Block length selection in the bootstrap for time series

by

Peter Bühlmann and
Hans R. Künsch

Research Report No. 72
October 1994

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

Block length selection in the bootstrap for time series

Peter Bühlmann and
Hans R. Künsch

Seminar für Statistik
ETH Zentrum
CH-8092 Zürich, Switzerland

October 1994

Abstract

The blockwise bootstrap is a modification of Efron's bootstrap designed to give correct results for dependent stationary observations. One drawback of the method is that it depends critically on a block length which so far had to be chosen by the user. Here we propose a data-driven method to select this block length. It is based on the equivalence of the blockwise bootstrap variance to a lag weight estimator of a spectral density at the origin. The relevant spectral density is the one of the process given by the influence function of the statistic to be bootstrapped. In this equivalence the block length is the inverse of the bandwidth. We thus apply a recently developed local bandwidth selection procedure to the time series given by the estimated influence function. The lag window corresponding to the blockwise bootstrap is the triangular or Bartlett window. In order to obtain other windows we introduce a correlated weights bootstrap where observations are weighted by an MA-process. The correlations of the weight process correspond to the lag weights.

Keywords: Blockwise bootstrap, Correlated weights bootstrap, Spectral estimation, Influence function, Local bandwidth, Iterative plug-in method.

1 Introduction

Efron's (1979) bootstrap has become a flexible and powerful tool for estimating standard errors and — more ambitiously — for constructing confidence intervals. But it has to be modified when observations are dependent like in time series problems. One possible modification which retains the nonparametric nature of Efron's original proposal is the blockwise bootstrap of Künsch (1989), see also Liu and Singh (1992). Since its introduction several papers have been written which show and extend its validity, see e.g. Bühlmann (1994a, 1995), Bühlmann and Künsch (1995), Götze and Künsch (1993), Naik-Nimbalkar and Rajarshi (1994), Politis and Romano (1992, 1993, 1994).

However one drawback of the procedure is its dependence on a block length ℓ to be chosen by the user which has to increase with sample size n. Asymptotic theory tells us only that the optimal block length for estimating standard errors is of the order $O(n^{1/3})$. The constant belonging to $n^{1/3}$ is a function of the statistic considered and the dependence among the observations. Although $\ell \approx n^{1/3}$ seems to work quite well in many cases, the situation remains unsatisfactory.

In this paper we propose a fully data driven procedure for the selection of ℓ. It is based on an equivalence of the blockwise bootstrap variance of the mean to a lag weight estimator of the spectral density at zero with bandwidth ℓ^{-1}. This makes it possible to apply an iterative plug in method to select the local bandwidth. This method has been introduced by Brockmann et al. (1993) for nonparametric regression and adapted to spectral estimation in Bühlmann (1994b). We summarize this plug in method in Section 5.

The window of the lag weight estimator to which the bootstrap variance is equivalent is Bartlett's triangular window. In spectral estimation this window is not very popular since it is not smooth enough at the origin and its Fourier transform has rather big sidelobes . Other windows can be obtained by a variant of the blockwise bootstrap, called the correlated weights bootstrap, which we discuss in Section 3. This bootstrap weights the original observations with random weights which are correlated over a distance ℓ playing the role of the block length. This variant has already been mentioned briefly in Künsch (1989), p. 1221. Here we discuss it in more detail. In particular we derive a correction factor which is in general necessary to obtain a valid procedure.

Our method can be used for a general statistic, not just the mean. We only require that the statistic is asymptotically equivalent to an arithmetic mean of a function of the observations, the so-called influence function, see Hampel et al. (1986). This means that we have to select the bandwidth for estimating the spectrum of the series given by the influence function. So we have to estimate this influence function first. Different ways to do this are given in Section 4.

Finally in Section 6 we present some simulation examples. They indicate that our procedure leads to a good choice of ℓ in a variety of situations.

2 The blockwise bootstrap

2.1 Definition

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary stochastic process with values in $\mathbb{R}^{\prime}$. Assume that $X_1,\dots,X_n$ have been observed and that we computed a real valued statistic

$$T_n = T_n(X_1,\dots,X_n).$$

The case of vector valued statistics will be discussed briefly at the end of Section 5.2. The aim is to estimate the variance of T_n (or some other parameter of its distribution). First we assume that T_n is symmetric, i.e. invariant under permutations of $(X_1,\dots,X_n)$. Examples include estimates of location, scale and correlations between components of X_t. The blockwise bootstrap resamples blocks of length ℓ from the original data, i.e. the bootstrap sample is

$$X^*_{(j-1)\ell+t} = X_{S_j+t} \quad (1\le j\le b=\lceil n/\ell\rceil; 1\le t\le \ell)$$

where $\lceil x\rceil$ denotes the smallest integer greater or equal to x and

$$S_1,\dots,S_b \quad \text{i.i.d., uniform on } \{0,1,\dots,n-\ell\}.$$

Then we put

$$T_n^* = T_n(X_1^*,\dots,X_n^*)$$

and

$$\hat\sigma^2_{\text{Boot}} = \text{Var}^*(T_n^*)$$

where Var^* denotes the variance of T_n^* conditional on $X_1,\dots,X_n$ (i.e. only $S_1,\dots,S_b$ are random). If n is not a multiple of ℓ, our definition uses only $n+\ell-b\ell$ observations from the last block. In order to make notations easier, we assume in the sequel that $n=b\ell$.

Because T_n is assumed to be independent of the time order of the data, we only need to know how often each X_t appears in the bootstrap sample $X_1^*,\dots,X_n^*$. So let

$$M_t = |\{j|1\le j\le b, t-\ell\le S_j\le t-1\}|$$

Then

$$T_n^* = T_n(\underbrace{X_1,\dots,X_1}_{M_1-\text{times}}, X_2,\dots, \qquad X_{n-1},\underbrace{X_n,\dots,X_n}_{M_n-\text{times}}).$$

This will be the basis for the correlated weights bootstrap in Section 3.

When T_n is a statistic measuring serial dependence of (X_t), it depends on the time order of the observations. In such a case one should not apply the above definitions directly because the junction of independent blocks introduces artifacts in the distribution of T_n^*. But usually after a minor modification at the boundary T_n is a symmetric function of the m-tuples $X_t' = (X_t, X_{t+1},\dots,X_{t+m-1})$ for suitable m. Thus a natural remedy suggested in Künsch (1989) is to apply the above procedure on the level of X_t', i.e. to resample blocks of blocks. This procedure has been further investigated by Politis and Romano (1992) and by Bühlmann and Künsch (1995). At least for fixed m, our results here remain valid.

2.2 Bootstrap of the mean as spectral estimation

Let us consider real valued X_t and the arithmetic mean

$$T_n = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

It is well known that

$$\mathrm{Var}(T_n) = \frac{1}{n} \sum_{k=-n+1}^{n-1} (1 - \frac{|k|}{n}) \mathrm{Cov}(X_0, X_k) \sim \frac{2\pi}{n} f(0)$$

where $f(\lambda)$ denotes the spectral density of (X_t).

Moreover

$$T_n^* = \frac{1}{n} \sum_{t=1}^{n} M_t X_t$$

so that the bootstrap moments of T_n^* are easily calculated (cf. Künsch (1989), Sections 3.1 - 3.2).

Lemma 1: If $\ell \leq (n+1)/2$, then

i) $E^*[T_n^*] = \sum_{t=1}^{n} \alpha_t X_t$ with $\alpha_t = (n - \ell + 1)^{-1} \min(t/\ell, 1, (n+1-t)/\ell)$,

ii) $\mathrm{Var}^*(T_n^*) = (n(n-\ell+1))^{-1} \sum_{s,t=1}^{n} \beta_{st} (X_t - E^*[T_n^*])(X_s - E^*[T_n^*])$ where

$$\beta_{t,s} = \begin{cases} \min(t,s)/\ell & \text{if } t \leq \ell & \text{and } s \leq \ell \\ (n+1-\max(t,s))/\ell & \text{if } t \geq n+1-\ell & \text{and } s \geq n+1-\ell \\ \max(1 - |t-s|/\ell, 0) & \text{otherwise} & \end{cases}$$

Proof: see Appendix.

Hence up to some boundary effects

$$E^*[T_n^*] = T_n$$

and

$$\mathrm{Var}^*(T_n^*) = \frac{1}{n} \sum_{k=-\ell+1}^{\ell-1} (1 - \frac{|k|}{\ell}) \hat{R}(k)$$

where $\hat{R}(k)$ are the empirical covariances. In other words $n\mathrm{Var}^*(T_n^*)/(2\pi)$ is a lag weight estimate of the spectrum at zero with the triangular or Bartlett window and the block length ℓ is the inverse of the bandwidth. Moreover, as the proof of the Lemma shows, the triangular window arises as $\mathrm{Corr}^*(M_t, M_s)/n$.

3 The correlated weights bootstrap

In spectral estimation one usually prefers a window w which is smoother at the origin than the triangular window. As the previous discussion suggests, this can be achieved by modifying (M_t) such that

$$\mathrm{Cov}^*(M_t, M_s) \propto w((t-s)/\ell).$$

We have not been able to achieve this with integer M_t's. So we restrict the class of statistics T_n by assuming that T_n is defined for all weighted samples $(x_t, p_t)_{1 \leq t \leq n}$ with $x_t \in \mathbb{R}^{\prime}$ and $p_t \in [0,1]$ satisfying $\sum_{t=1}^n p_t = 1$. This is no serious limitation. We generate M_t as a moving average

$$M_t = \sum_{k=-\ell}^{\ell} b_k \xi_{t-k}$$

with $b_k \geq 0, b_{-k} = b_k, \xi_t \geq 0$ i.i.d.. The bootstrap statistic is then

$$T_n^* = T_n((X_t, M_t / \sum_{s=1}^n M_s)_{t=1,\ldots n}).$$

However for the estimation of $\mathrm{Var}(T_n)$ we should multiply $\mathrm{Var}^*(T_n^*)$ with a correction factor. To explain why, consider the arithmetic mean $T_n = n^{-1} \sum X_t$ where

$$T_n^* = \sum_{t=1}^n M_t X_t / \sum_{t=1}^n M_t.$$

Even in this simple case we cannot compute the bootstrap moments of T_n^* exactly because the denominator is also random. So we rely on a linearization. Without loss of generality we take

$$\sum_{k=-\ell}^{\ell} b_k = 1, \;\; E[\xi_t] = 1$$

so that $E[M_t] = 1$. Then for $j = 1, 2$

$$(\sum_{t=1}^n M_t)^{-j} \approx n^{-j}(1 - jn^{-1} \sum_{t=1}^n (M_t - 1)).$$

Under the assumptions on (b_k) and ξ_t we are going to make, the error in the approximation above will be $O_{p^*}(\ell n^{-j-1})$. Using this we obtain with $\mu = E[X_t]$

$$T_n^* \approx U_n^* = \mu + \frac{1}{n} \sum_{t=1}^n M_t (1 - \frac{1}{n} \sum_{s=1}^n (M_s - 1))(X_t - \mu)$$

and

$$\begin{aligned} (T_n^* - T_n)^2 &= (\sum_{t=1}^n M_t (X_t - T_n))^2 (\sum_{t=1}^n M_t)^{-2} = (\sum_{t=1}^n (M_t - 1)(X_t - T_n))^2 (\sum_{t=1}^n M_t)^{-2} \\ &\approx V_n^* = \frac{1}{n^2} \sum_{t,s=1}^n (M_t - 1)(M_s - 1)(1 - \frac{2}{n} \sum_{u=1}^n (M_u - 1))(X_t - T_n)(X_s - T_n). \end{aligned}$$

From this we obtain approximate bootstrap moments. Let

$$v_k = \sum_{j=\max(-\ell,-\ell-k)}^{\min(\ell,\ell-k)} b_j b_{j+k}$$

be the convolution of (b_k) with itself. Then we have

Lemma 2: Let $\mathrm{Var}(\xi_t) = \sigma_\xi^2$, $E[(\xi_t - 1)^3] = \kappa_\xi$ and assume $4\ell \le n$. Then

i) $E^*[U_n^*] = T_n - \sigma_\xi^2 n^{-2} \sum \alpha_t (X_t - \mu)$ where

$$\alpha_t = \begin{cases} \sum_{k=-2\ell}^{t-1} v_k & \text{if} \quad t \le 2\ell \\ 1 & \text{if} \quad 2\ell < t < n+1-2\ell \\ \sum_{k=t-n}^{2\ell} v_k & \text{if} \quad t \ge n+1-2\ell \end{cases}$$

ii) $E^*[V_n^*] = n^{-1} \sum_{k=-2\ell}^{2\ell} v_k \hat{R}(k)(\sigma_\xi^2 - 2\kappa_\xi/n) + 2n^{-3}\kappa_\xi \sum \gamma_{t,s}(X_t - T_n)(X_s - T_n)$ where

$$\gamma_{t,s} = \sum_{j=\max(-\ell,-\ell-t+s)}^{\min(\ell,\ell-t+s)} b_j b_{j+t-s} \left(\sum_{k=-\ell}^{j-t} b_k + \sum_{k=n-t+j}^{\ell} b_k \right)$$

Proof: see Appendix.

Note that $\gamma_{t,s} = 0$ if $|t-s| > 2\ell$ or $2\ell < t < n+1-2\ell$. Hence the contribution of the last term to $E^*[V_n^*]$ is small.

Thus if we define,

$$\hat{\sigma}^2_{\mathrm{Boot}} = E^*[(T_n^* - T_n)^2](v_0(\sigma_\xi^2 - 2\kappa_\xi/n))^{-1} \tag{3.1}$$

we have

$$\hat{\sigma}^2_{\mathrm{Boot}} \approx \frac{1}{n} \sum_{k=-2\ell}^{2\ell} v_k/v_0 \hat{R}(k).$$

We use (3.1) also for a general T_n, not only for the arithmetic mean.

By an appropriate choice of b_k we can obtain the desired behavior of v_k. We will always take

$$b_k \sim \beta(k/\ell)/\ell \tag{3.2}$$

with $\int_{-1}^{1} \beta(x)dx = 1$. Then

$$v_k/v_0 \sim w(k/\ell) \tag{3.3}$$

where $w(x) = \int_{-1}^{1} \beta(y)\beta(y+x)dy(\int_{-1}^{1} \beta(y)^2 dy)^{-1}$.

There is still much freedom for choosing the law F of ξ_t. If we want the correction factor in (3.1) to disappear for $n \to \infty$, we need

$$\sigma_\xi^2 = v_0^{-1} = (\sum_{k=-\ell}^{\ell} b_k^2)^{-1} \sim \ell(\int_{-1}^{1} \beta(x)^2 dx)^{-1}.$$

This can be achieved by taking $\xi_t \sim \Gamma(v_0, v_0)$ at the expense of having a very skewed distribution. An alternative would be to take the ξ_t uniform on [0.5, 1.5]. At the moment we lack the experience to make definite recommendations, but we conjecture that this choice will be of secondary importance.

The parameter ℓ in the correlated weights bootstrap is the correlation length of the weight sequence M_t. Its role is completely analogous to the block length in the blockwise bootstrap. So we continue to call ℓ the block length. It is the basic tuning constant of the bootstrap. We discuss the selection of ℓ in Section 5. Note that the coefficients (b_k) and the weights (M_t) can be defined also for a real valued ℓ.

4 Linearization of a general statistic

So far our discussion of the properties of the bootstrap has been restricted to the mean. The importance of this special case comes from the fact that most statistics of interest behave asymptotically like an arithmetic mean. We assume from now on that the statistic T_n converges a.s. to an asymptotic value T_∞ and that it has an influence function IF with $E[\mathrm{IF}(X_t)] = 0$ such that

$$T_n = T_\infty + \frac{1}{n}\sum_{t=1}^{n} \mathrm{IF}(X_t) + o_p(n^{-1/2}) \tag{4.1}$$

and

$$T_n^* = T_n + \frac{1}{n}\sum_{t=1}^{n} M_t(\mathrm{IF}(X_t) - \frac{1}{n}\sum_{s=1}^{n} \mathrm{IF}(X_s)) + o_{p^*}(n^{-1/2}). \tag{4.2}$$

If this is true, then the asymptotic variances of T_n and T_n^* depend only on the linear parts $n^{-1}\sum \mathrm{IF}(X_t)$ and $n^{-1}\sum M_t \mathrm{IF}(X_t)$ respectively. If in addition $n\mathrm{Var}^*(T_n^*)$ converges to the asymptotic variance, then - by the discussion in Section 2 - $n\mathrm{Var}^*(T_n^*)$ is equivalent to a lag weight estimator of the spectral density at zero of the transformed variables $Y_t = \mathrm{IF}(X_t)$. This justifies our procedure to determine the block length ℓ by using a data driven choice of the bandwidth for a lag weight estimator of the spectrum of (Y_t). This bandwidth choice will be presented in Section 5.

The influence function $\mathrm{IF}(X_t)$ depends on the unknown underlying distribution P of X_t. So we have to estimate it. Usually there are several ways to do this. Often T_n is obtained by applying a functional T to the empirical distribution P_n:

$$T_n = T(P_n),\ P_n = \frac{1}{n}\sum_{t=1}^{n} \Delta_{X_t}$$

with Δ_x denoting a point mass at x. Then

$$\mathrm{IF}(x) = \mathrm{IF}(x, P, T) = \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)P + \varepsilon\Delta_x) - T(P)}{\varepsilon} \tag{4.3}$$

cf. Hampel et al. (1986). In this case we can estimate IF by replacing P by P_n in (4.3):

$$\hat{\mathrm{IF}}(x) = \mathrm{IF}(x, P_n, T) = \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)P_n + \varepsilon\Delta_x) - T(P_n)}{\varepsilon}.$$

Alternatively we might use the jackknife which replaces the derivative by a finite difference

$$\begin{aligned}
\hat{\mathrm{IF}}(X_t) &= n[T((1-\frac{1}{n})\frac{1}{n-1}\sum_{s\neq t}\Delta_{X_s}+\frac{1}{n}\Delta_{X_t})-T(\frac{1}{n-1}\sum_{s\neq t}\Delta_{X_s})] \\
&= n[T_n(X_1,\ldots,X_n)-T_{n-1}(X_1,\ldots,X_{t-1},X_{t+1},\ldots,X_n)].
\end{aligned}$$

This makes sense even if T_n is not of the form $T(P_n)$.

Sometimes it is important to notice that we have to estimate the influence function only up to a proportionality constant, i.e. if $IF(x,P,T) = M(P)^{-1}IF_{\mathrm{prop}}(x,P,T)$, it is sufficient to estimate $IF_{\mathrm{prop}}(x)$. This is true since our bandwidth choice is independent of the scale of the data. For instance in the case T_n is an M-estimator, i.e. defined implicitly by

$$\sum_{t=1}^{n}\psi(X_t,T_n)=0,$$

the influence function is (see Hampel et al., 1986)

$$\mathrm{IF}(x)=\mathrm{IF}(x,P,T)=-(\frac{\partial}{\partial\theta}E[\psi(X_t,\theta)]|_{\theta=T(P)})^{-1}\psi(x,T(P)).$$

Hence it is sufficient to estimate $\psi(x,T(P))$, e.g. by $\psi(x,T_n)$. This is especially useful in the case of the α-quantile which is an M-estimator with

$$\psi(x,\theta)=(1-\alpha)1_{[x<\theta]}-\alpha 1_{[x>\theta]}.$$

Thus

$$\mathrm{IF}(x,T,P)=\begin{cases} -(1-\alpha)/p(T(P)) & x<T(P) \\ \alpha/p(T(P)) & x>T(P)\end{cases}$$

where p is the density of P (assumed to exist). It would be awkward if we had to estimate this density.

As we have seen the computation and estimation of IF is usually not difficult. However a rigorous proof of (4.1) and (4.2), i.e. to obtain information on the order of the remainder from the linearization, is much more complicated. An elegant and powerful technique is to combine empirical process theory with differential calculus in function spaces. In i.i.d. settings this has become well established, see e.g. Gill (1989) or Arcones and Giné (1992). Under dependence only the empirical process theory has to be developed, the differential calculus part remains unchanged. Relevant references to empirical processes are Billingsley (1968), Deo (1973), Bickel and Wichura (1971), Yoshihara (1975) and Arcones and Yu (1994) for (X_t) and Naik-Nimbalkar and Rajarshi (1994) and Bühlmann (1993, 1994a, 1995) for (X_t^*). Taking all these results into account, we obtain (4.1) and (4.2) for a large number of interesting statistics.

5 Iterative plug-in estimation of tuning constants

5.1 Local bandwidth estimation for spectra

The method we are going to use is an adaptation of a method proposed by Brockmann et al. (1993) for nonparametric regression to the problem of spectral estimation. This adaptation (including proofs) is the subject of Bühlmann (1994b), but in order to make this paper self-contained we give here a description of the procedure.

Let (Y_t) be a real valued stationary process with $E[Y_t] = 0$, autocovariances $R(k)$ and spectrum

$$f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} R(k) e^{-ik\lambda} \ (-\pi \leq \lambda \leq \pi).$$

We denote by $\hat{R}(k)$ the empirical autocovariances and by $I_n(\lambda)$ the periodogram:

$$I_n(\lambda) = (2\pi n)^{-1} |\sum_{t=1}^{n} Y_t e^{-i\lambda t}|^2 = (2\pi)^{-1} \sum_{k=-n+1}^{n-1} \hat{R}(k) e^{-ik\lambda}.$$

It is well known that $I_n(\lambda)$ is not consistent for $f(\lambda)$, but the following lag weight estimator is

$$\hat{f}(\lambda; b) = f_w(\lambda; b) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \hat{R}(k) w(kb) e^{-ik\lambda}.$$

We call b the bandwidth and assume that $b \to 0, nb \to \infty$ as $n \to \infty$. For the window w we assume that $0 \leq w \leq 1, w(-x) = w(x), w(x) = 0$ for $|x| > c$ and

$$w(x) = 1 - \beta(w)|x|^j + O(|x|^{j+1}) \text{ as } x \to 0$$

for some $j \in \mathbb{N}$ and $\beta(w) > 0$. We call j the characteristic exponent of w. It is usually 1 or 2. Main examples with $c = 1$ are for $j = 1$ the Bartlett window

$$w(x) = 1 - |x| \quad (|x| \leq 1)$$

and for $j = 2$ the Tukey-Hanning window

$$w(x) = (1 + \cos(\pi x))/2 \quad (|x| \leq 1).$$

Under suitable regularity conditions (cf. Priestley, 1981) we have the following expressions for the asymptotic bias and variance of $\hat{f}(\lambda; b)$:

$$E[\hat{f}(\lambda; b)] - f(\lambda) \sim -b^j \beta(w) f^{(j)}(\lambda),$$

$$\mathrm{Var}(\hat{f}(\lambda; b)) \sim (bn)^{-1} B(w) \iota(\lambda) f(\lambda)^2,$$

where

$$f^{(j)}(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} |k|^j R(k) e^{-ik\lambda}$$

is a generalized derivative,

$$B(w) = \int_{-c}^{c} w(x)^2 dx$$

and

$$\iota(\lambda) = \begin{cases} 2 & \lambda \in \{-\pi, 0, \pi\} \\ 1 & \text{otherwise.} \end{cases}$$

This leads in the usual way to the asymptotic mean square error at λ (AMSE $(\lambda; b)$) and the asymptotic mean integrated square error (AMISE(b)). The optimal local bandwidth $b_{\text{opt}}(\lambda)$ is obtained by minimising AMSE $(\lambda; b)$. The result is

$$b_{\text{opt}}(\lambda) = n^{-1/(2j+1)}\left(\frac{B(w)\iota(\lambda)f(\lambda)^2}{2j\beta(w)^2 f^{(j)}(\lambda)^2}\right)^{1/(2j+1)}$$

Similarly the optimal global bandwidth is

$$b_{\text{opt}} = n^{-1/(2j+1)}\left(\frac{B(w)\int f(\lambda)^2 d\lambda}{2j\beta(w)^2 \int f^{(j)}(\lambda)^2 d\lambda}\right)^{1/(2j+1)},$$

the integrals being over $[-\pi, \pi]$. These expressions can however not be used directly because the right hand sides depend also on the unknown $f(\lambda)$. So the idea is to plug in estimates for the terms on the right hand side. For the integral $\int f(\lambda)^2 d\lambda$ we do not need a bandwidth. It can be estimated consistently by

$$\int I_n(\lambda)^2 d\lambda/2.$$

For the other terms we use lag weight estimators, but typically with different windows and bandwidths. Define

$$\hat{f}_w^{(j)}(\lambda; b) = (2\pi)^{-1}\sum |k|^j \hat{R}(k) w(kb) e^{ik\lambda}.$$

Our iterative plug in method is defined as follows for $j = 1$ or $j = 2$ respectively:

(I) Start with $b_0 = n^{-1/j}$, independently of the data.

(II) Iterate the global bandwidth for $i = 1, 2, 3, 4$:

$$b_i = n^{-1/(2j+1)}\left(\frac{B(w)\int I_n(\lambda)^2 d\lambda}{4j\beta(w)^2 \int \hat{f}_{\overline{w}}^{(j)}(\lambda; b_{i-1}n^{\zeta})^2 d\lambda}\right)^{1/(2j+1)}.$$

(III) Estimate the local bandwidth by

$$\hat{b}(\lambda) = n^{-1/(2j+1)}\left(\frac{B(w)f_{\tilde{w}}(\lambda; b_4 n^{\zeta})^2 \iota(\lambda)}{2j\beta(w)^2 \hat{f}_{\overline{w}}^{(j)}(\lambda, b_4 n^{\zeta})^2}\right)^{1/(2j+1)}.$$

Here $\tilde{w}$ is the Tukey-Hanning window, $\overline{w}$ is the split cosine window

$$\overline{w}(x) = \begin{cases} 1, |x| \le 0.8 \\ (1 + \cos(5(x - 0.8)\pi))/2, \quad 0.8 \le |x| \le 1 \end{cases}$$

and $\zeta = 4/21$ for $j = 1$ and $\zeta = 4/45$ for $j = 2$ respectively.

Some parts of the procedure like the number of iterations in (II) or the inflation factor n^{ζ} are motivated by asymptotic theory. Others like the window shape $\overline{w}$ or that we do not iterate the local bandwidth are based on empirical evidence. More background and discussion can be found in Bühlmann (1994b). There it is also proved that under conditions on the summability of higher order cumulants of (Y_t)

$$\hat{b}(\lambda) = b_{\text{opt}}(\lambda)(1 + O_p(n^{-2/7}))$$

for $j = 1$ and

$$\hat{b}(\lambda) = b_{\text{opt}}(\lambda)(1 + O_p(n^{-2/9}))$$

for $j = 2$.

5.2 Block length choice for the bootstrap

Summarizing the discussions from the previous sections, our procedure is as follows

(I) Estimate the influence function up to a proportionality constant as $\hat{\text{IF}}_{\text{prop}}(X_t)$ by any procedure proposed in Section 4.

(II) Compute the estimated optimal bandwidth $\hat{b}(0)$ according to Section 5.1 for $Y_t = \hat{\text{IF}}_{\text{prop}}(X_t)$. The relevant window w is the Bartlett window in case of the blockwise bootstrap, whereas for the correlated weights bootstrap w is given in (3.2)-(3.3). In particular if we take the moving average weights b_k for M_t to be the triangular weights, then w is the Parzen window (Priestley, 1981, p. 444) which has characteristic exponent $j = 2$.

(III) Choose as the esimated block length $\hat{\ell}$ the integer closest to $\hat{b}(0)^{-1}$ for the blockwise bootstrap and $\hat{\ell} = \hat{b}(0)^{-1}$ for the correlated weights bootstrap respectively.

The effect of step (I) on the procedure is typically asymptotically negligible. A sufficient condition is that there exists a bounded function h and a random variable $U_n = O_p(n^{-1/2})$ such that

$$\sup_x |M(P) \cdot \text{IF}(x) - \hat{\text{IF}}_{\text{prop}}(x) - h(x)U_n| = O_p(n^{-1}).$$

This implies that the empirical covariances of $\hat{\text{IF}}_{\text{prop}}(X_t)$ are equal to those of $M(P)\text{IF}(X_t)$ up to an error of order $O_p(n^{-1})$. Going through the steps in the proofs of Bühlmann (1994b) one can see that such an error is asymptotically negligible. These arguments show that the following is true.

Theorem 5.1 *Let $(X_t)_{t\in Z}$ be stationary with $\sum_{k=0}^{\infty}(k+1)^6\alpha^{\delta/(6+\delta)}(k) < \infty$ where $(\alpha(k))$ are the strong (Rosenblatt) mixing coefficients of (X_t). Furthermore let $Y_t = IF(X_t)$ have spectral density f and assume that $f^{(j)}(0) \neq 0$ and that $E|Y_t|^{8+\delta} < \infty$ $(\delta > 0)$. Finally denote by ℓ_{opt} the block length minimising the asymptotic mean square error (AMSE) of $\hat{\sigma}^2_{\text{Boot}}$ for estimating $\sum_{k=-\infty}^{\infty} E[Y_0Y_k]$. Then*

$$\hat{\ell} = \ell_{\text{opt}}(1 + O_p(n^{-2/7})) \text{ for } j = 1,$$
$$\hat{\ell} = \ell_{\text{opt}}(1 + O_p(n^{-2/9})) \text{ for } j = 2.$$

Finally for vector valued estimators $T_n = (T_{n,1}, \ldots, T_{n,p})'$ we suggest a similar approach. We propose to minimize

$$\sum_{i=1}^{p} E|n\hat{\sigma}^2_{\mathrm{Boot},i} - \sigma^2_{\infty,i}|^2/\sigma^4_{\infty,i},$$

where $\sigma^2_{\infty,i} = \lim nVar(T_{n,i})$ and $\hat{\sigma}^2_{\mathrm{Boot},i}$ is the bootstrap variance of $T^*_{n,i}$. Then by a linearization, as discussed in Section 4, the criterion above becomes

$$\sum_{i=1}^{p} E|n\hat{\sigma}^2_{\mathrm{Boot},i} - \sigma^2_{\infty,i}|^2/\sigma^4_{\infty,i} \approx \sum_{i=1}^{p} E|\hat{f}_i(0) - f_i(0)|^2/f_i(0)^2,$$

where f_i is the spectral density of the i-th component of IF. The optimal local bandwidth is thus

$$b_{\mathrm{opt}}(\lambda) = n^{-1/(2j+1)}\left(\frac{B(w)p\iota(\lambda)}{2j\beta(w)^2\sum_{i=1}^{p}(f_i^{(j)}(0)/f_i(0))^2}\right)^{1/(2j+1)}.$$

In analogy to our former algorithm we propose the following procedures for $j = 1, 2$

(I) Estimate the influence function.

(II) Start with $b_{0,1} = \ldots = b_{0,p} = n^{-1/j}$ independently of the data.

(III) Iterate the global bandwidths separately for each component of the estimated influence function as in the one-dimensional case, giving the results $b_{4,1}, \ldots, b_{4,p}$.

(IV) Estimate the local bandwidth as

$$\hat{b}(0) = n^{-1/(2j+1)}\left(\frac{pB(w)}{j\beta(w)^2\sum_{i=1}^{p}(\hat{f}^{(j)}_{\overline{w},i}(0; b_{4,i}n^{\zeta})/\hat{f}_{\tilde{w},i}(0; b_{4,i}n^{\zeta}))^2}\right)^{1/(2j+1)}.$$

(V) Choose the block length $\hat{\ell}$ as the inverse of $\hat{b}(0)$ (or the closest integer thereof).

The component-wise approach in step (III) should adapt better to the variable behavior of f_i in different components.

6 Simulations

We investigate our procedures for the mean and the median. For the median, we base our estimate for the block length on $\hat{IF}_{\mathrm{prop}}(X_t) = \mathrm{sign}(X_t - \mathrm{med}(X_i)_{i=1}^n)$ (see Section 4).

For the correlated weights bootstrap we generated the weights by taking $b_k = \max(1 - |k|/\ell, 0)/\ell$ and $\xi_t \sim \Gamma(v_0, v_0)$, where $v_0 = 2/(3\ell) + 1/(3\ell^3)$. The function w is then the Parzen window on $[-2, 2]$ and the correction factor in (3.1) equals $(1 - 2\kappa_\xi v_0/n)^{-1} = (1 - 4/(nv_0))^{-1}$ (see Section 3), i.e. the factor might correct the error terms but is not necessary for consistency.

We consider the following autoregressive models $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$:
(M1) AR(1): $\phi_1 = 0.8$, ε_t i.i.d. $\mathcal{N}(0,1)$.
(M2) AR(2): $\phi_1 = 1.372$, $\phi_2 = -0.677$, ε_t i.i.d. $\mathcal{N}(0, 0.4982)$.
(M3) AR(5): $\phi_1 = 0.9$, $\phi_2 = -0.4$, $\phi_3 = 0.3$, $\phi_4 = -0.5$, $\phi_5 = 0.3$, ε_t i.i.d. $\mathcal{N}(0,1)$.
These are the same models as in Bühlmann (1994b), (M1) and (M2) are also investigated in Künsch (1989) and Bühlmann (1993,1994a). By choosing $\sigma^2 = 0.4982$ in (M2) we force $\mathrm{Var}(X_t)$ in (M2) to be the same as $\mathrm{Var}(X_t)$ in (M1). The models (M2) and (M3) exhibit a 'damped-periodic' autocorrelation function, where the correlations can be negative; this phenomenon does not occur in (M1), where all autocorrelations are positive. We choose the sample sizes $n = 480$ and $n = 120$.

All our results are based on 200 simulations and 500 bootstrap replicates. For each of the 200 simulations we generate the replicates by an individual data-driven block length ℓ. We only report the bootstrap estimates for the variance. The bootstrap for the skewness and curtosis is in these cases not very accurate (cf. Bühlmann (1993)). We denote by $\sigma_n^2 = nVar(T_n)$, $(\sigma_n^2)^* = nVar^*(T_n^*)$, $RMSE^* = MSE((\sigma_n^2)^*)/\sigma_n^4$ and indicate by $*bl$, $*cw$ the correspondence to the blockwise or correlated weights bootstrap respectively. The true variance σ_n^2 is based on 1000 simulations.

$T_n = n^{-1} \sum_{t=1}^{n} X_t$:

$n = 480$	σ_n^2	$E[(\sigma_n^2)^{*bl}]$	$S.D.((\sigma_n^2)^{*bl})$	$RMSE^{*bl}$	$E[(\sigma_n^2)^{*cw}]$	$S.D.((\sigma_n^2)^{*cw})$	$RMSE^{*cw}$
(M1)	24.47	18.72	5.89	0.113	20.80	5.34	0.070
(M2)	5.79	6.40	1.47	0.076	6.19	1.47	0.069
(M3)	6.40	5.05	1.04	0.071	4.99	1.25	0.087

$n = 120$	σ_n^2	$E[(\sigma_n^2)^{*bl}]$	$S.D.((\sigma_n^2)^{*bl})$	$RMSE^{*bl}$	$E[(\sigma_n^2)^{*cw}]$	$S.D.((\sigma_n^2)^{*cw})$	$RMSE^{*cw}$
(M1)	26.04	12.93	6.85	0.323	17.07	9.51	0.252
(M2)	5.81	7.02	2.27	0.196	7.83	2.60	0.322
(M3)	6.14	4.90	1.52	0.102	5.46	1.74	0.093

$T_n = \mathrm{med}\{X_t\}_{t=1}^{n}$:

$n = 480$	σ_n^2	$E[(\sigma_n^2)^{*bl}]$	$S.D.((\sigma_n^2)^{*bl})$	$RMSE^{*bl}$	$E[(\sigma_n^2)^{*cw}]$	$S.D.((\sigma_n^2)^{*cw})$	$RMSE^{*cw}$
(M1)	26.31	21.54	8.90	0.147	24.06	8.70	0.117
(M2)	8.14	9.61	3.24	0.191	9.52	3.22	0.185
(M3)	7.69	7.08	2.03	0.076	7.22	2.49	0.108

$n = 120$	σ_n^2	$E[(\sigma_n^2)^{*bl}]$	$S.D.((\sigma_n^2)^{*bl})$	$RMSE^{*bl}$	$E[(\sigma_n^2)^{*cw}]$	$S.D.((\sigma_n^2)^{*cw})$	$RMSE^{*cw}$
(M1)	28.45	15.50	10.16	0.335	21.03	15.13	0.351
(M2)	7.66	10.27	4.82	0.512	12.16	5.87	0.932
(M3)	7.56	6.82	3.24	0.193	7.06	3.98	0.282

From a mean square errror point of view, both bootstrap procedures have about the same performance and it seems difficult to detect a general superiority of one method over the other. Though in general the correlated weights bootstrap for fixed block lengths is better, a slower error rate in the estimation of the optimal block length might explain the almost equal performance.

As expected, the theory is reflected best in model (M1), where all autocorrelations are positive. In this case, the correlated weights bootstrap has a smaller bias at the expense of variability. Model (M2) seems to be most difficult to estimate. The correlated weights bootstrap has here the problem, that, except for the mean with $n = 480$, the correction factor in (3.2) corrects towards the wrong direction and makes the estimate worse than without correction.

To get an idea about the effect of nonlinearity, we computed also the bootstrap variances for the linearization of the median under the knowledge of the true marginal distribution of (X_t). This procedure (which cannot be used in practice) has comparable bias but leads to much smaller standard deviations for $(\sigma_n^2)^*$. Therefore we believe that the bootstrap pays a price mainly for estimating the true underlying marginal distribution of (X_t) and not because of nonlinearity of the median.

Appendix: Proofs

Proof of Lemma 1:

By definition

$$M_t = \sum_{j=1}^{b} \sum_{s=\max(t-\ell,0)}^{\min(t-1,n-\ell)} 1_{[S_j=s]}.$$

and

$$P^*[S_j = s] = \ (n-\ell+1)^{-1} \ (0 \le s \le n-\ell).$$

Hence i) follows immediately. Furthermore by the independence of the S_j's

$$\begin{aligned} \mathrm{Cov}^*(M_t, M_u) &= b \sum_{s=\max(t-\ell,0)}^{\min(t-1,n-\ell)} \sum_{r=\max(u-\ell,0)}^{\min(u-1,n-\ell)} \mathrm{Cov}^*(1_{[S_j=s]}, 1_{[S_j=r]}) \\ &= (n-\ell+1)^{-1} b \sum_{s=\max(t-\ell,0)}^{\min(t-1,n-\ell)} \sum_{r=\max(u-\ell,0)}^{\min(u-1,n-\ell)} \delta_{sr} - n\ell\alpha_t\alpha_u, \end{aligned}$$

δ_{sr} being the Kronecker delta. The first term on the right is easily seen to be equal to $n(n-\ell+1)^{-1}\beta_{tu}$. From this ii) follows because

$$\begin{aligned} \mathrm{Var}^*(T_n^*) &= \mathrm{Var}^*(n^{-1} \textstyle\sum_{t=1}^n M_t(X_t - E^*[T_n^*])) \\ &= n^{-2} \textstyle\sum_{t,u=1}^n \mathrm{Cov}^*(M_t, M_u)(X_t - E^*[T_n^*])(X_u - E^*[T_n^*]) \end{aligned}$$

and $\sum_{t=1}^n \alpha_t = 1$. □

Proof of Lemma 2:

It holds that

$$\mathrm{Cov}^*(M_t, M_s) = \sigma_\xi^2 v_{t-s}$$

Hence by the definition of U_n^*

$$E^*[U_n^*] = \mu + n^{-1} \sum_{t=1}^{n} (1 - n^{-1}\sigma_\xi^2 \sum_{s=1}^{n} v_{t-s})(X_t - \mu).$$

Because $v_k = 0$ for $|k| > 2\ell$ and $\sum_{k=-2\ell}^{2\ell} v_k = (\sum_{k=-\ell}^{\ell} b_k)^2 = 1$, i) follows easily.

For ii) note that

$$E^*[(M_t - 1)(M_s - 1)(M_u - 1)] = \kappa_\xi \sum_j b_j b_{t-s+j} b_{t-u+j}$$

where we put $b_k = 0$ for $|k| > \ell$. Since $\sum b_k = 1$ we obtain

$$\sum_{u=1}^{n} E^*[(M_t - 1)(M_s - 1)(M_u - 1)] = \kappa_\xi (v_{t-s} - \gamma_{t,s}).$$

From this ii) follows easily. □

References:

Arcones, M.A. and Giné, E. (1992). *On the bootstrap of M-estimators and other statistical functionals.* In: Exploring the Limits of Bootstrap, edited by LePage, R. and Billard, L. Wiley, New York. pp. 13-47.

Arcones, M.A. and Yu, B. (1994). *Central limit theorems for empirical and U-processes of stationary mixing sequences.* J. Theoret. Probab. **7** 47-71.

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.

Brockmann, M., Gasser, Th. and Herrmann, E. (1993). *Locally adaptive bandwidth choice for kernel regression estimators.* J. Amer. Statist. Assoc. **88** 1302-1309.

Bühlmann, P. (1993) *The blockwise bootstrap in time series.* Ph.D. thesis, ETH Zürich.

Bühlmann, P. (1994a). *Blockwise bootstrapped empirical process for stationary sequences.* To appear in Ann. Statist.

Bühlmann, P. (1994b) *Locally adaptive lag-window spectral estimation.* Tech. Rep., Statistics Dept., Univ. of Calif. Berkeley.

Bühlmann, P. (1995) *The blockwise bootstrap for general empirical processes of stationary sequences.* Submitted for publication in Stoch. Processes and their Appl.

Bühlmann, P. and Künsch, H.R. (1995). *The blockwise bootstrap for general parameters of a stationary time series.* To appear in Scand. J. Statist.

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

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

Götze, F. and Künsch, H.R. (1993). *Second order correctness of the blockwise bootstrap for stationary observations.* Preprint 93-061, Universität Bielefeld.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

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

Liu,R.Y., and Singh, K. (1992). *Moving blocks jackknife and bootstrap capture weak dependence*, in *Exploring the limits of the bootstrap* (Le Page, R., and Billard, L., eds.). Wiley, New York.

Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1994). *Validity of block-wise bootstrap for empirical processes with stationary observations.* To appear in Ann. Statist.

Politis, D.N. and Romano, J.P. (1992). *A general resampling scheme for triangular arrays of α-mixing random variables.* Ann. Statist. **20**, 1985-2007.

Politis, D.N. and Romano, J.P. (1993). *Nonparametric resampling for homogeneous strong mixing random fields.* J. Multiv. Anal. **47**, 301-328.

Politis, D.N. and Romano, J.P. (1994). *The stationary bootstrap.* J. Amer. Stat. Assoc., to appear.

Priestley, M.B. (1981). *Spectral Analysis and Time Series* **1**. Academic, New York.

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