Discussiones Mathematicae
Probability and Statistics 21 (2001) 121–147

ON CONSTRUCTION OF CONFIDENCE INTERVALS FOR A MEAN OF DEPENDENT DATA

JAN ĆWIK
Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland
e-mail: jc@ipipan.waw.pl

AND

JAN MIELNICZUK
Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland
Polish-Japanese Institute of Computer Technologies
Koszykowa 86, 02-008 Warszaw, Poland
e-mail: miel@ipipan.waw.pl

Abstract

In the report, the performance of several methods of constructing confidence intervals for a mean of stationary sequence is investigated using extensive simulation study. The studied approaches are sample reuse block methods which do not resort to bootstrap. It turns out that the performance of some known methods strongly depends on a model under consideration and on whether a two-sided or one-sided interval is used. Among the methods studied, the block method based on weak convergence result by Wu (2001) seems to perform most stably.

Keywords: confidence intervals; short-range dependence; reuse block methods; normal approximation; iterated random function sequence.

2000 Mathematics Subject Classification: Primary 62G15; Secondary 62G07.

1. INTRODUCTION

Let $\{X_i\}_{i=1}^{\infty}$ be a stationary time series with mean μ and autocovariance at lag i $r(i) = E(X_1 - \mu)(X_{1+i} - \mu)$. It is assumed throughout that $\{X_i\}_{i=1}^{\infty}$ is a short-range dependent (short-memory) sequence in the sense that $\sum_{i=1}^{\infty} |r(i)| < \infty$. Although studying the opposite case of the so called long-range dependence (long-memory) when autocovariances are not absolutely summable is becoming increasingly popular (see e.g. Beran (1994)), the short-range dependent case is likely to be the most common case of dependence in time series. In the paper, we study several methods of constructing confidence intervals and their performance for a variety of dependence structures. As the frame of reference we consider a normal approximation method for which tapered plug-in estimator of the variance of a sample mean is used. Performance of the normal approximation method is compared with three block methods. The first one is based on a weak convergence result by Wu (2001). The second is a percentile method pertaining to the approach of Hall and Jing (1996). The last studied method is a studentized version of the second approach which additionaly incorporates the Richardson correction. This is a slightly modified Hall and Jing's original proposal. The only change consists in using an alternative estimator of a variance of a sample mean which does not require assuming that the sequence is m-dependent for some known $m \in \mathbb{N}$. The main difference between the first and the two remaining methods is the use of overlapping, as in the case of Hall and Jing's approach, or non-overlapping blocks.

Performance of confidence intervals is studied by means of simulations for a variety of models including iterated random function sequences defined in Section 2. Previously, the study of performance of confidence intervals for dependent data was limited to few simple models, predominantly linear autoregressive, which do not reflect adequately potential difficulties which

can occur when dealing with estimation in more realistic situations. Here, by working with a greater variety of models, we were able to delineate more clearly some drawbacks of the methods used.

Resampling techniques are also popular for the construction of interval estimates. However, as Singh (1981) pointed out, ordinary bootstrap fails even for m-dependent sequences due to inconsistent estimation of the variance of $\sqrt{n}\bar{X}_n$, $\bar{X}_n$ being a sample mean. Indeed, the variance of a bootstrapped mean $\bar{X}_n^*$ is equal to $n^{-2}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, which multiplied by n converges to the variance $r(0)$ of a marginal distribution instead of to σ_∞^2 defined in (2). In order to overcome this problem and account for the dependence, usually blocks of observations from original data are randomly selected and then merged into a new series to form a bootstrap sample. After having repeated this procedure several times and having obtained an empirical distribution of the bootstrapped means its quantiles are used to construct confidence interval for μ. This general line of tackling the dependence problem is called block bootstrap. Again, non-overlapping (see e.g. Carlstein (1986)) or overlapping (Künsch (1989)) blocks are used for random selection. Shao and Tu (1995) give an overview of this approach in Chapter 9. This technique suffers from the drawback that although the original dependence structure is maintained within blocks, it is destroyed at their boundaries. This is the main reason behind the result of Davison and Hall (1992) which states that regardless what type of blocks is used, the distribution of block bootstrapped studentized mean is no more accurate than that obtained by normal approximation. In order to overcome this problem a much more careful variance estimation method has to be employed. That is why we focus here on conceptually simpler methods described in Section 3. The adopted approach was initiated by Hall and Jing (1996) where sample reuse methods were proposed to construct a confidence interval for mean without resorting to bootstrap.

2. PRELIMINARIES

In view of stationarity of $\{X_i\}_{i=1}^\infty$ it is easily seen that the variance of $\sqrt{n}\bar{X}_n$ is equal

$$(1)\quad \sigma_n^2 := n\mathrm{Var}(\bar{X}_n) = n^{-1} \sum_{1\le i,j\le n} \mathrm{Cov}(X_i, X_j) = r(0) + 2\sum_{i=1}^{n-1} \Big(1 - \frac{i}{n}\Big) r(i)$$

and thus for the short-range dependent $\{X_i\}_{i=1}^{\infty}$ we have as $n \to \infty$

$$(2)\qquad \sigma_n^2 \to \sigma_\infty^2 = r(0) + 2\sum_{i=1}^{\infty} r(i) < \infty.$$

Construction of confidence intervals for μ based on the sample mean $\bar{X}_n$ is intimately related to estimation of variance σ_n^2. Let us note that estimation of σ_n^2 is also relevant for other related inferential problems. For example, assesment of performance of $\hat{\theta} = h(\bar{X}_n)$ as an estimator of $\theta = h(\mu)$ for a differentiable function h is usually based on approximate equalities for the bias and the variance of $\hat{\theta}$: $\mathbb{E}(\hat{\theta}) - \theta \approx (1/2)h''(\mu)\mathrm{Var}(\bar{X}_n)$ and $\mathrm{Var}(\hat{\theta}) \approx h'(\mu)^2\mathrm{Var}(\bar{X}_n)$ which easily follow from Taylor expansion. In order to make them applicable we must first reliably estimate σ_n^2 or its limit σ_∞^2. Construction of confidence interval for μ has an obvious relevance in practice. Let us mention that constructing confidence intervals based on disjoint portions of data and checking whether they overlap is an effective method for detecting a possible change point of a mean function.

Suppose that we observe first n variables $X_1, X_2, \ldots, X_n$ of the sequence $\{X_i\}_{i=1}^{\infty}$. Under short-range dependence, a spectral density $f(\cdot)$ of $\{X_n\}_{n=1}^{\infty}$ exists and is given by (see e.g. Brockwell and Davis (1998), Theorem 4.3.2)

$$(3)\qquad f(s) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} r(k)e^{-iks} = (2\pi)^{-1} \left\{ r(0) + 2\sum_{k=1}^{\infty} r(k)e^{-iks} \right\}$$

for $-\pi \le s \le \pi$. Thus $\sigma_\infty^2 = 2\pi f(0)$ and estimation of limiting variance of $\sqrt{n}\bar{X}_n$ is equivalent to estimation of the spectral density at the origin. An intuitive way to achieve this is to plug sample autocovariances defined by

$$\hat{r}(i) = (n - |i|)^{-1} \sum_{j=1}^{n-|i|} (X_i - \bar{X}_n)(X_{j+|i|} - \bar{X}_n), \quad i \in \mathbb{Z},$$

in the definition of σ_∞^2 or equivalently in equation (3). It turns out, however, that precision of estimation of $r(i)$ decreases dramatically with increasing lag i. We feel that the usual recommendation which allows for estimation of $r(i)$ for $i \leq n/4$ whenever $n \geq 50$ (see e.g. Box and Jenkins (1976) p. 33) is too liberal, especially when mild nonstationarity is encountered as in the case of iterated random function sequences. We use a more restrictive condition $i \leq \sqrt{n}$ in definition of $\bar{\sigma}_n^2$ below.

We shall consider two (overlapping) classes of dependent sequences. The first one is that of two-sided linear processes for which

$$X_n = \mu + \sum_{j=-\infty}^{\infty} a_j \varepsilon_{n-j}, \quad i = 1, 2, \ldots \tag{4}$$

where $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ are i.i.d. innovations with mean 0 and finite variance and $\{a_j\}_{j=-\infty}^{\infty}$ are weights satisfying $\sum_{i=-\infty}^{\infty} |a_i| < \infty$. The last condition roughly corresponds to short-range dependence case. This is an important class of processes in view of Wold decomposition (c.f. Brockwell and Davis (1998), Theorem 5.7.1) stating that any purely non-deterministic stationary process satisfies decomposition (3) with $a_i = 0$ for $i < 0$ and *uncorrelated* innovations. In particular, if spectral density is continuous and positive on $[-\pi, \pi]$ the process is purely non-deterministic. The class includes ARMA and ARIMA procesess among others.

The second class of dependent sequences considered here is a class of iterated random functions given by

$$X_n = F(X_{n-1}, \theta_n), \quad n = 1, 2, \ldots$$

with $X_0 = x_0$, F is a given function and θ_n is an i.i.d. sequence with a common distribution H such that θ_n is independent of σ-field $\sigma(X_0, X_1, \ldots, X_{n-1})$ generated by $X_0, X_1, \ldots, X_{n-1}$ which pertains to the history of the process up to but not including time n. Mild conditions

under which iterated random function processes have a stationary distribution are known. A typical assumption here is $\int \log(L_\theta) H(d\theta) < 0$, where $L_\theta = \sup_{x \neq x'} |F(x,\theta) - F(x',\theta)|/|x - x'|$ which means that F is contracting on average. For an excellent review of known results see Diaconis and Freedman (1999). The models we consider will always satisfy conditions of asymptotic stability. In particular, this means in that after rejection of its initial part the process may be considered as approximately stationary.

3. CONFIDENCE INTERVALS

Let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution. Throughout all limits are considered with $n \to \infty$ unless stated otherwise. We describe now confidence intervals studied in the paper. We state the form of two-sided intervals only, their one-sided versions are defined analogously with an appropriate modification of an order of the pertaining quantile.

Method I

This is a commonly used method considered here as the frame of reference. Its justification follows from Theorem 1 below. The result yields asymptotically exact $1 - \alpha$ confidence level interval for μ of the form

$$\mathrm{CI}_1 = \Big(\bar{X}_n - \frac{\bar{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}_n + \frac{\bar{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}\Big), \tag{5}$$

where z_β is a standard normal quantile of order β and

$$\bar{\sigma}_n^2 = \sum_{|k| \leq m_n} w(k/m_n)\hat{r}(k),$$

where $w(\cdot)$ is a symmetric function supported on $[-1,1]$ with mode at 0 equal to $w(0) = 1$ and m_n is a sequence of integers tending to ∞ and such that $m_n/n \to 0$. Introduction of factors $w(k/m_n)$ in the definition of $\bar{\sigma}_n^2$ plays an obvious role of weighing down values of $\hat{r}(k)$ for large k when sample autocovariances might not be sufficiently precise estimators of their theorethical counterparts. In view of (3) it comes as no suprise

that $\bar{\sigma}_n^2 = 2\pi\hat{f}(0)$, where $\hat{f}(\cdot)$ is a smoothed periodogram corresponding to weight function $w(\cdot)$. In our study we used Bartlett kernel $w(x) = (1 - |x|)I\{|x| \leq 1\}$ since weight $w(k/m_n)$ yields analogous decay factor to $(1 - i/n)$ appearing in (1). Moreover, we choose $m_n = \sqrt{n}$. Thus for $n = 100, 300$ and 500 parameter m_n was equal to 10, 17 and 22, respectively. This corresponds reasonably well to the range of k for which $r(k)$ is sensibly estimated. CI_1 in (5) attains asymptotically confidence level $1 - \alpha$ due to the following result.

Theorem 1. *Assume that X_i is a linear process such that $\sum_{i=-\infty}^{\infty} |a_i| i^{1/2} < \infty$, $\sum_{i=-\infty}^{\infty} a_i \neq 0$ and $E\varepsilon_1^4 < \infty$, then*

$$\sqrt{n}(\bar{X}_n - \mu)/\bar{\sigma}_n \xrightarrow{\mathcal{D}} N(0,1).$$

The result follows from Theorems 7.1.2 and 10.4.1 in Brockwell and Davis (1998). Namely, $n^{1/2}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma_\infty^2)$ provided $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, where it is easily seen that $\sigma_\infty^2 = (\sum_{i=-\infty}^{\infty} a_i)^2 \neq 0$. Thus it is enough to check that $\bar{\sigma}_n^2$ is a weakly consistent estimate of σ_∞^2. This, however, follows from Theorem 10.4.1.

Method II

Let b be a chosen number such that n is divisible by b and $l = n/b$. Number b signifies a number of disjoint blocks into which the sample $\{X_1, X_2, \ldots, X_n\}$ is divided and l their common length. Consider i^{th} consecutive subsample of size l $\{X_{(i-1)b+j}, j = 1, \ldots, l\}$ and let $S_i = l^{-1} \sum_{j=1}^{l} X_{(i-1)b+j}$ be its sample mean. The second studied interval which asymptotically attains $1 - \alpha$ confidence level is defined as

$$\mathrm{CI}_2 = \left(\bar{X}_n - \frac{\hat{\sigma}_{n,l}}{\sqrt{n}} t_{1-\alpha/2}^{b-1}, \bar{X}_n + \frac{\hat{\sigma}_{n,l}}{\sqrt{n}} t_{1-\alpha/2}^{b-1}\right), \tag{6}$$

where t_β^k is quantile of order β of Student t-distribution with k degrees of freedom and $\hat{\sigma}_{n,l}^2 = \hat{\sigma}_n^2$ is an estimator of the variance of $\sqrt{n}\bar{X}_n$ defined as follows

$$\hat{\sigma}^2_{n,l} = \frac{n}{b(b-1)} \sum_{i=1}^{b} (S_i - \frac{1}{n} \sum_{i=1}^{b} S_i)^2 = \frac{l}{b-1} \sum_{i=1}^{b} (S_i - \bar{X}_n)^2. \tag{7}$$

Estimator $\hat{\sigma}^2_{n,l}$ is called grouped jacknife variance estimator in Shao and Tu (1995). Note that $\hat{\sigma}^2_n$ is equal to l times a sample variance of $\{S_1, S_2, \ldots, S_b\}$. Its idea may be heuristically explained in the following way. An appeal to law of large numbers suggests that

$$\frac{1}{b-1} \sum_{i=1}^{b} (S_i - \bar{X}_n)^2 \sim \sigma^2_{S_1} = \frac{1}{l} \Big\{ r(0) + 2 \sum_{j=1}^{l-1} \Big(1 - \frac{j}{l}\Big) r(j) \Big\},$$

where $\sigma^2_{S_1}$ denotes the variance of block mean S_1. Thus $\hat{\sigma}^2_n \sim r(0) + 2\sum_{j=1}^{l-1}(1 - j/l)r(j) \sim \sigma^2_n \sim \sigma^2_\infty$ for large l and n.

Intuitively, increasing block length l should result in decreasing bias of $\hat{\sigma}^2_{n,l}$ as an estimator of σ^2_∞, as subsamples $\{X_{(i-1)b+j}, j = 1, \ldots, l\}$ resemble $\{X_1, X_2, \ldots, X_n\}$ more closely for large l. On the other hand, decreasing block length results in a larger number of blocks and smaller variance of $\hat{\sigma}^2_{n,l}$. Thus we have the usual trade-off between magnitudes of the variance and the squared bias which sum up to the mean squared error (MSE) of $\hat{\sigma}^2_{n,l}$. Formula for a block length which minimizes the asymptotic MSE is unknown for a general process $\{X_i\}$. For an autoregressive process of order 1 defined as $X_t = \phi X_{t-1} + \varepsilon_t$, Carlstein (1986) found that the optimal block length is equal to

$$l_{opt} = \Big(\frac{2\phi}{(1-\phi)(1+\phi)} \Big)^{2/3} n^{1/3} \tag{8}$$

and thus the optimal number of blocks $b_{opt} \sim n^{2/3}$.

Below we state the weak convergence result due to Wu (2001) which justifies validity of $1-\alpha$ level confidence interval in (6). It is supplemented by results on the asymptotic behaviour of bias and variance of $\hat{\sigma}_n^2$. The fact that $\mathrm{Var}\hat{\sigma}_n^2$ behaves asymptotically as an inverse of the sample size is proved for process $\{X_i\}$ such that cumulant $\mathrm{cum}(X_i, X_j, X_k, X_l) = 0$ for any $i, j, k, l \in \mathbb{N}$ .This is true in particular for a Gaussian stationary process. We conjecture that the result is valid in more general situation under summability conditions of the cumulants of order four. Similar results are available in the literature for m-dependent processes (see e.g. Shao and Tu (1995), p. 389).

Let $T_i = \sum_{j=1}^{i} X_j$ and $W(\cdot)$ is a Wiener process on [0,1].

Theorem 2. (a) *Assume that*

$$\frac{T_{[nt]} - [nt]\mu}{\sqrt{n}\sigma_\infty} \xrightarrow{\mathcal{D}} W(t)$$

with $\xrightarrow{\mathcal{D}}$ *denoting convergence in* $\mathcal{D}[0,1]$ *for some* $\sigma > 0$. *Then for fixed* $b \in \mathbb{N}$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} \xrightarrow{\mathcal{D}} t_{b-1}.$$

(b)

$$\mathbb{E}\hat{\sigma}_n^2 = \sigma_n^2 + \frac{2b}{b-1}\left\{\sum_{j=1}^{l}\left(\frac{j}{n} - \frac{j}{l}\right) r(j) - \sum_{j=l+1}^{n-1}\left(1 - \frac{j}{n}\right) r(j)\right\}.$$

(c) *Assume that* $\{X_i\}$ *is short-range dependent, its cumulants of order* 4 *are* 0 *and* $\min(b, l) \to \infty$. *Then*

$$\mathrm{Var}\hat{\sigma}_n^2 = 2(\sigma_\infty^2)^2/b + o(b^{-1}).$$

The result is proved in the appendix. Note that from (b) it follows that if $r(i) \geq 0$ for all i then $\mathbb{E}\hat{\sigma}_{n,l}^2 \leq \sigma_n^2$ and the opposite inequality holds if $r(i) \leq 0$.

Corollary. *If $\{X_i\}$ is short-range dependent and $l \to \infty$ then $\hat{\sigma}^2_{n,l}$ is asymptotically unbiased estimator of σ^2_∞. Moreover, if cumulants of order 4 are 0 and $b = n/l \to \infty$ then $\hat{\sigma}^2_{n,l}$ is* MSE*-consistent.*

Remark. In view of Ho and Hsing (1997) Theorem 2(a) is applicable for any one-sided linear process such that $\sum_{i=0}^{\infty} |a_i| < \infty$ and $\sum_{i=0}^{\infty} a_i \neq 0$. Moreover, in view of Theorem 1 in Csörgő and Mielniczuk (1995) the result is valid if $\{X_i\}$ is obtained as an instanteonous transformation of Gaussian stationary sequence $\{Z_i\}$, that is $X_i = G(Z_i)$ provided $\{X_i\}_{i=1}^{\infty}$ is short-range dependent and $EX_i^4 < \infty$.

Choice of b. In view of Carlstein's result (8) on MSE-optimal block length for autoregressive AR(1) model, it may be conjectured that in general optimal block length $b_{opt}(n) = Cn^{2/3}$ where C is some constant pertaining to dependence structure of underlying process. Thus, under validity of the conjecture,

$$b_{opt}(n) = b_{opt}(100)(n/100)^{2/3}.$$

Therefore, one can estimate $b_{opt}(n)$ by plugging $\hat{b}_{opt}(100)$ into the last formula. In order to estimate $b_{opt}(100)$ the following recalibration method was used. We choose $n \gg 100$, $n = 1000$, say, and apply formula (7) to get an estimate of $\hat{\sigma}^2_{1000,100} := \hat{\sigma}^2$ based on 10 blocks of length 100. Then for every $i = 1, 2, \ldots, 10$ we estimate σ^2_{100} using i^{th} block only. This pertains to dividing i^{th} block into k smaller blocks and applying (7) again. Let call the estimate $\hat{\sigma}^2_{100,k}$ based on i^{th} block by $\hat{\sigma}^2_i(k)$. The estimated mean squared error of $\hat{\sigma}^2_{100,k}$ is defined as

$$\widehat{\mathrm{MSE}}(k) = \frac{1}{10}\sum_{i=1}^{k}(\hat{\sigma}^2 - \sigma^2_i(k))^2.$$

We choose as $\hat{b}_{opt}(100)$ the value of k which minimizes the above expression. In the simulation study below we compare $\hat{b}_{opt}(100)$ with b yielding the smallest discrepancy of empirical coverage from an assumed coverage level.

Method III

For a given block length $l \leq n$ and $i = 1, 2, \ldots, n-l+1$, consider $n-l+1$ overlapping blocks $\{X_i, X_{i+1}, \ldots, X_{i+l-1}\}$. Let S_i be a sample mean of i^{th} block and define $Y_i = \sqrt{l}(S_i - \bar{X}_n)$. We define the following percentile confidence interval

$$\mathrm{CI}_3 = \left(\bar{X}_n - \frac{q_{1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{q_{\alpha/2}}{\sqrt{n}}\right),$$

where q_β is an empirical quantile of order β from the sample Y_i, $i = 1, 2, \ldots, n-l+1$. The following argument is helpful to understand the role played by the empirical distribution of $\{Y_i\}$. Denote by $\mu_n(P)$ the distribution of $\sqrt{n}(\bar{X}_n - \mu)$ and suppose that $\mu_n(P)$ converges in distribution to some probability measure ν. Then for each $i = 1, 2, \ldots, n-l+1$, $Y_i^* = \sqrt{l}(S_i - \mu)$ is distributed as $\mu_l(P)$ which for large l should well approximate $\mu_n(P)$ in view of weak convergence of the last sequence of measures. Moreover, as $\sqrt{l}(\bar{X}_n - \mu) \to 0$ when $l/n \to 0$ it is permissible to replace μ with $\bar{X}_n$ in the definition of Y_i^*. The reasoning justifies the definition of Y_l given above and was used by Politis and Romano (1994) to motivate a resample method in which all $\binom{n}{l}$ subsamples of size l are used to estimate the distribution $\mu_n(P)$. They proved that under some conditions on mixing coefficients of sequence $\{X_i\}$ the empirical distribution of $\{Y_i\}$ has the same distributional limit as $\mu_n(P)$. Here, instead of an estimate based on all subsamples which is necessarily computationally expensive we concentrate on its counterpart based on $n-l+1$ overlapping blocks. Observe that the empirical variance $\tilde{\sigma}_n^2$ of $\{Y_1, Y_2, \ldots, Y_{n-l+1}\}$ is equal to

$$\frac{1}{n-l}\sum_{i=1}^{n-l+1}\left(\sqrt{l}(S_i - \bar{X}_n) - \frac{1}{n-l+1}\sum_{j=1}^{n-l+1}\sqrt{l}(S_j - \bar{X}_n)\right)^2 =$$

$$= \frac{l}{n-l}\sum_{i=1}^{n-l+1}\left(S_i - \frac{1}{n-l+1}\sum_{j=1}^{n-l+1} S_j\right)^2$$

and is an analogue of $\hat{\sigma}_n^2$ in (7) for overlapping samples. The heuristic reasoning given below (7) to jusify $\hat{\sigma}_n^2$ as an estimator of σ_∞^2 can also be used in case of $\tilde{\sigma}_n^2$. Estimate $\tilde{\sigma}_n^2$ with the normalizing factor $l/(n-l+1)$ instead of $l/(n-l)$ was considered by Künsch (1989). Obviously, there is a great deal of redundancy involved in the definition of $\tilde{\sigma}_n^2$ as close subsamples do not contribute much information about sampling variability.

Method IV

The last method is a studentized analogue of the previous method which additionally incorporates Richardson correction. It consists in considering studentized versions of Y_i defined as $\tilde{Y}_i^* = \sqrt{l}(S_i - \bar{X}_n)/\tilde{\sigma}_n$ where $\tilde{\sigma}_n^2$ is Künsch's estimator of the variance σ_n^2. Morever, an empirical distribution function of $\tilde{Y}_i^*$ is introduced

$$\tilde{F}_1(x) = \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} I\{\tilde{Y}_i^* \leq x\}.$$

Instead of directly using quantiles of $\tilde{F}_1$ in order to construct confidence interval for μ, Hall and Jing (1996) introduced distribution estimate $\hat{F}$ incorporating Richardson extrapolation. Namely, $\hat{F}$ is defined as a convex combination of $\tilde{F}$ and a cumulative distribution function of standard normal distribution Φ

$$\hat{F}(x) = \left(\frac{l}{n}\right)^{1/2} \Phi(x) + \left(1 - (\frac{l}{n})^{1/2}\right) \tilde{F}_1(x).$$

Then

$$\mathrm{CI}_4 = \left(\bar{X}_n - \frac{\tilde{\sigma}_n \hat{q}_{1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{\tilde{\sigma}_n q_{\alpha/2}}{\sqrt{n}}\right),$$

where $\hat{q}_\beta$ is quantile of order β pertaining to $\hat{F}$. Richardson extrapolation is introduced in order to cancel out the first order terms in $\tilde{F} - F$, where

$F(x) = P((S_i - \bar{X}_n)/\tilde{\sigma}_n \leq x)$. For details see Hall and Jing (1996). The only difference between the studied method and original Hall and Jing's method is the use of $\bar{\sigma}_n^2$ instead of σ_n^{*2} for standardization. The definition of σ_n^{*2} is based on the assumption that the sequence $\{X_i\}$ is m-dependent for known m. This method works well for moving averages of known order considered as examples in Hall and Jing but our simulations indicate that its performance degrades dramatically for more general time series such as iterated random functions.

4. Simulation study

The following iterated random function sequences studied in "Cwik *et al.* (2000) in a different context were considered
Model A. (Linear autoregressive, AR(1))

$$X_t = 0.8X_{t-1} + \epsilon_t.$$

Model B. (Exponential autoregressive)

$$X_t = (0.8 - 1.1\exp\{-30X_{t-1}^2\})X_{t-1} + \epsilon_t.$$

Model C. (Threshold autoregressive)

$$X_t = 0.8X_{t-1}I\{X_{t-1} > 0\} - 0.3X_{t-1}I\{X_{t-1} \leq 0\} + \epsilon_t.$$

Model C–.

$$X_t = -0.8X_{t-1}I\{X_{t-1} > 0\} + 0.3X_{t-1}I\{X_{t-1} \leq 0\} + \epsilon_t.$$

Model D. (Random coefficient autoregressive)

$$X_t = (1 + \tilde{\epsilon_t})(0.8X_{t-1}I\{X_{t-1} > 0\} - 0.3X_{t-1}I\{X_{t-1} \leq 0\}) + \epsilon_t$$

with $\tilde{\epsilon_t}$ independent of ϵ_t and $\sigma(X_0, X_1, \ldots, X_{t-1})$, and assuming value 0.25 with probability 1/2 and value -0.25 with probability 1/2.

Model E. (Discontinuous)

$$X_t = \frac{2\text{sgn}X_{t-1}}{3 + |X_{t-1}|} + \epsilon_t$$

where $\epsilon_t \sim N(0, 0.5^2)$.

In all the above models except in model E, we assume that $\epsilon_t \sim N(0, 0.01)$. Moreover, moving average models M1 and M2 with innovations distributed as chi square distribution with 1 degree of freedom were considered. More specifically,

Model Mi.

$$X_t = (\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-i} - (i+1))/\sqrt{i+1},$$

where independent $\varepsilon_t \sim \chi^2_1$. Models M$i$ $(i = 1, 2, 3)$ were considered in the simulation study of Hall and Jing (1996).

The means in case A, M1 and M2 are clearly 0. The mean μ of the stationary distribution in case of B, C, C–, D, E was empirically determined as an average of 5 sample means of sample paths consisting of 100000 observations generated from the respective model after discarding initial 1000 observations. All the values obtained are as follows

$$\mu_A = \mu_B = 0, \quad \mu_C = 0.091, \quad \mu_{C-} = -0.039, \quad \mu_D = 0.095,$$

$$\mu_E = \mu_{M1} = \mu_{M2} = 0.$$

Observe that model D strictly speaking does not belong to the class of iterated random function processes as defined in Section 2. Hovewer, it belongs to a generalized class of these processes when univariate innovations ε_t are replaced by bivariate ones, namely $(\varepsilon_t, \tilde{\varepsilon}_t)$. It can be also viewed as a special case of the hidden Markov chain model. Note that the conditional mean function $\mathbb{E}(X_t | X_{t-1} = x)$ is in model D the same as in model C. The conditional mean of model B differs from that in model A in nonlinear factor $-1.1\exp(-30x^2)$ which is significant for small x only.

All the simulations were performed using R statistical software. Their results are given below. They are based on $k = 3000$ repetitions and the sample sizes are equal to $n = 100, 300$ and 500. Here we were mostly concerned with the coverage accuracy understood as a discrepancy between the actual coverage probability and the assumed confidence level $1 - \alpha = 0.95$. Coverage probabilities listed in Tables 1 and 2 correspond to block lengths and numbers of blocks yielding the highest coverage accuracy. Results for two-sided and both one-sided intervals are given, since as it turns out their performance may crucially differ.

As Method IV performs similarly to Method III (but yielding slightly shorter confidence intervals) we focus here on the first three methods. Let us however point out that Method IV performs notably worse than its analogue proposed by Hall and Jing for which knowledge of the order of the model under consideration is assumed (see Table 4 in Hall and Jing (1996)).

Table 1. Coverage probabilities for $n = 100$.

		Model							
Interval	Method	A	B	C	C–	M1	M2	D	E
	I	.8092	.9132	.8450	.9192	.8944	.8896	.8406	.8532
two-	II	.9382	.9556	.9358	.9700	.9434	.9428	.9248	.9346
sided	III	.7672	.9230	.8020	.9040	.8400	.8372	.7886	.7474
	IV	.7562	.8952	.7964	.8884	.8416	.8302	.7846	.7676
	I	.8636	.9238	.8552	.9378	.8840	.8844	.8302	.8878
lower	II	.9374	.9522	.9126	.9704	.9268	.9268	.8932	.9388
	III	.8650	.9378	.8226	.9400	.8462	.8494	.7894	.8558
	IV	.8600	.9250	.8258	.9352	.8506	.8472	.7972	.8622
	I	.8684	.9212	.9326	.9274	.9518	.9440	.9422	.8978
upper	II	.9400	.9456	.9698	.9574	.9624	.9610	.9512	.9426
	III	.8684	.9384	.9402	.9168	.9560	.9518	.9544	.8602
	IV	.8596	.9164	.9280	.9080	.9474	.9426	.9444	.8630

Table 2. Coverage probabilities for $n = 300$.

		Model							
Interval	Method	A	B	C	C–	M1	M2	D	E
	I	.8812	.9250	.8940	.9344	.9186	.9274	.9018	.9052
one-	II	.9384	.9502	.9363	.9496	.9448	.9518	.9374	.9444
sided	III	.8710	.9482	.8610	.9318	.8884	.8994	.8640	.8508
	IV	.8636	.9294	.8607	.9222	.8830	.8974	.8632	.8556
	I	.9066	.9284	.8927	.9444	.9150	.9196	.8880	.9212
lower	II	.9398	.9478	.9237	.9542	.9346	.9380	.9120	.9468
	III	.9102	.9418	.8673	.9516	.8874	.8962	.8548	.9054
	IV	.9064	.9342	.8723	.9492	.8910	.8968	.8612	.9044
	I	.9050	.9308	.9427	.9368	.9484	.9542	.9476	.9210
upper	II	.9414	.9494	.9577	.9506	.9586	.9628	.9532	.9452
	III	.9092	.9456	.9510	.9314	.9466	.9554	.9628	.9082
	IV	.9040	.9346	.9447	.9274	.9484	.9494	.9566	.9056

Comments. For two-sided intervals and $n = 100$ Method II performs overall much better than Methods I and III. Moreover, this is true not only for the number of blocks yielding the highest coverage accuracy but also for several values even fairly distant from it (see Figures 1–6). The distinctive edge of Method II over the other methods becomes less significant for $n = 300$ and all methods perform almost identically for $n = 500$. For $n = 100$ Method I is outperformed in most cases by all block methods when block lengths are suitably chosen. Most remarkable feature of Method II is that discrepancies between coverage probabilities of one-sided and two-sided confidence intervals are much smaller than for other methods. As to performance of one-sided intervals it may well happen that these constructed by Method III work better than those by Method II.

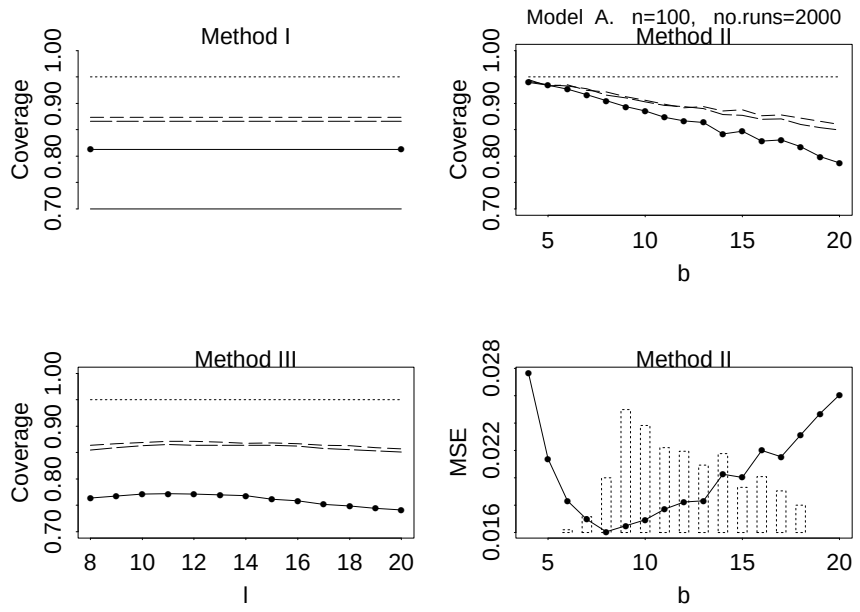

Figure 1. Coverage probability for model A, $n = 100$. Continuous line corresponds to a two-sided interval and long-dashed and short-dashed one to an upper and lower interval, respectively. Dotted line pertains to the target value 0.95.

However, for the models under consideration this is true only for one type of it (i.e. lower or upper) with Method II prevailing in the case of the other type. Moreover, the type of interval which performs better depends strongly on the model considered: it is *upper* confidence interval constructed by Method III in the case of model M1 and M2, but *lower* confidence interval for model C–. Thus it seems impossible to make use of the better performance without knowing the model we consider. Consequently, it might be advisable to consider two-sided confidence intervals for this type of data when high coverage accuracy is needed. Quite surprisingly it turns out that the presence of an additional noise in model D when compared to model C does not necessarily lead to a smaller coverage accuracy in as the case of upper confidence intervals for $n = 100$ indicates.

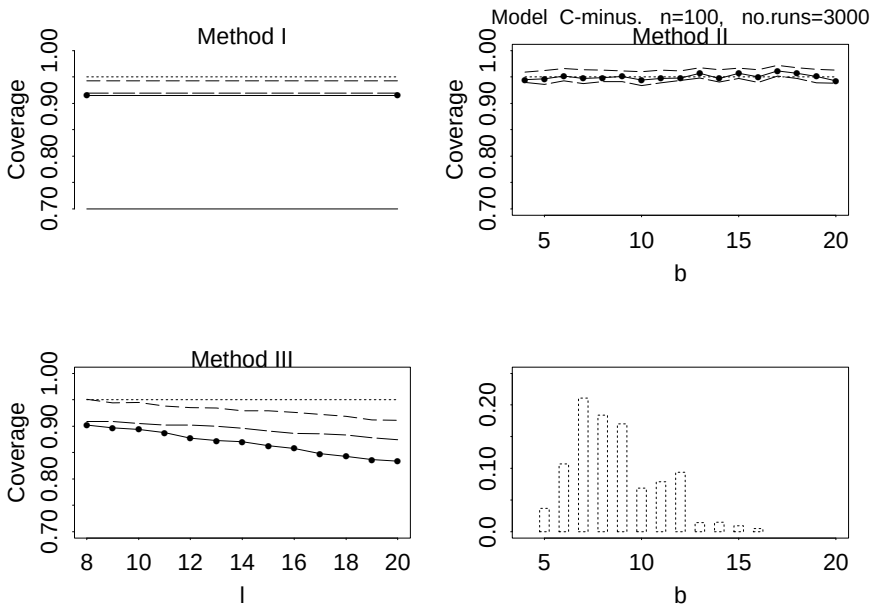

Figure 2. Coverage probability for model C–, $n = 100$.

The performance of two-sided interval constructed by Method III or IV is disastrous for discontinuous model E and, astoundingly, for linear autoregression model A.

Figure 7 shows the price paid for the better performance of two-sided confidence intervals constructed by Method II. Namely, they are in general wider than their counterparts constructed by other methods. Note however that this phenomenon is not associated with the coverage probability consistently higher than 0.95, i.e., Method II confidence intervals are not conservative. The moderate increase of numbers of blocks in Method II makes the average length of confidence intervals decrease without adversely affecting coverage accuracy of the method.

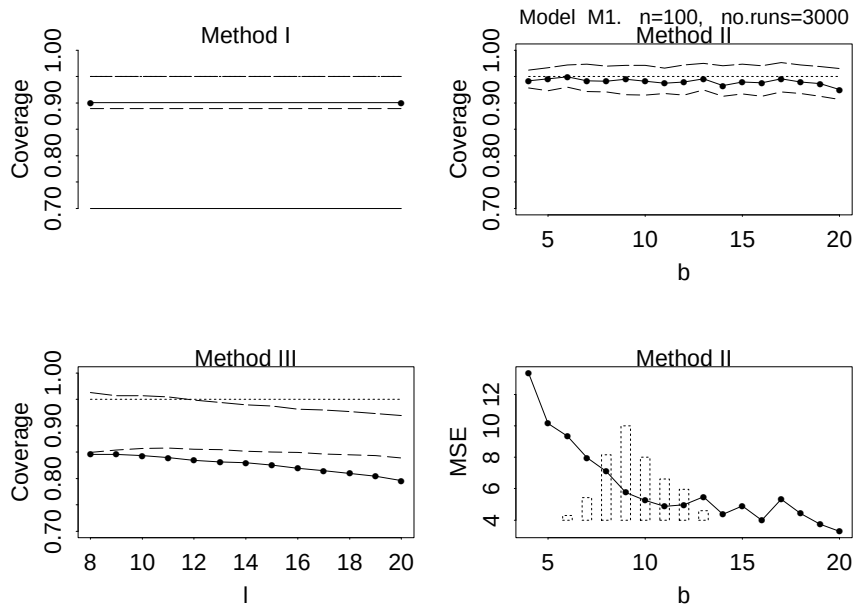

Figure 3. Coverage probability for model M1, $n = 100$.

Additionally, histograms of $\hat{b}_{opt}$ for Method II are shown in Figures 1-6. In the case when the value of σ_∞^2 is known (models A, M1 and M2) the plot of empirical MSE($\hat{\sigma}_n^2$) is superimposed on them. Histograms of $\hat{b}_{opt}$ are usually unimodal with a pronounced mode location which corresponds fairly well to the location of minimum of empirical MSE in the case of model A. This is obviously a desirable feature. However, $\hat{b}_{opt}$ *does not* correspond to the block length yielding highest coverage accuracy. This was also the case for optimal block length $\hat{l}_{opt}$ for Method III chosen by a similar recalibration method. That is why in Tables 1 and 2 coverage probabilities are given for optimal parameters chosen by examining the performance of confidence intervals over wide range of l and b.

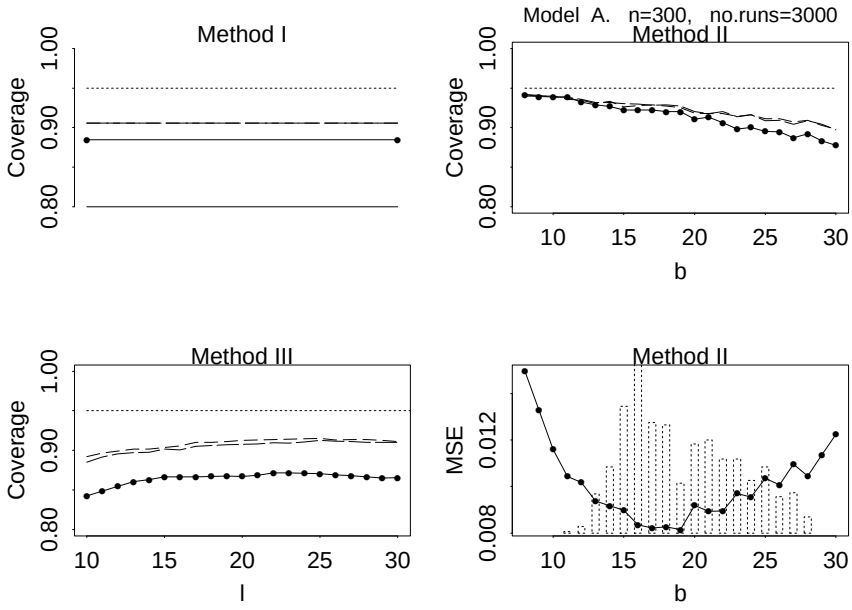

Figure 4. Coverage probability for model A, $n = 300$.

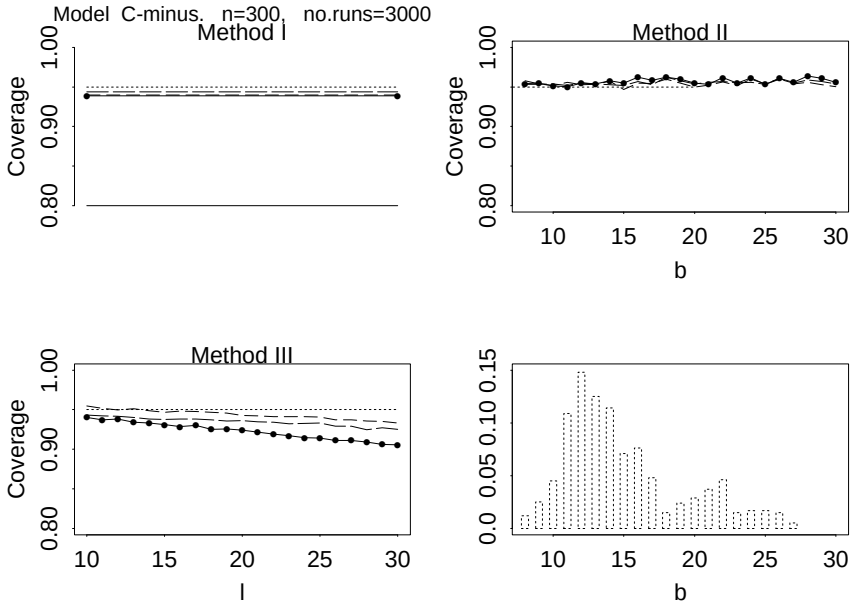

Figure 5. Coverage probability for model C–, $n = 300$.

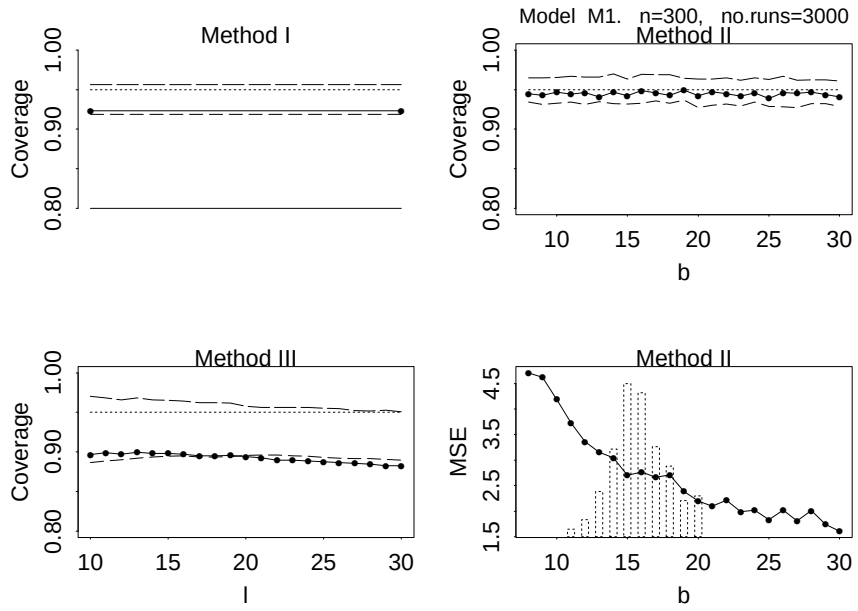

Figure 6. Coverage probability for model M1, $n = 300$.

Conclusions. The simulation study indicates that coverage inaccurracy of commonly used confidence intervals may be considerable, especially when processes other than finite order moving averages are considered and the order of dependence is not assumed to be known. This should be a matter of concern, especially so when one-sided confidence intervals are used. Two-sided confidence intervals based on limit result of Wu (2001) turn out to perform most stably in comparison with the other investigated methods. Moreover, it turns out that this method is fairly insensitive to the number of blocks used.

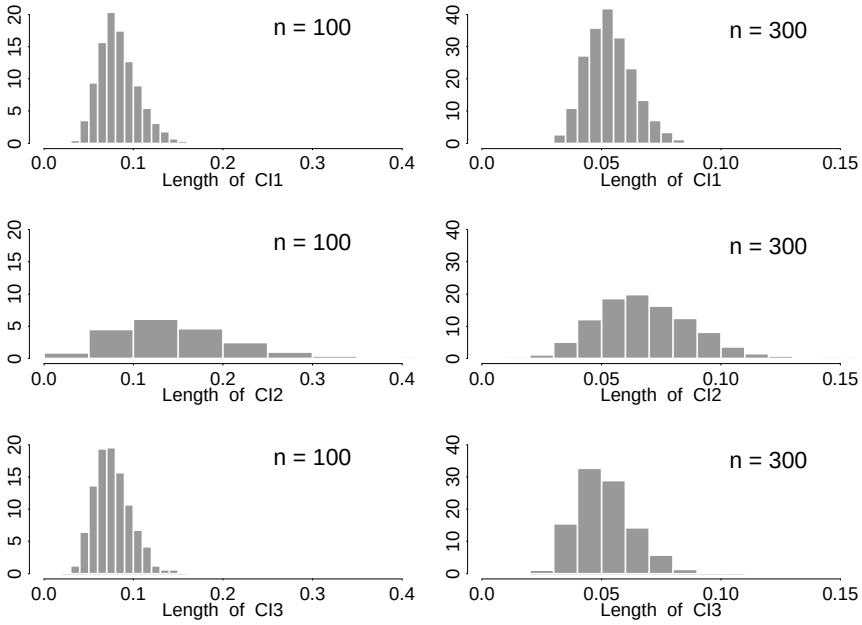

Figure 7. Histograms of lengths of confidence interval for model C, Methods I, II and III, $n = 100$ and $n = 300$. Optimal values of b and l were used.

Appendix

***Proof of Theorem* 2.** (a) For the sake of completeness we give a short proof of (a). From assumed weak convergence it follows that

$$\frac{\sqrt{n}}{\sigma_\infty}\left(\bar{X}_n-\mu, S_i-\mu, 1\le i\le b\right) \xrightarrow{\mathcal{D}} \left(W(1), b\left(W\left(\frac{i}{b}\right)-W\left(\frac{i-1}{b}\right)\right), 1\le i\le b\right).$$

Denote the limit vector by $(Z, Z_1, \ldots, Z_b)$. As $Z_1, Z_2, \ldots, Z_b$ are independent and identically distributed with normal distribution and $Z = b^{-1}\sum_{i=1}^{b} Z_i$, we have

$$\frac{\sqrt{b}Z}{\{(b-1)^{-1}\sum_{i=}^{b}(Z_i-\bar{Z})^2\}^{1/2}} \sim t_{b-1}.$$

Thus continuous mapping theorem implies

$$\frac{\sqrt{b}(\bar{X}_n-\mu)}{\{(b-1)^{-1}\sum_{i=1}^{b}(S_i-\mu-(\bar{X}_n-\mu))\}^{1/2}} = \frac{\sqrt{n}(\bar{X}_n-\mu)}{\hat{\sigma}_n} \xrightarrow{\mathcal{D}} t_{b-1}.$$

(b) Obviously

$$\frac{b(b-1)}{n}\hat{\sigma}_n^2 = \sum_{i=1}^{b}(S_i-\mu)^2 - b(\bar{X}_n-\mu)^2.$$

Thus

$$\mathbb{E}\frac{b(b-1)}{n}\hat{\sigma}_n^2 = b\mathbb{E}(S_1-\mu)^2 - b\mathbb{E}(\bar{X}_n-\mu)^2 =$$

$$= b\left(\frac{1}{l}-\frac{1}{n}\right)\left\{r(0)+2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)r(j)\right\}$$

$$+\frac{2b}{l}\left(\sum_{j=1}^{l}\left\{\left(1-\frac{j}{l}\right)-\left(1-\frac{j}{n}\right)\right\}r(j)-\sum_{j=l+1}^{n-1}\left(1-\frac{j}{n}\right)r(j)\right).$$

As $b(l^{-1}-n^{-1}) = b(b-1)/n$ and $2n/(b-1)l = 2b/(b-1)$, the result follows.

(c) Observe that $\mathrm{Var}(\sum_{j=1}^{b}(S_j - \bar{X}_n)^2)$ equals to

$$\mathrm{Var}\left(\sum_{j=1}^{b}(S_j-\mu)^2\right)+\mathrm{Var}(b(\bar{X}_n-\mu)^2)-2b\mathrm{Cov}\left(\sum_{j=1}^{b}(S_j-\mu)^2,(\bar{X}_n-\mu)^2\right). \tag{9}$$

We will show that the first term yields the main contribution to the variance and the two remaining ones are negligible. As cumulants of order 4 of the sequence $\{X_i\}$ are 0, Theorem 2 p. 34 in Rosenblatt (1985) implies that

$$\begin{aligned}
&\mathrm{Cov}\left((S_1-\mu)^2,(S_j-\mu)^2\right)\\
&=2\left(\mathrm{Cov}\left(S_1-\mu,S_j-\mu\right)\right)^2+\mathrm{cum}\left(S_1-\mu,S_1-\mu,S_j-\mu,S_j-\mu\right)\\
&=2\left(\mathrm{Cov}\left(S_1,S_j\right)\right)^2.
\end{aligned}$$

Thus

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{j=1}^{b}(S_j-\mu)^2\right)\\
&=b\mathrm{Var}(S_1-\mu)^2+2\sum_{i=1}^{b-1}(b-i)\mathrm{Cov}((S_1-\mu)^2,(S_{i+1}-\mu)^2) \qquad (10)\\
&=2b(\mathrm{Var}(S_1))^2+4\sum_{i=1}^{b-1}(b-i)(\mathrm{Cov}(S_1,S_{1+i}))^2.
\end{aligned}$$

It is easy to see that as $l \to \infty$

$$2b(\mathrm{Var}(S_1))^2 = \frac{2b}{l^2}(\sigma_\infty^2)^2 + o\left(\frac{b}{l^2}\right).$$

Moreover, note that

$$|\mathrm{Cov}(S_1, S_2)| \leq \frac{2}{l^2}\left\{\sum_{j=1}^{l-1} j|r(j)| + l\sum_{j=l}^{2l-1}|r(j)|\right\}$$

and

$$|\mathrm{Cov}(S_1, S_{i+1})| \leq \frac{1}{l^2}\left\{\sum_{j=il+1}^{(i+2)l-1} l|r(j)|\right\}$$

for $i \geq 2$. Thus the second term on the right hand of (10) may be bounded by

$$\frac{16b}{l^4}\left(\sum_{j=1}^{l-1} j|r(j)| + l\sum_{j=l}^{\infty}|r(j)|\right)^2 = o\left(\frac{b}{l^2}\right)$$

as $\sum_{j=1}^{l-1}\frac{j}{l}|r(j)| \to 0$ as $l \to \infty$ provided $\sum_{i=1}^{\infty}|r(i)| < \infty$. Using the same argument we show that the second term in (10) is $O(b^2/n^2) = o(b/l^2)$ as $b \to \infty$. Thus in view of Cauchy inequality the third term is also $o(b/l^2)$ and the result follows.

REFERENCES

[1] J. Beran, *Statistics for Long-Memory Processes*, Chapman and Hall, New York 1994.

[2] G. Box and G. Jenkins, *Time Series Analysis*, Holden Day 1976.

[3] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, 6th edition, Springer 1998.

[4] J. Ćwik, J. Koronacki and J. Mielniczuk, *Testing for a difference between conditional variance functions of nonlinear time series*, Control and Cybernetics **29** (2000), 33–50.

[5] E. Carlstein, *The use of subseries methods for estimating the variance of a general statistics from a stationary time series*, Ann. Statist **14** (1986), 1171–1179.

[6] S. Csörgő, and J. Mielniczuk, *Close short-range dependent sums and regression estimation*, Acta. Sci. Math. (Szeged) **60** (1995), 177–196.

[7] A. Davison and P. Hall, *On studentizing and blocking methods for implementing the bootstrap with dependent data*, Austr. J. Statist. **35** (1992), 215–224.

[8] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Review **41** (1999), 41–76.

[9] P. Hall and B. Jing, *On sample reuse methods for dependent data*, J. R. Statist. Soc. B (1996), 727–737.

[10] H.C. Ho and T. Hsing, *Limit theorems for functionals of moving averages*, Ann. Probab. **25** (1997), 1636–1669.

[11] H. Künsch, *The jacknife and the bootstrap for general stationary observations*, Ann. Statist. **17** (1989), 1217–1241.

[12] Politis and Romano, *Large sample confidence regions based on subsamples under minimal asuumptions*, Ann. Statist. **22** (1994), 2031–2050.

[13] M. Rosenblatt, *Stationary Sequences and Random Fields*, Birkhäuser, Boston 1985.

[14] J. Shao and D. Tu, *The Jacknife and Bootstrap*, Springer 1995.

[15] K. Singh, *On the asymptotic accuracy of Efron's bootstrap*, Ann. Statist. 9 (1981), 1187–1195.

[16] W.B. Wu, *Studies in time series and random dynamics*, Ph. D. thesis, University of Michigan 2001.

Received 9 January 2001