > \kappa\right] < \left(G(B) + \delta^p\right)\eta. \tag{13}$$

We identify the distribution function G with the corresponding probability measure. Without loss of generality we assume $\delta \in \mathbb{Q}, \mathbf{b}_0 \in \mathbb{Q}^p \cap E^p$ [cf. Billingsley (1968), proof of Theorem 8.3]. For condition (13) see Sen [(1974), formula (2.5)]; note that only a countable intersection of sets $C(\mathbf{b}_0)$ is involved.

In the following we fix the hyperquadrangle B. We consider the following grid (for a fixed n):

$$\mathbf{b}_n(\mathbf{i}) = \mathbf{b}_0 + \frac{\kappa}{6p} n^{-1/2}\mathbf{i}, \qquad \mathbf{0} \leq \mathbf{i} \leq \mathbf{m}_n, \qquad \kappa > 0,$$

where $\mathbf{m}_n = m_n\mathbf{1}$, $m_n = [6p\delta n^{1/2}/\kappa] - 1$ and, for $\mathbf{i}$ with $i^{(j)} = m_n + 1$, we define $\mathbf{b}_n(\mathbf{i})^{(j)} = \mathbf{b}_0^{(j)} + \delta$, where $\mathbf{z}^{(j)}$ denotes the jth component of a $p \times 1$ vector $\mathbf{z}$.

LEMMA 4.

$$\sup_{\mathbf{t}\in B} |W_n^*(\mathbf{t}) - W_n^*(\mathbf{b}_0)| \leq 3 \max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n} \left|W_n^*\left(\mathbf{b}_n(\mathbf{i}+\mathbf{1})\right) - W_n^*(\mathbf{b}_0)\right|$$
$$+ n^{1/2} \max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n} \left|\mu_n^*\left(\mathbf{b}_n(\mathbf{i}+\mathbf{1})\right) - \mu_n^*\left(\mathbf{b}_n(\mathbf{i})\right)\right|.$$

PROOF. For $\mathbf{t} \in \{\mathbf{s}; \mathbf{b}_n(\mathbf{i}) \leq \mathbf{s} \leq \mathbf{b}_n(\mathbf{i}+\mathbf{1})\}$ a few routine steps lead to

$$\left|W_n^*(\mathbf{t}) - W_n^*\left(\mathbf{b}_n(\mathbf{i})\right)\right| \leq \left|W_n^*\left(\mathbf{b}_n(\mathbf{i}+\mathbf{1})\right) - W_n^*\left(\mathbf{b}_n(\mathbf{i})\right)\right|$$
$$+ n^{1/2}\left|\mu_n^*\left(\mathbf{b}_n(\mathbf{i}+\mathbf{1})\right) - \mu_n^*\left(\mathbf{b}_n(\mathbf{i})\right)\right|$$

[see Billingsley (1968), (22.17)].

This implies the desired result. □

In the next step we will analyze $\max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n} |W_n^*(\mathbf{b}_n(\mathbf{i}+\mathbf{1})) - W_n^*(\mathbf{b}_0)|$ and prove that, for n sufficiently large,

$$P^*\left[\max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n} \left|W_n^*\left(\mathbf{b}_n(\mathbf{i}+\mathbf{1})\right) - W_n^*(\mathbf{b}_0)\right| > \frac{\kappa}{6}\right] < \left(G(B) + \delta^P\right)\frac{\eta}{2} \quad \text{a.s.} \tag{14}$$

[cf. (13)].

Let $H_i(n), i = 1, \dots, N(n)$, denote a p-dimensional hyperrectangle whose corner points are points from the grid of B. In the following we identify functions with hyperrectangles H as arguments in the obvious way via the basic quantity $1_{[\mathbf{Y} \in H]}$. In the spirit of Bickel and Wichura (1971) we first show $E^*|W_n^*(H_i(n))|^{2r+2} \leq [\text{const.}\mu(H_i(n))]^{\beta}, \beta > 1$, almost surely, for an appropriate $r \in \mathbb{N}, \mu$ denoting a finite, nonnegative measure in E^p (see Lemma 7). Let $D_j(H_i(n)) = l^{-1}\sum_{t=S_j+1}^{S_j+l} 1_{[\mathbf{Y}_t \in H_i(n)]}, j = 1, \dots, k$. Then $W_n^*(H_i(n)) = n^{1/2}k^{-1} \times \sum_{j=1}^{k}(D_j(H_i(n)) - E^*[D_j(H_i(n))])$. Note that $D_1, \dots, D_k$ are i.i.d. under P^*. The key idea is to study the $(2r+2)$th moment of $W_n^*(H_i(n))$ in terms of the second centered moment of $D_1(H_i(n))$. The main difficulty is to derive a bound for $E^*|D_1(H_i(n)) - E^*[D_1(H_i n))]|^2$ which holds uniformly in the index $1 \leq i \leq N(n) = O(n^p)$, where $N(n)$ denotes the number of hyperrectangles $H_i(n)$.

LEMMA 5. *Assume that* $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$, *and* $A_{16(p+1)}(\alpha) < \infty$. *Then*

$$nk^{-1}E^*\Big|D_1\big(H_i(n)\big) - E^*\Big[D_1\big(H_i(n)\big)\Big]\Big|^2 \leq G\big(H_i(n)\big)^{1/20(p+1)}\big(\text{const.} + \Delta_i(n)\big),$$

where $\sup_{i \leq N(n)} |\Delta_i(n)| = o(1)$ *almost surely.*

PROOF. We split the left-hand side into a centered sum and an expectation, that is, $nk^{-1}E^*|D_1(H_i(n)) - E^*[D_1(H_i(n))]|^2 = R_n + E[R_n]$, where $R_n = l(n-l+1)^{-1} \times \sum_{j=0}^{n-l} Z_j(H_i(n)) - lV(H_i(n))$, where $Z_j(H_i(n))$ and $V(H_i(n))$ are defined as in the proof of Lemma 2, replacing $I(\mathbf{s}, \mathbf{t})$ by $1_{[\mathbf{Y} \in H_i(n)]}$ and so on. The main difficulty is to derive the following key inequality:

$$E\left|G\big(H_i(n)\big)^{-1/(10r)}\sum_{j=0}^{n-l} Z_j\big(H_i(n)\big)\right|^{2r} = O(n^r l^{-r}), \qquad r \in \mathbb{N}. \tag{15}$$

Then by choosing r large enough we can get a uniform estimate by a Borel–Cantelli argument. To prove (15), we proceed similarly as in Yokoyama (1980) using the bound $\|Z_1(H_i(n))\|_m \leq \text{const.}l^{-1}G(H_i(n))^{1/(2m)}$. Details are given in Bühlmann (1992). □

LEMMA 6. *Assume that* $\{Z_i\}_{i \in \mathbb{Z}}$ *are i.i.d. random variables with* $E[Z_1] = 0$ *and* $|Z_i| \leq 1$. *Let* $S_n = \sum_{i=1}^{n} Z_i$. *Then*

$$E|S_n|^{2r} \leq K_r\Big\{nE|Z_1|^2 + \cdots + n^r\big(E|Z_1|^2\big)^r\Big\}, \qquad r \in \mathbb{N},\ K_r \textit{ a constant.}$$

PROOF. Since $E|Z_i|^k \leq E|Z_1|^2$ for $k \geq 2$ and using the independence assumption, this lemma is obvious. □

LEMMA 7. *Assume that* $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$, *and* $A_{16(p+1)} < \infty$. *Let* $\mu(C) = G(C) + \lambda(C)$, *where* λ *denotes the Lebesque measure in* E^p, *and* C *a Borel set in*

E^p. Then there exists an $r = r(p) \in \mathbb{N}$ such that

$$E^*\left|W_n^*(H_i(n))\right|^{2r+2} \leq \left[\left(\text{const.} + \zeta_i(n)\right)\mu(H_i(n))\right]^{\beta}, \qquad \beta > 1$$

where $\sup_{i \leq N(n)} |\zeta_i(n)| \leq K$ *almost surely* (K *a constant*).

PROOF. Let $\tau = nk^{-1}E^*|D_1(H_i(n)) - E^*[D_1(H_i(n))]|^2$. Lemma 6 yields

$$\begin{aligned} E^*&\left|W_n^*(H_i(n))\right|^{2r+2} \\ &= k^{-r-1}E^*\left|\sum_{j=1}^{k} n^{1/2}k^{-1/2}\left\{D_j(H_i(n)) - E^*\left[D_j(H_i(n))\right]\right\}\right|^{2r+2} \\ &\leq K_r k^{-r-1}\sum_{j=1}^{r+1}(k\tau)^j = K_r\sum_{j=1}^{r+1}\left(ln^{-1}\right)^{r+1-j}\tau^j. \end{aligned}$$

Since $\lambda(H_i(n)) \geq \text{const.}n^{-p/2}$ we have, for n sufficiently large, $ln^{-1} \leq \lambda(H_i(n))^{1/p}$ for $i \leq N(n)$. Therefore $(ln^{-1})^{p+s} \leq [\lambda(H_i(n))]^{\beta_1}$, $\beta_1 > 1$, for $i \leq N(n), s \geq 1$, n sufficiently large. On the other hand we have, by Lemma 5, $\tau^{20(p+1)+s} \leq [\text{const.}\mu(H_i(n))]^{\beta_2}, \beta_2 > 1$, for all $i \leq N(n)$ a.s., for $s \geq 1$. Take $r = 20(p+1) + p$. This finishes the proof. □

Now we are able to give a bound for

$$\max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n}\left|W_n^*(\mathbf{b}_n(\mathbf{i}+\mathbf{1})) - W_n^*(\mathbf{b}_0)\right|$$

in the sense of (14).

LEMMA 8. *Assume that* $l(n) = O(n^{1/2-\varepsilon}), \varepsilon > 0$, *and* $A_{16(p+1)} < \infty$. *Then* $\forall\, \kappa > 0$, $\forall\, \eta > 0$, $\exists\, \delta \in \mathbb{Q}$, $0 < \delta < 1$, *and* $\exists\, n_0(\omega) \in \mathbb{N}$ *such that*, $\forall\, n \geq n_0(\omega)$,

$$P^*\left[\max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n}\left|W_n^*(\mathbf{b}_n(\mathbf{i}+\mathbf{1})) - W_n^*(\mathbf{b}_0)\right| > \frac{\kappa}{6}\right] < \left(G(B) + \delta^p\right)\frac{\eta}{2} \quad \textit{almost surely.}$$

PROOF. Following Bickel and Wichura (1971), we look at the modulus $M''(W_n^*)$. Unfortunately W_n^* does not vanish at the lower boundary of B, that is, there exists a $\mathbf{t}$ with $\mathbf{t}^{(j)} = \mathbf{b}_0^{(j)}$ for some $1 \leq j \leq p$, such that $W_n^*(\mathbf{t}) \neq 0$. For this reason we cannot directly apply the fluctuation inequality from Bickel and Wichura (1971). However, bounds of M'' give rise to bounds of maxima:

$$\begin{aligned} &\max_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}_n}\left|W_n^*(\mathbf{b}_n(\mathbf{i}+\mathbf{1})) - W_n^*(\mathbf{b}_0)\right| \\ &\quad\leq (2^p - 1)M''(W_n^*) + \sum_{\mathbf{u}(\delta)}\left|W_n^*(\mathbf{u}(\delta)) - W_n^*(\mathbf{b}_0)\right|, \end{aligned} \tag{16}$$

where $\mathbf{u}(\delta)$ is a corner point of B.

Using Lemma 7 we can apply Theorem 1 in Bickel and Wichura (1971) (this theorem remains true in our case although W_n^* does not vanish at the lower boundary) to bound the modulus $M''(W_n^*)$. The term $\Sigma_{\mathbf{u}(\delta)}|W_n^*(\mathbf{u}(\delta)) - W_n^*(\mathbf{b}_0)|$ does not involve any maxima; therefore it is much easier to find an appropriate bound for this term. Details can be found in Bühlmann (1992). □

In the last step we consider $n^{1/2}\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n}|\mu_n^*(\mathbf{b}_n(\mathbf{i}+\mathbf{1})) - \mu_n^*(\mathbf{b}_n(\mathbf{i}))|$ (cf. Lemma 4). Let $\mu_n^*(\mathbf{t},\mathbf{s}) = \mu_n^*(\mathbf{t}) - \mu_n^*(\mathbf{s})$, and so on. We have

$$
\begin{aligned}
&\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|\mu_n^*\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big|\\
&\quad\le \max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|\mu_n^*\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big|\\
&\qquad+ \max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big|\\
&\qquad+ \max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big|\\
&\quad\le O(ln^{-1/2}) + \max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_m} n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),b_n(\mathbf{i})\big)\big|\\
&\qquad+ \kappa/3.
\end{aligned}
\tag{17}
$$

Note that we have used

$$\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|\mu_n^*\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| = O(ln^{-1/2})$$

[see Künsch (1989), (3.14)]. In order to analyze $\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n}|G_n(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})) - G(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i}))|$, we establish an extension of Lemma 6 to the α-mixing case.

LEMMA 9. *Assume that $\{Z_i\}_{i\in\mathbb{Z}}$ is a stationary α-mixing sequence with $E[Z_1] = 0$, $|Z_i|\le 1$, and $A_{4r-2}(\alpha)<\infty$. Let $S_n = \sum_{i=1}^n Z_i$, and let $\tau = E|Z_1|^2$. Then*

$$E|S_n|^{2r} \le K_{\alpha,r}\Big[n\tau^{1/2} + \cdots + \big(n\tau^{1/2}\big)^r\Big], \qquad r\in\mathbb{N}.$$

PROOF. We use Lemma 1(a) with $q_1 = 2$, $q_2 = \infty$, $q_3 = 2$ applied to the sequence $\{Z_i\}_{i\in\mathbb{Z}}$. Then we can proceed in exactly the same manner as in the proof of Lemma 3.1 in Sen (1974). □

LEMMA 10. *Assume that $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon>0$, and $A_{8p+18}<\infty$. Then*

$$\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|\mu_n^*\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1})\big) - \mu_n^*\big(\mathbf{b}_n(\mathbf{i})\big)\big| < \kappa/3 + o(1),$$

where the $o(1)$ term is almost surely.

PROOF. We will show that

$$\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| = o(1) \quad \text{a.s.} \tag{18}$$

We have

$$n^{1/2}\Big(G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\Big)$$
$$= n^{-1/2}\sum_{j=1}^{n}\Big(\mathbf{I}_j\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(i)\big)\Big),$$

$\mathbf{I}_j$ as in the proof of Lemma 2.

Let $Z_j(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})) = I_j(\mathbf{b}_n(\mathbf{i}+\mathbf{1})),\mathbf{b}_n(\mathbf{i})) - G(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i}))$. Then

$$\tau := E\big|Z_j\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big|^2 \le G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)$$
$$\le \frac{\kappa}{3n^{1/2}} = \text{const.}n^{-1/2}.$$

Applying Lemma 9 to the random variables $\{Z_j(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i}))\}_{j=1}^{n}$ we get

$$P\Big[n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| > \rho\Big]$$
$$\le \text{const.}\rho^{-2r}\big[n^{-r+1}n^{-1/4} + \cdots + n^{-r/4}\big], \qquad \rho > 0$$

We choose $r = 2p + 5$. Then

$$P\Big[n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| > \rho\Big] \le \text{const.}n^{-p/2-5/4},$$

for all $\mathbf{i}$ and $\mathbf{b}_0$, and

$$P\Big[\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| > \rho\Big]$$
$$\le \sum_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} P\Big[n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| > \rho\Big]$$
$$\le \text{const.}m_n^p n^{-p/2-5/4} = O\big(n^{-5/4}\big).$$

Therefore, for any $\rho > 0$,

$$\sum_{n=1}^{\infty} P\Big[\max_{\mathbf{0}\le\mathbf{i}\le\mathbf{m}_n} n^{1/2}\big|G_n\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big) - G\big(\mathbf{b}_n(\mathbf{i}+\mathbf{1}),\mathbf{b}_n(\mathbf{i})\big)\big| > \rho\Big] < \infty.$$

This proves (18). Relation (18) together with (17) finishes the proof of the lemma. □

LEMMA 11. *Assume that* $l(n) = O(n^{1/2-\varepsilon}), \varepsilon > 0$, *and* $A_{16(p+1)} < \infty$. *Then for every* $\mathbf{0} \le \mathbf{b}_0 \le \mathbf{1}, \mathbf{b}_0 \in \mathbb{Q}^p \cap E^p$, *there exists a set* $C = C(\mathbf{b}_0)$ *with* $P[C] = 1$ *such that on* C *we have the following:* $\forall\kappa > 0, \forall\eta > 0, \exists\delta > 0$ *and* $\exists n_0(\omega) \in \mathbb{N}$ *such that, for* $n \ge n_0(\omega)$,

$$P^*\Big[\sup_{\mathbf{t}\in B}\big|W_n^*(\mathbf{t}) - W_n^*(\mathbf{b}_0)\big| > \kappa\Big] \le \big(G(B) + \delta^p\big)\eta \quad \textit{almost surely.}$$

Moreover, the following holds: $\forall\, \kappa > 0$, $\forall\, \eta > 0$, $\exists\, \delta > 0$ *and* $\exists\, n_0(\omega) \in \mathbb{N}$ *such that, for* $n \geq n_0(\omega)$,

$$P^* \left[w_\delta(W_n^*) > \kappa \right] < \eta \quad \textit{almost surely.}$$

PROOF. The first part is a direct consequence of Lemmas 4, 8 and 10. For the second part we set $\widetilde{C} = \cap_{\mathbf{b}_0 \in \mathbb{Q}^p \cap E^p} C(\mathbf{b}_0)$. Noting that $P[\widetilde{C}] = 1$, the statement can be proved like Theorem 8.3 in Billingsley (1968) [see also Sen (1974), (2.5)]. □

By virtue of Lemmas 3 and 11 we have proved Theorem 1.

5. Bootstrapping GM-estimates. We consider a class of robust estimates for the parameters of an autoregressive model of order p[AR(p)]. The $AR(p)$ model is

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t, \tag{19}$$

where the innovations ϵ_t are assumed to be i.i.d. with $E[\epsilon_t] = 0, E[\epsilon_t^2] = \sigma^2 < \infty$. The assumption of a known location $\mu = 0$ is made to simplify the exposition. Furthermore, we assume that the AR(p)-process is stationary.

The Mallows variant of the GM-estimate $\widehat{\phi}_n$ for the parameter $\phi = (\phi_1, \dots, \phi_p)'$ (and $\widehat{\sigma}_n$ for the nuisance parameter σ) [see Bustos (1982) or Martin and Yohai (1986)] is implicitly defined by

$$\sum_{t=p+1}^{n} \psi\big((X_t - \widehat{\phi}_1 X_{t-1} - \cdots - \widehat{\phi}_p X_{t-p})\widehat{\sigma}_n^{-1}\big) w\big(\|A_n \mathbf{X}_{t-1}^p\|\big)\mathbf{X}_{t-1}^p = 0,$$

$$\sum_{t=p+1}^{n} \chi\big((X_t - \widehat{\phi}_1 X_{t-1} - \cdots - \widehat{\phi}_p X_{t-p})\widehat{\sigma}_n^{-1}\big) = 0,$$

where ψ: $\mathbb{R} \to \mathbb{R}$ and χ: $\mathbb{R} \to \mathbb{R}$ are bounded robustifying psi- and chi-functions, w: $\mathbb{R} \to \mathbb{R}^+$ is a positive weight function and A_n is a sequence of nonsingular $p \times p$ matrices; often A_n is a robust estimate of $(\mathrm{Cov}(\mathbf{X}_{t-1}^p))^{-1/2}$. We let $\mathbf{X}_t^{p+1} = (X_t, \dots, X_{t-p})'$ and let $\| \cdot \|$ denote the Euclidean norm in $\mathbb{R}^p$.

Furthermore, we assume the existence of a nonsingular $p \times p$ matrix A with $A_n - A = O_p(n^{-1/2})$. Then the limiting distribution of $(\widehat{\phi}_n, \widehat{\sigma}_n)$ depends besides ψ, w and χ, only on σ and A, that is,

$$n^{1/2}\big[(\widehat{\phi}_n', \widehat{\sigma}_n)' - (\phi', \sigma)'\big] \to_d \mathcal{N}(0, V), \qquad V = V(A, \sigma; \psi, w, \chi),$$

and, moreover,

$$n^{1/2}\big[\widehat{\phi}_n' - \phi'\big] \to_d \mathcal{N}(0, \widetilde{V}), \qquad \widetilde{V} = \widetilde{V}(A, \sigma; \psi, w)$$

[see Bustos (1982), Theorem 2.3].

The latter shows that treating $\widehat{\sigma}_n$ as fixed σ does not change the asymptotic results. If we do not bootstrap the estimate $\widehat{\sigma}_n$ of the nuisance parameter, we can therefore consider without loss of generality the GM-estimator $\widehat{\phi}_n$ defined by

$$\sum_{t=p+1}^{n} \psi\big((X_t - \widehat{\phi}_1 X_{t-1} - \cdots - \widehat{\phi}_p X_{t-p})\sigma^{-1}\big) w\big(\|A\,\mathbf{X}_{t-1}^p\|\big)\mathbf{X}_{t-1}^p = 0. \tag{20}$$

These estimators can be written after an asymptotically negligible modification at the boundary as statistical functionals $\widehat{\phi}_n = T(F_n^{p+1})$ by defining $T(F^{p+1})$ via

$$\int \Psi\big(\mathbf{X}_t^{p+1}, T\big)\, dF^{p+1}\big(\mathbf{X}_t^{p+1}\big) = \mathbf{0}, \tag{21}$$

where $\Psi\colon \mathbb{R}^{p+1} \times \Theta \to \mathbb{R}^p$,

$$\Psi\big(\mathbf{X}_t^{p+1}, \phi\big) = \psi\big((X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})\sigma^{-1}\big) w\big(\|A\mathbf{X}_{t-1}^p\|\big)\mathbf{X}_{t-1}^p,$$

and F_n^{p+1} denotes the $(p+1)$-dimensional empirical distribution function of $\{\mathbf{X}_t^{p+1}\}_{t=p+1}^n$.

The hope is that bootstrapping GM-estimates yields an asymptotically valid approximation for their true limiting distribution if the associated functional is differentiable.

LEMMA 12. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be the stationary* AR(p) *model* (19). *Let F^{p+1} be the distribution function of $(X_t, \ldots, X_{t+p})'$. Consider the statistical functional T defined by* (21). *We assume that $T(F^{p+1}) = \phi$ (Fisher consistency) and $\int \partial\Psi(\mathbf{X}_t^{p+1}, \tau)/\partial\tau\, dF(\mathbf{X}_t^{p+1})$ is nonsingular at $\tau = \phi$. Furthermore, we assume that ψ and w fulfill the following conditions*:

(a) *ψ is odd, bounded, continuously differentiable.*
(b) *w is even, bounded, continuously differentiable, $w \geq 0$;*

$$w\big(\|x\|\big) \leq \text{const.}\|x\|^{-2} + o\big(\|x\|^{-2}\big), \qquad \|x\| \to \infty.$$

Then T is Fréchet-differentiable at F^{p+1} with respect to the sup-norm.

REMARK. T is Fisher consistent if the innovations ϵ_t are symmetric around 0.

PROOF OF LEMMA 12. A straightforward calculation shows that, under the assumptions of the lemma, the conditions A from Clarke (1983) hold. Since Ψ [cf. (21)] is of bounded variation, condition (5.1) from Clarke (1983) is satisfied with respect to the sup-norm. We achieve the proof by Theorem 5.1 of Clarke (1983). □

THEOREM 2. *Consider the stationary* AR(p) *model* (19). *Assume that the distribution of the innovation* ϵ_t *is dominated by the Lebesgue measure and that the distribution function* F^1 *of* X_t *is continuous . Let* $\widehat{\phi}_n$ *be the GM-estimator for* ϕ *defined by* (20). *Assume that the assumptions of Lemma* 12 *for the associated statistical functional T hold. Denote by* $\widehat{\phi}_n^*$ *the bootstrapped GM-estimator, based on a blocklength l with* $l(n) = O(n^{1/2-\varepsilon})$, $\varepsilon > 0$.

Then

$$\sup_{x \in \mathbb{R}^p} \left| P^*\left[\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n) \le x\right] - P\left[\sqrt{n}(\widehat{\phi}_n - \phi) \le x\right] \right| \to 0 \quad \textit{in probability}.$$

PROOF. The assumption about the distribution of the innovation ϵ_t implies that $\{X_t\}_{t \in \mathbb{Z}}$ is strong-mixing with geometrically decreasing mixing coefficients [see, e.g., Doukhan (1992)], that is,the mixing condition in Theorem 1 holds for any p.

Let T be the functional associated to the GM-estimator. Since T is differentiable at F^{p+1} (Lemma 12), we have

$$T(F_n^{p+1*}) - T(F^{p+1}) = \int \mathrm{IF}(x; F^{p+1})\, dF_n^{p+1*}(x) + o\left(\|F_n^{p+1*} - F^{p+1}\|_\infty\right),$$

$$T(F_n^{p+1}) - T(F^{p+1}) = \int \mathrm{IF}(x; F^{p+1})\, dF_n^{p+1}(x) + o\left(\|F_n^{p+1} - F^{p+1}\|_\infty\right),$$

where F_n^{p+1} and F_n^{p+1*} denote the $(p+1)$-dimensional empirical distribution function and its bootstrap, respectively; IF denotes the influence function. Now the triangle inequality yields

$$\begin{aligned}\sqrt{n}\left(T(F_n^{p+1*}) - T(F_n^{p+1})\right) = \sqrt{n} \int \mathrm{IF}(x; F^{p+1})\, d\left[F_n^{p+1*}(x) - F_n^{p+1}(x)\right] \\ + o\left(\sqrt{n}\|F_n^{p+1*} - F_n^{p+1}\|_\infty\right) \\ + o\left(\sqrt{n}\|F_n^{p+1} - F^{p+1}\|_\infty\right).\end{aligned}$$

The last two terms on the right-hand side are of the order $o_P(1)$ since the empirical process and its bootstrapped process converge weakly. The first term on the right hand side is the standardized bootstrapped mean based on the variables $\mathrm{IF}(\mathbf{X}_t^{p+1}; F^{p+1})$. Since IF is bounded, conditions (B1)–(B3) in Künsch (1989) are satisfied under the assumptions of the theorem (the analogue of Lemma 2 holds, since IF is bounded). Together with Slutsky's theorem this implies that $\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ converges to the limiting distribution of $\sqrt{n}(\widehat{\phi}_n - \phi)$ [cf. Künsch (1989), Theorem 3.5]. □

We close this section with a small simulation study. Consider the GM-estimators of the AR (p)-model (19). In this study, for simplicity we fit only AR (1)-models to the data. By Lemma 12 the GM-estimator for ϕ can be written as

$$T_n = (n-1)^{-1} \sum_{t=2}^{n} \mathrm{IF}(\mathbf{X}_t^2; F^2) + R_n, \qquad R_n = o_P(n^{-1/2}). \tag{22}$$

If the data follow an AR (1)-model, the influence functions are uncorrelated, that is, $E[\mathrm{IF}(\mathbf{X}_t^2; F^2), \mathrm{IF}(\mathbf{X}_s^2; F^2)] = 0$ for $t \neq s$ [see Künsch (1984), (1989)]. In this case the optimal choice for the block length with respect to the mean square error of the bootstrap variance is $l \equiv 1$ (Efron's bootstrap), but as soon as the model does not hold we have to increase $l = l(n)$ with the sample size to get an asymptotically valid bootstrap approximation.

For our study we generated the data from different models with two sample sizes, $n = 481$ and $n = 65$: (a) from an AR (1)-model with $\phi = 0.8$, innovations $\epsilon_t \sim \mathcal{N}(0,1)$ or $\epsilon_t \sim 0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,10)$; (b) from an AR(2)-model with $\phi_1 = 1.372, \phi_2 = -0.677$, innovations, $\epsilon_t \sim \mathcal{N}(0, 0.4982)$.

We have chosen the AR-parameters as in Künsch [(1989), section 5]. The lag(1)-correlations in (a) and (b) are approximately equal, whereas the lag(2)-correlations are considerably different. Moreover, by choosing $\sigma^2 = 0.4982$ in (b) we obtain that $\mathrm{Var}(X_t)$ in the AR(2)-model equals $\mathrm{Var}(X_t)$ in the AR(1)-model with $\epsilon_t \sim \mathcal{N}(0,1)$. In (a) we fit the right model, whereas in (b) the fitted model is wrong.

Let $\sigma_n^2 = n\mathrm{Var}(T_n), \gamma_n = n^{3/2}E[(T_n - E[T_n])^3]$ (skewness) and $\kappa_n = n^2E|T_n - E[T_n]|^4 - 3\sigma_n^2$ (kurtosis). The true quantities have been estimated from 1000 simulations; the bootstrap quantities have been computed with 500 replicates and estimated from 300 simulations (for $n = 481$) or 200 simulations (for $n = 65$), respectively. [Computed with ROBETH; $\psi(x) = \min[\max(x, -c), c]$, $c = 1.345, w(x) = \psi(x)/x$.] (See Table 1.)

As an overall result we have that for appropriate choices of $l > 1$ the non-normality of the true distribution is picked up to a satisfactory extent by the resampling procedure.

In (a), that is, we fit the right model, Efron's bootstrap ($l \equiv 1$) seems to be the best for estimating the variance. This was to be expected according to the theory. However, the resampling procedure still works reasonably for $l(n) \approx cn^{1/3}$, $c \in [1, 1.5]$; this is the optimal order [w.r.t. MSE $(\hat{\sigma}_n^2)$] in the case of the mean [see Künsch (1989)]. Moreover, the estimate for the skewness is far better with $l > 1$. For l too large the procedure is inefficient: $\mathrm{Var}(\hat{\sigma}_n^2)$ is large and even $|\mathrm{Bias}(\hat{\sigma}_n^2)|$ is larger than for some smaller l.

In (b), where we fit a wrong model, the blockwise bootstrap clearly out performs Efron's bootstrap. The bias of $\hat{\sigma}_n^2$ decreases with growing block size, agreeing with the case of the mean, where $\mathrm{Bias}(\hat{\sigma}_n^2) \sim \mathrm{const.}l^{-1}$ $\mathrm{Var}(\hat{\sigma}_n^2)$ seems to be fairly stable with respect to l; for $n = 65$, $\mathrm{Var}(\hat{\sigma}_n^2)$ even decreases with bigger l; for the mean Künsch (1989) showed that $\mathrm{Var}(\hat{\sigma}_n^2) \sim \mathrm{const.}ln^{-1}$. The optimal block length seems to be bigger than $cn^{1/3}$, $c \approx 1$, which is often a good choice in the case of the mean.

We wondered if there is a relevant effect due to the nonlinearity of the estimator. We looked at the linear part of the estimator [see (22)]

$$T_{n;\mathrm{lin}} = (n-1)^{-1} \sum_{t=2}^{n} \mathrm{IF}(\mathbf{X}_t^2; F^2)$$

and computed $\mathrm{Var}^*(T^*_{n;\mathrm{lin}})$ [This can be done without any bootstrap replicates

TABLE 1

	$E[\widehat{\sigma}_n^2]$	S.D.$(\widehat{\sigma}_n^2)$	MSE$(\widehat{\sigma}_n^2)$	E$[\widehat{\gamma}_n]$	S.D.$(\widehat{\gamma}_n)$	E$[\widehat{\kappa}_n]$	S.D$(\widehat{\kappa}_n)$
			AR(1)-model with $\epsilon_t \sim \mathcal{N}(0,1)$, $n = 481$				
True	0.53			−0.089		−0.063	
$l = 1$	0.56	0.10	0.01	−1.86e − 03	2.21e − 03	2.41e − 03	5.84e − 03
$l = 8$	0.56	0.13	0.02	−0.087	4.54e − 03	0.015	8.44e − 03
			AR(1)-model with $\epsilon_t \sim 0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,10)$, $n = 481$				
True	0.45			−0.097		0.084	
$l = 8$	0.45	0.11	0.01	−0.065	2.97e − 03	0.023	4.97e − 03
$l = 12$	0.45	0.12	0.01	−0.072	3.03e − 03	0.019	5.52e − 03
$l = 60$	0.41	0.21	0.04	−0.083	0.014	0.015	0.017
			AR(2)-model with $\epsilon_t \sim \mathcal{N}(0, 0.4982)$, $n = 481$				
True	0.134			−8.17e − 03		7.77e − 04	
$l = 1$	0.507	0.057	0.143	3.55e − 04	1.93e − 03	−4.76e − 03	3.02e − 03
$l = 8$	0.267	0.044	0.020	−9.01e − 03	3.07e − 04	−3.85e − 04	3.31e − 04
$l = 20$	0.189	0.045	5.05e − 03	−8.98e − 03	1.64e − 04	8.31e − 04	9.62e − 05
$l = 30$	0.170	0.050	3.81e − 03	−7.76e − 03	1.26e − 04	−2.60e − 04	6.64e − 05
$l = 60$	0.144	0.061	3.79e − 03	−6.21e − 03	1.69e − 04	−1.22e − 03	5.70e − 05
			AR(1)-model with $\epsilon_t \sim 0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,10)$, $n = 65$				
True	0.54			−0.33		0.28	
$l = 1$	0.52	0.21	0.05	−5.94e − 03	0.01	0.05	0.2
$l = 4$	0.52	0.32	0.10	−0.15	0.06	0.15	0.11
$l = 8$	0.49	0.43	0.19	−0.20	0.15	0.15	0.37
$l = 16$	0.41	0.52	0.29	−0.21	0.39	0.16	1.88
			AR(2)-model with $\epsilon_t \sim \mathcal{N}(0, 0.4982)$, $n = 65$				
True	0.18			−0.0273		0.010	
$l = 1$	0.52	0.16	0.14	−7.01e − 03	5.10e − 03	0.031	9.31e − 03
$l = 4$	0.39	0.13	0.06	−0.0140	2.50e − 03	−5.03e − 04	3.34e − 03
$l = 8$	0.26	0.12	0.02	−0.0237	2.09e − 03	5.44e − 03	1.30e − 03
$l = 16$	0.18	0.12	0.01	−0.0227	2.73e − 03	3.03e − 03	1.90e − 03

since $T_{n;\mathrm{lin}}$ is an arithmetic mean; see Künsch (1989)]. It turns out that the results for $E[\widehat{\sigma}_n^2]$, S.D. $(\widehat{\sigma}_n^2)$ and MSE$(\widehat{\sigma}_n^2)$ are about the same for T_n and $T_{n;\mathrm{lin}}$. Therefore we believe that the error term R_n in the linearization (22) does not play a relevant role, and thus the optimal block length [w.r.t. MSE$(\widehat{\sigma}_n^2)$] would still be $l(n) \sim \mathrm{const}.n^{1/3}$. However, the constant depends on the autocovariances of IF$(\mathbf{X}_t^2; F^2)$, which can be quite different from the autocovariances of X_t.

The study confirms a nice robustness property. At the right model, there is not too much loss by choosing a block length $l(n)$ increasing with the sample size instead of the optimal [w.r.t. MSE$(\widehat{\sigma}_n^2)$] $l \equiv 1$, whereas in a misspecified model the blockwise bootstrap clearly performs better for many block lengths l. Nevertheless further research is needed for an adaptive choice of the block length l.

Acknowledgments. I would like to thank H. Künsch for much advice and many very helpful discussions. I also thank both the referees for their constructive comments.

REFERENCES

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

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

BÜHLMANN, P. (1992). Weak convergence of the bootstrapped multidimensional empirical process for stationary strong-mixing sequences. Research Report 68, Seminar für Statistik, ETH Zürich.

BUSTOS, O. H. (1982). General M-estimates for contaminated p-th order autoregressive processes; consistency and asymptotic normality. *Z. Wahrsch. Verw. Gebiete* **59** 491–504.

CLARKE, B. R (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11** 1196–1205.

DEO, C. (1973). A note on empirical processes of strong-mixing sequences. *Ann. Probab.* **1** 870–875.

DOUKHAN, P. (1994). Mixing: properties and examples. *Lecture Notes in Statist.* **85**. Springer, New York.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1). *Scand. J. Statist.* **16** 97–128.

IBRAGIMOV, I. and LINNIK, Y. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

KÜNSCH, H. R. (1984). Infinitesimal robustness for autoregressive processes. *Ann. Statist.* **12** 843–863.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

MARTIN, R. D. and YOHAI, V. J. (1986). Influence functionals for time series. *Ann. Statist.* **14** 781–818.

NAIK-NIMBALKAR, U. V. and RAJARSHI, M. B. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Ann. Statist.* **22** 980–994.

SEN, P. (1974). Weak convergence of multidimensional empirical processes for stationary ϕ-mixing processes. *Ann. Probab.* **2** 147–154.

YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. Gebiete* **52** 45–57.

YOSHIHARA, K. (1975). Weak convergence of multidimensional empirical processes for strong mixing sequences of stochastic vectors. *Z. Wahrsch. Verw. Gebiete* **33** 133–137.

SEMINAR FÜR STATISTIK
ETH ZENTRUM
CH-8092 ZÜRICH
SWITZERLAND