ECONOMETRIC REVIEWS
Vol. 23, No. 1, pp. 53–70, 2004

Automatic Block-Length Selection for the Dependent Bootstrap

Dimitris N. Politis[1,*] and Halbert White[2]

[1]Department of Mathematics and [2]Department of Economics,
University of California, San Diego, La Jolla,
California, USA

ABSTRACT

We review the different block bootstrap methods for time series, and present them in a unified framework. We then revisit a recent result of Lahiri [Lahiri, S. N. (1999b). Theoretical comparisons of block bootstrap methods, *Ann. Statist.* 27:386–404] comparing the different methods and give a corrected bound on their asymptotic relative efficiency; we also introduce a new notion of finite-sample "attainable" relative efficiency. Finally, based on the notion of spectral estimation via the flat-top lag-windows of Politis and Romano [Politis, D. N., Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* 16:67–103], we propose practically useful estimators of the optimal block size for the aforementioned block bootstrap methods. Our estimators are characterized by the fastest possible rate of convergence which is adaptive on the strength of the correlation of the time series as measured by the correlogram.

Key Words: Bandwidth choice; Block bootstrap; Resampling; Subsampling; Time series; Variance estimation.

*Correspondence: Dimitris N. Politis, Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA; E-mail: dpolitis@ucsd.edu.

DOI: 10.1081/ETC-120028836 0747-4938 (Print); 1532-4168 (Online)

1. INTRODUCTION

Implementation of block bootstrap methods for dependent data typically requires selection of b, a block length or an expected block length; cf. Künsch (1989), Liu and Singh (1992), Politis and Romano (1992), Politis and Romano (1994)—see also the related work of Carlstein (1986). Apart from specifying the rate at which b must grow with the sample size, N, available results typically offer little guidance on how to choose b. Exceptions are the results of Hall et al. (1995) and Bühlmann and Künsch (1999) who provide data-dependent methods for selecting b for the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992); see also the review by Berkowitz and Kilian (2000).

In this note we review some different ways of implementing the block bootstrap for time series, and present them in a unified framework. We give a comparison between the moving blocks bootstrap and the stationary bootstrap, thus rectifying an incorrect claim by Lahiri (1999b). In addition, we provide a novel methodology of automatic selection/estimation of optimal block sizes; the methodology is based on the notion of spectral estimation via the flat-top lag-windows of Politis and Romano (1995) that possess many favorable properties. Finally, we present some illustrative simulations and introduce a new notion of finite-sample "attainable" relative efficiency for comparing different block bootstrap methods.

2. BASIC FRAMEWORK

Suppose $X_1, \ldots, X_N$ are observations from the (strictly) stationary real-valued sequence $\{X_n, n \in \mathbf{Z}\}$ having mean $\mu = EX_t$, and autocovariance sequence $R(s) = E(X_t - \mu)(X_{t+|s|} - \mu)$. Both μ and $R(\cdot)$ are unknown, and the objective is to obtain an approximation to the sampling distribution of $\overline{X}_N = N^{-1}\sum_{t=1}^{N} X_t$. Since typically $\overline{X}_N$ is asymptotically normal, estimating the variance $\sigma_N^2 = Var(\sqrt{N}\,\overline{X}_N) = R(0) + 2\sum_{s=1}^{N}(1 - s/N)R(s)$ is important.

Sufficient conditions for the validity of a central limit theorem for $\overline{X}_N$ are given by a moment condition and a mixing (weak dependence) condition that is conveniently defined by means of the strong mixing coefficients; see e.g., Rosenblatt (1985). In particular, we say that the series $\{X_t, t \in \mathbf{Z}\}$ is strong mixing if $\alpha_X(k) \to 0$, as $k \to \infty$, where $\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$, and $A \in \mathcal{F}_{-\infty}^{0}$, $B \in \mathcal{F}_{k}^{\infty}$ are events in the σ-algebras generated by $\{X_n, n \leq 0\}$ and $\{X_n, n \geq k\}$ respectively. If in addition

$$E|X_1|^{2+\delta} < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_X^{\delta/(2+\delta)}(k) < \infty \tag{1}$$

for some $\delta > 0$, then the limit of σ_N^2 exists (denoted by $\sigma_\infty^2 = \sum_{s=-\infty}^{\infty} R(s)$), and in addition, $\sqrt{N}(\overline{X}_N - \mu) \overset{\mathcal{L}}{\Longrightarrow} N(0, \sigma_\infty^2)$ as $N \to \infty$; see Ibragimov and Linnik (1971).

Many estimators of σ_∞^2 have been proposed in the literature; see e.g., Politis et al. (1999) for some discussion. In the next section, we focus on estimators constructed via block resampling in two popular forms: the circular/moving blocks bootstrap and the stationary bootstrap,[a] and we address the important practical problem of estimation of the optimal block size. An illustration of the proposed block selection algorithm and some examples are given in Sec. 4. Technical proofs are provided in the Appendix.

3. PRACTICAL BLOCK SIZE CHOICE

3.1. Brief Review of Block Bootstrap Methods

A general block bootstrap algorithm can be defined as follows:

(1) Start by "wrapping" the data $\{X_1, \ldots, X_N\}$ around a circle, i.e., define the new series $Y_t := X_{t_{mod(N)}}$, for $t \in \mathbf{N}$, where $mod(N)$ denotes "modulo N".
(2) Let $i_0, i_1, \ldots,$ be drawn i.i.d. with uniform distribution on the set $\{1, 2, \ldots, N\}$; these are the starting points of the new blocks.
(3) Let $b_0, b_1, \ldots,$ be drawn i.i.d. from some distribution $F_b(\cdot)$ that depends on a parameter b (that may depend on N and will be specified later); these are the block sizes.
(4) Construct a bootstrap pseudo-series $Y_1^*, Y_2^*, \ldots,$ as follows. For $m = 0, 1, \ldots,$ let

$$Y^*_{mb_m+j} := Y_{i_m+j-1} \qquad \text{for } j = 1, 2, \ldots, b_m.$$

This procedure defines a probability measure (conditional on the data $X_1, \ldots, X_N$) that will be denoted P^*; expectation and variance with respect to P^* are denoted E^* and Var^* respectively.
(5) Finally, we focus on the first N points of the bootstrap series and construct the bootstrap sample mean $\overline{Y}^*_N = N^{-1} \sum_{i=1}^N Y_i^*$. The corresponding estimate of the asymptotic variance of the sample mean is then given by $Var^*(\sqrt{N}\,\overline{Y}^*_N)$.

We will explicitly address two interesting cases:

(A) The distribution F_b is a unit mass on the positive integer b; this is the *circular bootstrap* (CB) of Politis and Romano (1992). Its corresponding estimate of σ^2_∞ will be denoted $\sigma^2_{b,CB}$.
(B) The distribution F_b is a Geometric distribution with mean equal to the real number b; this is the *stationary bootstrap* (SB) of Politis and Romano (1994). Its corresponding estimate of σ^2_∞ will be denoted $\sigma^2_{b,SB}$.

---

[a]There is yet another block bootstrap methodology that has been recently introduced, namely the *tapered* block bootstrap. Tapering the blocks before allowing them to be included in a bootstrap pseudo-sample has many favorable properties including a faster rate of convergence; for more details see Paparoditis and Politis (2001, 2002).

The circular bootstrap is an asymptotically equivalent variation of the moving blocks (MB) bootstrap of Künsch (1989) and Liu and Singh (1992) whose estimate of σ_∞^2 may be simply written as $\hat{\sigma}_{b,MB}^2 = (b/Q)\sum_{i=1}^{Q}(\overline{X}_{i,b} - \overline{X}_N)^2$; here $\overline{X}_{i,b} = b^{-1}\sum_{t=i}^{i+b-1} X_t$, and $Q = N - b + 1$. Note that the estimator $\hat{\sigma}_{b,MB}^2$ is found in the literature in many asymptotically equivalent variations and under many different names, including the following: Bartlett spectral density estimator (at the origin)—Bartlett (1946, 1950); moving block bootstrap—Künsch (1989), Liu and Singh (1992); full-overlap subsampling—Politis and Romano (1993); and overlapping batch means estimator—Schmeiser (1982, 1990).

Note that both the circular bootstrap and the stationary bootstrap share with the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) the property of higher-order[b] accurate estimation of the distribution of the sample mean after standardization/studentization; see Lahiri (1991, 1999a), Politis and Romano (1992), and Götze and Künsch (1996).

Under mixing and moment conditions, consistency of both $\hat{\sigma}_{b,CB}^2$ and $\sigma_{b,SB}^2$ was shown in Politis and Romano (1992, 1994). In a recent paper, Lahiri (1999b) provides a detailed approximation to the first two moments of $\hat{\sigma}_{b,CB}^2$ and $\sigma_{b,SB}^2$ that is very useful and is given below. To state it, we define the spectral density function as $g(w) := \sum_{s=-\infty}^{\infty} R(s)\cos(ws)$.

Theorem 3.1 (Lahiri, 1999b). *Assume* $E|X_t|^{6+\delta} < \infty$, *and* $\sum_{k=1}^{\infty} k^2(\alpha_X(k))^{\frac{\delta}{6+\delta}} < \infty$ *for some* $\delta > 0$. *If* $b \to \infty$ *as* $N \to \infty$ *but with* $b = o(N^{1/2})$, *then we have*:

$$Bias(\hat{\sigma}_{b,CB}^2) = E\hat{\sigma}_{b,CB}^2 - \sigma_\infty^2 = -\frac{1}{b}G + o(1/b); \qquad (2)$$

$$Var(\hat{\sigma}_{b,CB}^2) = \frac{b}{N}D_{CB} + o(b/N); \qquad (3)$$

$$Bias(\hat{\sigma}_{b,SB}^2) = E\hat{\sigma}_{b,SB}^2 - \sigma_\infty^2 = -\frac{1}{b}G + o(1/b); \qquad (4)$$

$$Var(\hat{\sigma}_{b,SB}^2) = \frac{b}{N}D_{SB} + o(b/N); \qquad (5)$$

in the above, $D_{CB} = \frac{4}{3}g^2(0), D_{SB} = \left(4g^2(0) + \frac{2}{\pi}\int_{-\pi}^{\pi}(1 + \cos w)g^2(w)dw\right)$, *and* $G = \sum_{k=-\infty}^{\infty} |k|R(k)$.

From the above theorem it is apparent that the SB is less accurate than the CB for estimating σ_∞^2. Although the two methods have similar bias (to the first order), the SB has higher variance due to the additional randomization involved in drawing the random block sizes.

[b]Higher-order accuracy is typically defined by a comparison to the Central Limit Theorem that is concurrently available under (1); thus, the aforementioned bootstrap schemes are all more accurate as compared to the benchmark of the standard normal approximation to the distribution of the standardized and/or studentized sample mean.

To compare the two methods, we may define the asymptotic relative efficiency (ARE) of the SB relative to the CB as

$$ARE_{CB/SB} := \lim_{N\to\infty} \frac{MSE_{opt,CB}}{MSE_{opt,SB}},$$

where $MSE_{opt,CB} := \inf_b MSE(\hat{\sigma}^2_{b,CB})$, and $MSE_{opt,SB} := \inf_b MSE(\hat{\sigma}^2_{b,SB})$.

From the previous remarks, it is intuitive that this $ARE_{CB/SB}$ is less than one. Nevertheless, contrary to a claim in Lahiri (1999b), this $ARE_{CB/SB}$ is *always* bounded away from zero; the subject of the following lemma is a corrected bound on the $ARE_{CB/SB}$.

Lemma 3.1. *Under the assumptions of Theorem* 3.1 *we have*:

$$0.331 \leq ARE_{CB/SB} \leq 0.481.$$

Lemma 3.1 gives a precise bound on the price we must pay in order to have a block-bootstrap method that generates *stationary* bootstrap sample paths; the stationarity of bootstrap sample paths is a convenient property—see e.g., Politis and Romano (1994) or White (2000).

Nevertheless, the above definition of asymptotic relative efficiency involves a comparison of the theoretically optimized (with respect to block size choice) SB and CB methods; but the optimal block size is *never* known in practice, and—more often than not—the block size used is suboptimal. Interestingly the SB method is *less* sensitive to block size misspecification as compared to CB and/or the moving blocks bootstrap—see Politis and Romano (1994). We achieve a more realistic comparison of the two methods based on the new notion of finite-sample attainable relative efficiency introduced in Sec. 4.

The problem of empirically optimizing the block size choice is as challenging as it is practically important. In the next two subsections a new method of optimal block size choice is put forth for both SB and CB methods.

3.2. Choosing the Expected Block Size for the Stationary Bootstrap

From Theorem 3.1 it follows that for the stationary bootstrap we have:

$$MSE(\hat{\sigma}^2_{b,SB}) = \frac{G^2}{b^2} + D_{SB}\frac{b}{N} + o(b^{-2}) + o(b/N).$$

It now follows that the large-sample $MSE(\sigma^2_{b,SB})$ is minimized if we choose

$$b_{opt,SB} = \left(\frac{2G^2}{D_{SB}}\right)^{1/3} N^{1/3}. \tag{6}$$

Using the optimal block size $b_{opt,SB}$ we achieve the optimal MSE, which is given by

$$MSE_{opt,SB} \approx \frac{3}{2^{2/3}} \frac{G^{2/3} D_{SB}^{2/3}}{N^{2/3}}. \tag{7}$$

The quantities G and D_{SB} involve the unknown parameters $\sum_{k=-\infty}^{\infty} |k|R(k)$, $\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} R(k) = g(0)$, and $\frac{1}{\pi}\int_{-\pi}^{\pi}(1+\cos w)g^2(w)dw$; these must be (accurately) estimated in order to obtain a practically useful procedure.

To achieve accurate estimation of the infinite sum $\sum_{k=-\infty}^{\infty} |k|R(k)$ above, as well as the infinite sum $\sum_{k=-\infty}^{\infty} R(k)\cos(wk)$ that equals the spectral density $g(w)$, we propose using the "flat-top" lag-window of Politis and Romano (1995). Thus, we estimate $\sum_{k=-\infty}^{\infty} |k|R(k)$ by $\sum_{k=-M}^{M} \lambda(k/M)|k|\widehat{R}(k)$, where $\widehat{R}(k) = N^{-1}\sum_{i=1}^{N-|k|}(X_i - \overline{X}_N)(X_{i+|k|} - \overline{X}_N)$, and the function $\lambda(t)$ has a trapezoidal shape symmetric around zero, i.e.,

$$\lambda(t) = \begin{cases} 1 & \text{if } |t| \in [0, 1/2] \\ 2(1-|t|) & \text{if } |t| \in [1/2, 1] \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we estimate $g(w) = \sum_{k=-\infty}^{\infty} R(k)\cos(wk)$ by $\hat{g}(w) = \sum_{k=-M}^{M} \lambda(k/M) \times\widehat{R}(k)\cos(wk)$. Plugging in our two estimators in the expressions for G and D_{SB}, we arrive at the estimators

$$\widehat{G} = \sum_{k=-M}^{M} \lambda(k/M)|k|\widehat{R}(k) \quad \text{and} \quad \widehat{D}_{SB} = \left(4\hat{g}^2(0) + \frac{2}{\pi}\int_{-\pi}^{\pi}(1+\cos w)\hat{g}^2(w)dw\right). \tag{8}$$

Thus, our estimator for the (expected) block size choice is given by:

$$\hat{b}_{opt,SB} = \left(\frac{2\widehat{G}^2}{\widehat{D}_{SB}}\right)^{1/3} N^{1/3}. \tag{9}$$

One reason for using the flat-top lag-window $\lambda(t)$ is that smoothing with the flat-top lag-window is highly accurate, taking advantage of a possibly fast rate of decay of the autocovariance $R(k)$, and thus achieving the best rate of convergence possible. In order to investigate the asymptotic performance of our suggested $\hat{b}_{opt,SB}$ we give the following result.

Theorem 3.2. *Assume the conditions of Theorem* 3.1 *hold.*

(i) *Assume that* $\sum_{s=-\infty}^{\infty} |s|^{(r+1)}|R(s)| < \infty$ *for some positive integer* r; *then taking* M *proportional to* $N^{1/(2r+1)}$ *yields*

$$\hat{b}_{opt,SB} = b_{opt,SB}(1 + O_P(N^{-r/(2r+1)})).$$

(ii) *If* $R(k)$ *has an exponential decay, then taking* $M \sim A\log N$, *for some given non-negative constant* A, *yields*

$$\hat{b}_{opt,SB} = b_{opt,SB}\left(1 + O_P\left(\frac{\sqrt{\log N}}{\sqrt{N}}\right)\right). \tag{10}$$

(iii) *If* $R(k) = 0$ *for* $|k|$ *greater than some integer* q, *then taking* $M = 2q$ *yields*

$$\hat{b}_{opt,SB} = b_{opt,SB}\left(1 + O_P\left(\frac{1}{\sqrt{N}}\right)\right).$$

Besides the fast convergence and adaptivity to the underlying correlation structure, another equally important reason for using the flat-top lag-window is that choosing the bandwidth M for the flat-top lag-window in practice is intuitive and feasible by a simple inspection of the correlogram, i.e., the plot of $\widehat{R}(k)$ vs. k. In particular, Politis and Romano (1995) suggest looking for the smallest integer, say $\widehat{m}$, after which the correlogram appears negligible, i.e., $\widehat{R}(k) \simeq 0$ for $k > \widehat{m}$. Of course, $\widehat{R}(k) \simeq 0$ is taken to mean that $\widehat{R}(k)$ is *not* significantly different from zero, i.e., an implied hypothesis test.[c] After identifying $\widehat{m}$ on the correlogram, the recommendation is to just take $M = 2\widehat{m}$.

We now further discuss the $M = 2\widehat{m}$ recommendation in the specific context of an exponential decay of $R(k)$; such a fast decay is often encountered, e.g., all stationary ARMA models are characterized by such a fast decay—cf. Brockwell and Davis (1991). First note that the "recipe" $M = 2\widehat{m}$, where $\widehat{m}$ is gotten by a correlogram inspection, does *not* contradict the recommendation $M \sim A \log N$ offered in Theorem 3.2(ii). On the contrary, the $M = 2\widehat{m}$ recipe should be viewed as an empirical way to obtain the optimal constant A in $M \sim A \log N$. To see this, recall that the autocovariance $R(k)$ of a stationary ARMA model satisfies $R(k) \simeq \text{const} \times \xi^k$, for large k, where ξ is essentially the modulus of the characteristic polynomial root that is closest to the unit circle. Let the autocorrelations be defined as $\rho_X(k) := R(k)/R(0)$; therefore, the estimated autocorrelations are given by $\hat{\rho}_X(k) := \widehat{R}(k)/\widehat{R}(0) \simeq C\xi^k + O_P(1/\sqrt{N})$ for some constant C. To say that $\widehat{R}(k) \simeq 0$ for $k > \widehat{m}$ means that $\hat{\rho}_X(\widehat{m}+1)$ is not significantly different from zero; this in turn means that $-c/\sqrt{N} < \hat{\rho}_X(\widehat{m}+1) < c/\sqrt{N}$ for some constant c. Putting this all together, it follows that with probability tending to one we have

$$A_1 \log N < \widehat{m} < A_2 \log N$$

for some positive constants A_1, A_2.

Perhaps the most attractive feature of the $M = 2\widehat{m}$ recipe is its adaptivity to different correlation structures. Arguments similar to those just given show that, if the autocovariance $R(k)$ has a polynomial (as opposed to exponential) decay, then $\widehat{m}$ grows at a polynomial rate, as is advisable in that case—see Theorem 3.2(i). In addition, if $R(k) = 0$ for $|k| > q$ (but $R(q) \neq 0$), then it is easy to see that

[c] A precise formulation of this implied hypothesis test is given in Politis (2001) and can be described as follows: Let $\rho(k) = R(k)/R(0)$, $\hat{\rho}(k) = \widehat{R}(k)/\widehat{R}(0)$, and let $\widehat{m}$ be the smallest positive integer such that $|\hat{\rho}(\widehat{m}+k)| < c\sqrt{\log N/N}$, for $k = 1, \ldots, K_N$, where $c > 0$ is a fixed constant, and K_N is a positive, nondecreasing integer-valued function of N such that $K_N = o(\log N)$. Taking log to denote logarithm with base 10, recommended practical values for the above are $c = 2$ and $K_N = \max(5, \sqrt{\log N})$.

$\widehat{m} \xrightarrow{P} q$; this corresponds to the interesting case of MA(q) models, i.e., the set-up of Theorem 3.2(iii). Thus, the recipe $M = 2\widehat{m}$, is an *omnibus* rule-of-thumb that automatically gives good bandwidth choices without having to prespecify the correlation structure. Finally, note that the simple, correlogram-based, $M = 2\widehat{m}$ recipe can *not* be applied to traditional lag-windows; it is *only* applicable in connection with the flat-top lag-windows of Politis and Romano (1995).

3.3. Choosing the Block Size for the Circular Bootstrap

Theorem 3.1 similarly implies that for the circular bootstrap we have:

$$MSE(\hat{\sigma}^2_{b,CB}) = \frac{G^2}{b^2} + D_{CB}\frac{b}{N} + o(b^{-2}) + o(b/N).$$

It now follows that the large-sample $MSE(\sigma^2_{b,CB})$ is minimized if we choose

$$b_{opt,CB} = \left[\left(\frac{2G^2}{D_{CB}}\right)^{1/3} N^{1/3}\right] \tag{11}$$

where $[x]$ indicates the closest integer to the real number x. Using the optimal block size $b_{opt,CB}$ we achieve the optimal MSE, which is now given by

$$MSE_{opt,CB} \approx \frac{3}{2^{2/3}} \frac{G^{2/3} D_{CB}^{2/3}}{N^{2/3}}. \tag{12}$$

Plugging in our estimator $\hat{g}$ for g in the expression for D_{CB} we obtain

$$\widehat{D}_{CB} = \frac{4}{3}\hat{g}^2(0). \tag{13}$$

Estimating G by $\widehat{G}$ as given in (8), we are led to the following optimal block size estimator:

$$\hat{b}_{opt,CB} = \left[\left(\frac{2\widehat{G}^2}{\widehat{D}_{CB}}\right)^{1/3} N^{1/3}\right]. \tag{14}$$

The behavior of $\hat{b}_{opt,CB}$ is similar to that of $\hat{b}_{opt,SB}$ as the following theorem shows.

Theorem 3.3. *Assume the conditions of Theorem* 3.1 *hold.*

(i) *Assume that* $\sum_{s=-\infty}^{\infty} |s|^{(r+1)}|R(s)| < \infty$ *for some positive integer* r; *then taking* M *proportional to* $N^{1/(2r+1)}$ *yields*

$$\hat{b}_{opt,CB} = b_{opt,CB}(1 + O_P(N^{-r/(2r+1)})).$$

(ii) *If $R(k)$ has an exponential decay, then taking $M \sim A \log N$, for some given non-negative constant A, yields*

$$\hat{b}_{opt,CB} = b_{opt,CB}\left(1 + O_P\left(\frac{\sqrt{\log N}}{\sqrt{N}}\right)\right). \tag{15}$$

(iii) *If $R(k) = 0$ for $|k|$ greater than some integer q, then taking $M = 2q$ yields*

$$\hat{b}_{opt,CB} = b_{opt,CB}\left(1 + O_P\left(\frac{1}{\sqrt{N}}\right)\right).$$

Note that the moving blocks bootstrap variance estimator $\hat{\sigma}^2_{b,MB}$ and the circular bootstrap variance estimator $\hat{\sigma}^2_{b,CB}$ have identical[d] (at least to first order) bias and variance; consequently, the large-sample optimal block size is the same, i.e., $b_{opt,MB} \equiv b_{opt,CB}$. Therefore, the estimator $\hat{b}_{opt,CB}$ can be considered to be an estimator of the optimal block size for the moving blocks bootstrap as well, i.e., $\hat{b}_{opt,MB} \equiv \hat{b}_{opt,CB}$. As Theorem 3.3 shows, our estimator $\hat{b}_{opt,MB}$ has a faster rate of convergence than that of the block size estimator proposed in Bühlmann and Künsch (1999), and the difference is especially pronounced when the autocovariance $R(k)$ has a fast decay. To elaborate, recall that the Bühlmann and Künsch (1999) block size estimator, denoted by $\bar{b}_{opt,MB}$, generally satisfies

$$\bar{b}_{opt,MB} = b_{opt,MB}(1 + O_P(N^{-2/7})).$$

By contrast, note that

$$\hat{b}_{opt,MB} = b_{opt,MB}(1 + O_P(N^{-1/3}))$$

under *any* of the autocovariance decay conditions considered in Theorem 3.3; this is true, for example, under the slowest decay condition, i.e., condition (i) with $r = 1$. If the autocovariance $R(k)$ happens to have a faster decay, then $\hat{b}_{opt,MB}$ becomes more accurate whereas the accuracy of $\bar{b}_{opt,MB}$ is *not* improved; in the interesting example of exponential decay of $R(k)$, Theorem 3.3(ii) shows that

$$\hat{b}_{opt,MB} = b_{opt,CB}\left(1 + O_P\left(\frac{\sqrt{\log N}}{\sqrt{N}}\right)\right).$$

Thus, $\bar{b}_{opt,MB}$ is outperformed by $\hat{b}_{opt,MB}$ under a wide range of conditions, namely any of the conditions considered in Theorem 3.3; the constrast is more dramatic under conditions (ii) and (iii).

Finally note that, although the subsampling/cross-validation method for block size selection of Hall et al. (1995) is intuitively appealing, no information on its rate of convergence (besides consistency) has yet been established.

---

[d]As shown in Künsch (1989), $\hat{\sigma}^2_{b,MB}$ satisfies Eqs. (2) and (3) with the *same* constants as given for the circular case. In other words, $Bias(\hat{\sigma}^2_{b,MB}) = -\frac{1}{b}G + o(1/b)$, and $Var(\hat{\sigma}^2_{b,MB}) = \frac{b}{N}\frac{4}{3}g^2(0) + o(b/N)$.

4. ILLUSTRATION OF BLOCK SELECTION ALGORITHM

Having presented the stationary and the circular bootstrap in a unified way, we have compared their performances in Lemma 3.1 which is a corrected version of earlier results by Lahiri (1999b). Noting that the performance of either method crucially depends on the block size used, we have presented a novel methodology of selection/estimation of optimal block sizes. The methodology is based on the notion of spectral estimation via the flat-top lag-windows of Politis and Romano (1995), and it is outlined below.

Block Selection Algorithm via Flat-Top Lag-Windows

(1) Identify the smallest integer, say $\widehat{m}$, after which the correlogram appears negligible, i.e., $\widehat{R}(k) \simeq 0$ for $k > \widehat{m}$, using the procedure introduced in Politis (2001) and outlined in the footnote to Sec. 3.2 in this paper.
(2) Using the value $M = 2\widehat{m}$, estimate G, D_{SB} and D_{CB} by $\widehat{G}$, $\widehat{D}_{SB}$ and $\widehat{D}_{CB}$ as given in (8) and (13).
(3) Estimate the optimal (expected) block size $\hat{b}_{opt,SB}$ for the stationary bootstrap as in (9), and the optimal block size $\hat{b}_{opt,CB}$ for the circular and/or moving blocks bootstrap as in (14).

Note that the above algorithm is fully automatic.[e] Indeed, Dr. Andrew Patton of the London School of Economics has compiled a Matlab computer code for implementing the above block selection algorithm via flat-top lag-windows; his code is now made publicly available from his website: http://fmg.lse.ac.uk/~patton/code.html.

Based on Dr. Patton's code a small simulation was conducted in which time series were generated of length N (with N being either 200 or 800), from the AR(1) model: $X_t = \rho\, X_{t-1} + Z_t$, with $\{Z_t\} \sim$ i.i.d. $N(0, 1)$. The values for the parameter ρ were chosen as 0.7, 0.1, and -0.4. For each ρ and N combination, 1000 series were generated. Table 1 contains the theoretical values of the optimal block sizes $b_{opt,SB}$ and $b_{opt,CB}$ that can be analytically calculated from (6) and (11) by our knowledge regarding the underlying AR(1) model.

Table 2 contains the mean, standard deviation, and Root Mean Squared Error (RMSE) computed over the 1000 replications of the quantity $\hat{b}_{opt,SB}/b_{opt,SB}$ in each of the different cases. Since the AR(1) model satisfies the assumptions of Theorem 3.2 (ii) we expect that $\hat{b}_{opt,SB}/b_{opt,SB} = 1 + O_P(\sqrt{\log N}/\sqrt{N})$. This theoretical result from Theorem 3.2 is supported by the simulation; in particular, note the approximate halving of the RMSE going from the case $N = 200$ to $N = 800$. Interestingly, in the case $\rho = 0.7$, the bias of $\hat{b}_{opt,SB}/b_{opt,SB}$ is significant, yielding an important contributing to the RMSE; by contrast, in the cases where ρ is 0.1 or -0.4, the bias seems negligible. For illustration purposes, Figure 1 shows a histogram of the

[e]Nevertheless, it should be stressed that valuable information will invariably be gained by looking at the *correlogram*, i.e., a plot of $\hat{\rho}(k)$ vs. k; the automatic procedure should complement—rather than replace—this correlogram examination.

Table 1. Theoretical optimal block sizes $b_{opt,SB}$ and $b_{opt,CB}$; the brackets [·] indicate "closest integer" to the entry.

		$b_{opt,SB}$	$b_{opt,CB}$
$\rho = 0.7$,	$N = 200$	12.0043	[18.52]
	$N = 800$	19.0557	[29.40]
$\rho = 0.1$,	$N = 200$	1.3106	[2.31]
	$N = 800$	2.0805	[3.66]
$\rho = -0.4$,	$N = 200$	2.7991	[5.70]
	$N = 800$	4.4432	[9.04]

Table 2. Empirical mean, standard deviation, and root mean squared error (RMSE) of the quantity $\hat{b}_{opt,SB}/b_{opt,SB}$.

$\hat{b}_{opt,SB}/b_{opt,SB}$		Mean	St. dev.	RMSE
$\rho = 0.7$,	$N = 200$	0.646	0.383	0.521
	$N = 800$	0.619	0.222	0.441
$\rho = 0.1$,	$N = 200$	1.030	0.858	0.858
	$N = 800$	0.827	0.421	0.455
$\rho = -0.4$,	$N = 200$	1.107	0.704	0.712
	$N = 800$	1.013	0.334	0.334

distribution of our estimator $\hat{b}_{opt,SB}$ for $\rho = 0.7$ in the two cases: Figure 1(a) for $N = 200$ and Figure 1(b) for $N = 800$.

Table 3 is similar to Table 2 but focuses instead on the quantity $\hat{b}_{opt,CB}/b_{opt,CB}$. Comparing line-by-line the entries of Table 3 to those of Table 2, we notice an important pattern: the RMSEs of Table 2 are much *smaller* than those of Table 3. Coupled with the fact that $b_{opt,SB}$ is invariably smaller than $b_{opt,CB}$—see Table 1—it follows that $\hat{b}_{opt,SB}$ is a more accurate estimator than $\hat{b}_{opt,CB}$. In other words, estimating the optimal (expected) block size in the stationary bootstrap seems to be an *easier* problem than estimating the optimal block size in the circular and/or moving blocks bootstrap. In addition, recall that the stationary bootstrap is less sensitive to block size misspecification; see e.g., Politis and Romano (1994).

The above considerations motivate the introduction of a new way of comparing the performance of the two methods. Thus, we now define the finite-sample "attainable" relative efficiency (FARE) of the SB relative to CB as

$$FARE_{CB/SB} := \frac{MSE_{\hat{b}_{opt},CB}}{MSE_{\hat{b}_{opt},SB}},$$

where $MSE_{\hat{b}_{opt},CB} := MSE(\hat{\sigma}^2_{\hat{b}_{opt},CB})$, and $MSE_{\hat{b}_{opt},SB} := MSE(\hat{\sigma}^2_{\hat{b}_{opt},SB})$. Note that the $FARE_{CB/SB}$ depends on the sample size N although it is not explicitly denoted. More importantly, the $FARE_{CB/SB}$ compares the performance of SB to that of CB when both are used in connection with *estimated* optimal block sizes which is the case of

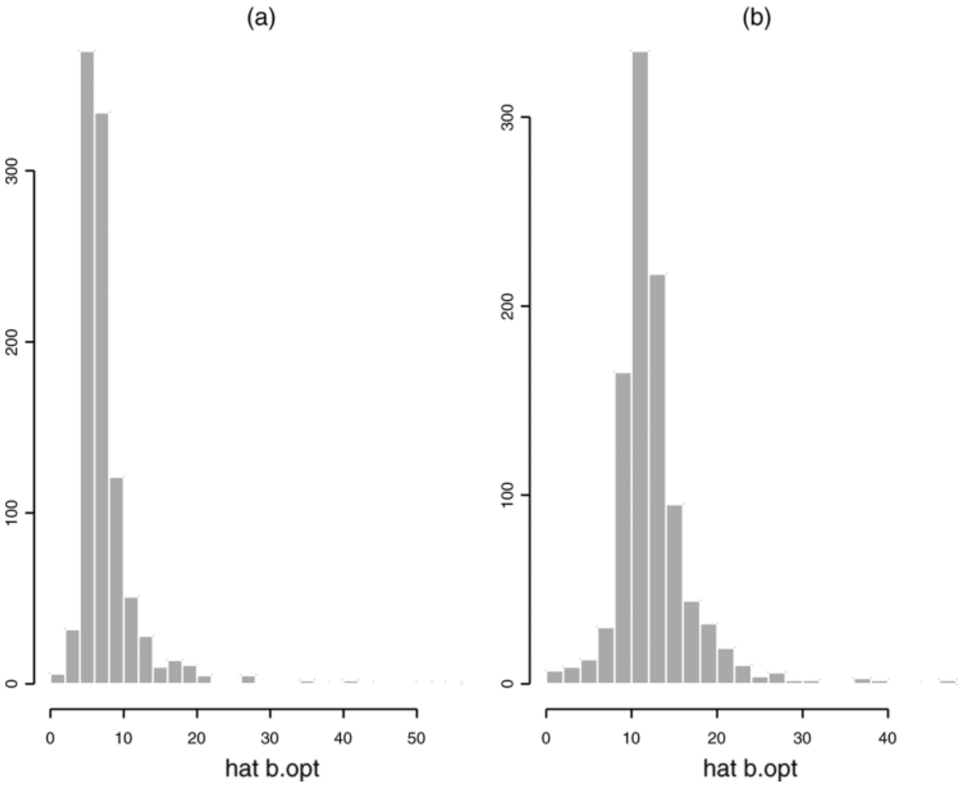

Figure 1. Histogram of $\hat{b}_{opt,SB}$ for $\rho = 0.7$; the two cases: (a) $N = 200$ and (b) $N = 800$. (*View this art in color in www.dekker.com.*)

practical interest; recall that the $ARE_{CB/SB}$ compared the MSEs of SB and CB when those were used in connection with the *true* optimal block sizes (assumed known).

It would be illuminating to be able to give some bounds on the $FARE_{CB/SB}$ in the spirit of Lemma 3.1 but this seems too difficult for the present moment. Nevertheless, from our previous remarks, it is expected that the $FARE_{CB/SB}$ will be greater than the $ARE_{CB/SB}$. Although theoretical analysis seems to be intractable, we can investigate the behavior of $FARE_{CB/SB}$ via simulation.

Table 3. Empirical mean, standard deviation, and root mean squared error (RMSE) of the quantity $\hat{b}_{opt,CB}/b_{opt,CB}$.

$\hat{b}_{opt,CB} / b_{opt,CB}$		Mean	St. dev.	RMSE
$\rho = 0.7$,	$N = 200$	0.523	0.656	0.811
	$N = 800$	0.471	0.186	0.561
$\rho = 0.1$,	$N = 200$	1.155	1.543	1.551
	$N = 800$	1.012	0.554	0.554
$\rho = -0.4$,	$N = 200$	1.868	2.311	2.469
	$N = 800$	1.371	0.565	0.676

Table 4. The true σ^2_∞, and the mean and MSE of its two estimators based on estimated block size; the last column indicates the finite-sample attainable relative efficiency (FARE) of the SB relative to the CB.

		σ^2_∞	$E\hat{\sigma}^2_{\hat{b}_{opt},SB}$	$E\hat{\sigma}^2_{\hat{b}_{opt},CB}$	$MSE_{\hat{b}_{opt},SB}$	$MSE_{\hat{b}_{opt},CB}$	$FARE_{CB/SB}$
$\rho = 0.7$,	$N = 200$	11.111	7.016	7.787	25.691	22.569	0.878
	$N = 800$	11.111	8.808	9.433	10.555	8.421	0.798
$\rho = 0.1$,	$N = 200$	1.235	1.063	1.132	0.059	0.055	0.940
	$N = 800$	1.235	1.101	1.157	0.030	0.021	0.712
$\rho = -0.4$,	$N = 200$	0.510	0.699	0.553	0.074	0.028	0.381
	$N = 800$	0.510	0.619	0.543	0.023	0.008	0.363

Table 4 reports the performance (bias, MSE and FARE) of the two methods based on estimated block sizes in the setting of our AR(1) example. To construct the entries of Table 4, the following procedure was followed: for each generated series, the estimated optimal block sizes (for SB and CB) were computed using the algorithm of this section; then the SB and CB estimators of σ^2_∞ for that series were computed using those estimated optimal block sizes that were specific to that particular series.

Table 4 is quite informative. First note that—except for the case of negative dependence—the FAREs are very large, definitely outside the maximum value of 0.481 prescribed for the AREs by Lemma 3.1. Interestingly, the two positive dependence cases ($\rho = 0.7$ and 0.1) yields FAREs close to unity in the small-sample case ($N = 200$); this is strong indication of the block size effects previously alluded to. The fact that the $FARE_{CB/SB}$ is small (and potentially quite close to $ARE_{CB/SB}$) when $\rho = -0.4$ could be attributed to a reduced sensitivity of the two estimators of σ^2_∞ to block size in this case.

We also note that in all cases the FAREs seem to drop when the sample size increases. To explain this phenomenon, we offer the following conjecture:

Conjecture. *Under the assumptions of Theorem* 3.2 (*with the possible exception of the* $r = 1$ *case in part* (i)), *we conjecture that* $FARE_{CB/SB} \to ARE_{CB/SB}$ *as* $n \to \infty$.

The rationale behind the above conjecture is the following; to fix ideas, consider the clauses of part (ii) of Theorem 3.2 that corresponds to the exponential decay associated with ARMA models—including our AR(1) example. We thus have

$$\hat{b}_{opt,SB} = b_{opt,SB}\left(1 + O_P\left(\frac{\sqrt{\log N}}{\sqrt{N}}\right)\right) = b_{opt,SB} + O_P\left(\frac{\sqrt{\log N}}{N^{1/6}}\right) \tag{16}$$

where we have used the fact that $b_{opt,SB}$ is of the order of $N^{1/3}$. Thus, we not only have that $\hat{b}_{opt,SB} / b_{opt,SB} \to 1$ in probability; we also have $\hat{b}_{opt,SB} - b_{opt,SB} \to 0$ albeit at a very slow rate. Therefore, for (really) large samples, the values of $\hat{b}_{opt,SB}$ and $b_{opt,SB}$ should approach each other. A similar behavior holds for $\hat{b}_{opt,CB}$, thus giving support to our conjecture.

However, note that the rate of the (alleged) convergence of $FARE_{CB/SB}$ to $ARE_{CB/SB}$ would be excruciatingly slow. To see this, note that this convergence is governed by the fact that $\sqrt{\log N}/N^{1/6}$ tends to 0 but very slowly. Furthermore, the $N^{-1/6}$ factor given above is under the scenario of exponential decay of the correlations; under the polynomial decay of part (i) of Theorem 3.2 the convergence is even slower (and may well break down in the case $r = 1$). It is for this reason that $N = 800$ does not seem to be a sample size large enough to ensure that $FARE_{CB/SB}$ is close to $ARE_{CB/SB}$.

Returning to Table 3, note that Theorem 3.3(ii) leads us to expect that $\hat{b}_{opt,CB}/b_{opt,CB} = 1 + O_P(\sqrt{\log N}/\sqrt{N})$. This fact is again generally supported by our simulation but special note must be made regarding the 3rd and 5th row of the table where the standard deviation seems too large. To fix ideas, we focus on the 3rd row as the phenomenon is similar for the 5th row.

Figure 2 shows a histogram of the distribution of our estimator $\hat{b}_{opt,CB}$ for $\rho = 0.1$ in the two cases: Figure 2(a) for $N = 200$ and Figure 2(b) for $N = 800$. In particular, the center of location—whether measured by a mean or median—of histogram 2(a) is approximately equal to 3 which is quite close to the true $b_{opt,CB}$. However, the histogram is somewhat heavy-tailed: about 5% of its values are bigger

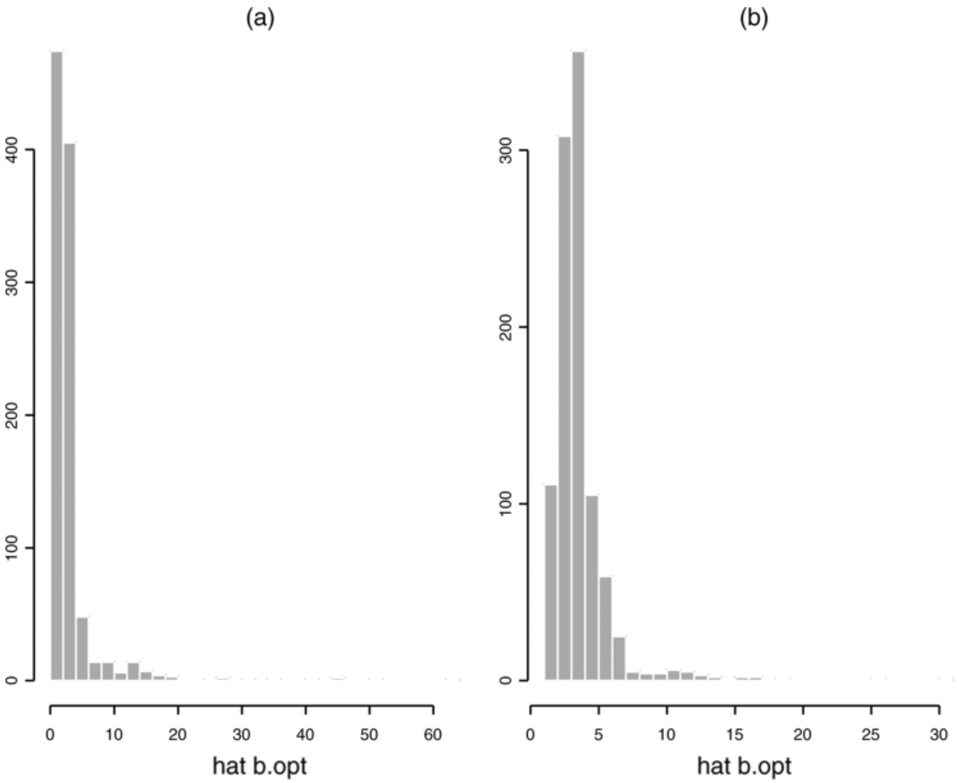

Figure 2. Histogram of $\hat{b}_{opt,CB}$ for $\rho = 0.1$; the two cases: (a) $N = 200$ and (b) $N = 800$. (*View this art in color in www.dekker.com.*)

or equal to 10, and the maximum value is 64 which is extreme relative to a sample size of 200.

By contrast, the histogram 2(b) is free from this undesirable existence of extreme values. For this reason, we believe that this phenomenon is related to the automatic nature of the simulation. As stressed in the footnote in Sec. 4, the rule for estimating $\widehat{m}$ should always be complemented by an examination of the correlogram. Indeed, such an examination is imperative in cases where $\hat{m}$ and the resulting $\hat{b}_{opt,CB}$ are unusually large, as is the case where the latter turns out to be 64.

For example, consider a "problematic" correlogram pictured in Fig. 3 that corresponds to an AR(1) model with $\rho = 0.3$ and $N = 500$. Superimposed are the bands $\pm c\sqrt{\log N/N}$ with $c = 2$ that was recommended in connection with $K_N = \max(5, \sqrt{\log N})$; recall that log denotes logarithm with base 10.

Following the rule proposed in Politis (2001) and outlined in the footnote to Sec. 3.2, we pick $\widehat{m}$ to be the smallest integer such that the correlogram remains within the bands for at least $K_N = 5$ lags after the lag $\widehat{m}$. By strict application of this rule, we should pick $\widehat{m} = 6$. But note that a little tweaking of the values of c and/or K_N yields radically different $\widehat{m}$'s, which is disconcerting. For instance, with $c = 2$ but $K_N = 6$, we would be led to $\widehat{m} = 12$. Alternatively, with $K_N = 5$ but c slightly bigger than 2, the bands would be slightly wider and we would be led to $\widehat{m} = 1$. A warning flag should be raised in such a case and the practitioner should be vigilant.

Note that the values $c = 2$ and $K_N = \max(5, \sqrt{\log N})$ are just recommendations, not absolute requirements. Thus, faced with a problematic correlogram such as in Fig. 3, the practitioner must make a decision drawing upon his/her experience and information concerning the dataset at hand. As a general guideline it should be noted that flat-top lag-window spectral estimators perform best with small values for M; this guideline is in accord with the famous "Okham's razor" that would favor the simplest/smallest among two models with comparable power of explaining the

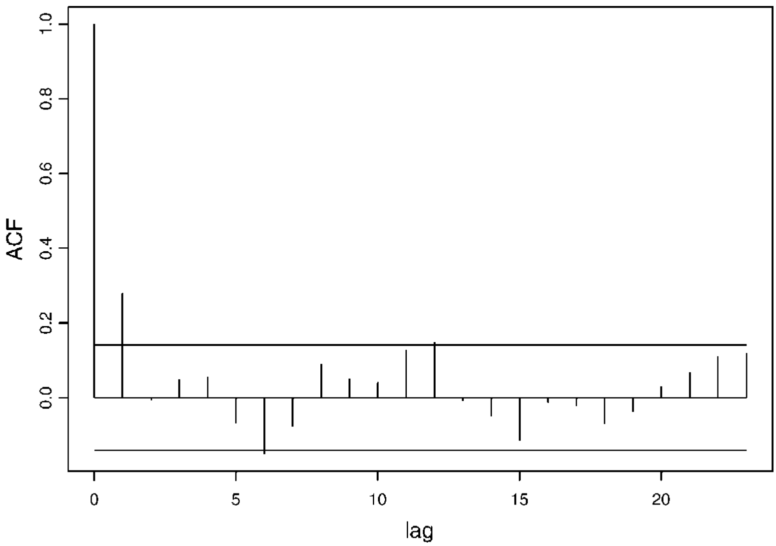

Figure 3. A "problematic" correlogram from an AR(1) model with $\rho = 0.3$ and $N = 500$.

real world. Thus, faced with a dilemma such as the one posed by the correlogram of Fig. 3, we would forego the recommendation $c = 2$ and $K_N = 5$, and instead opt for the simple choice $\widehat{m} = 1$.

5. APPENDIX: TECHNICAL PROOFS

Proof of Lemma 3.1. First note that by Eqs. (7) and (12) we have:

$$ARE_{CB/SB} := \lim_{N\to\infty} \frac{MSE_{opt,CB}}{MSE_{opt,SB}} = \frac{D_{CB}^{2/3}}{D_{SB}^{2/3}}.$$

Thus, to bound the ARE it is sufficient to relate the quantity D_{CB} to the quantity D_{SB}.

Claim. $4g^2(0) \leq D_{SB} \leq 7g^2(0)$.

Proof of Claim. The lower bound is obvious by the positivity of the integrand $(1 + \cos w)g^2(w)$ that features in D_{SB}. For the upper bound, note that by the Cauchy-Schwarz inequality it follows that the quantity

$$\int_{-\pi}^{\pi} (1 + \cos w)\frac{g^2(w)}{g^2(0)}\, dw$$

is maximized if and only if $\frac{g^2(w)}{g^2(0)} = c(1 + \cos w)$ for some constant c; letting $w = 0$ shows that $c = 1/2$. A simple calculation of the integral completes the proof of the claim.

From the claim, it now follows that $3 \leq \frac{D_{SB}}{D_{CB}} \leq 5.25$, and the Lemma is proven. □

Proof of Theorem 3.2. We give the proof of part (ii); the other parts are proven in the same manner. Observe that under the assumed conditions of part (ii) we have that

$$\sum_{k=-M}^{M} \lambda(k/M)\widehat{R}(k)\cos(wk) = \sum_{k=-\infty}^{\infty} R(k)\cos(wk) + O_P(\sqrt{\log N}/\sqrt{N}),$$

i.e., $\hat{g}(w) = g(w) + O_P(\sqrt{\log N}/\sqrt{N})$; see Politis and Romano (1995). Since $g(w)$ is (uniformly) bounded, and the term $O_P(\sqrt{\log N}/\sqrt{N})$ is uniform in w, it follows that[f]

$$\int_{-\pi}^{\pi} (1 + \cos w)\hat{g}^2(w)dw = \int_{-\pi}^{\pi} (1 + \cos w)g^2(w)dw + O_P(\sqrt{\log N}/\sqrt{N}),$$

i.e., $\widehat{D}_{SB} = D_{SB} + O_P(\sqrt{\log N}/\sqrt{N})$.

[f]The quantity $\int_{-\pi}^{\pi}(1 + \cos w)g^2(w)dw$ can also be accurately estimated using the (unsmoothed) periodogram in place of the unknown spectral density g; we use the (smoothed) estimator $\hat{g}$ instead, mainly because $\hat{g}$ has to be calculated anyway for the purposes of estimating D_{SB}.

In the same vein, we can similarly show that

$$\sum_{k=-M}^{M} \lambda(k/M)|k|\widehat{R}(k) = \sum_{k=-\infty}^{\infty} |k|R(k) + O_P(\sqrt{\log N}/\sqrt{N}).$$

An application of the delta method completes the proof. □

Proof of Theorem 3.3. Similar to the proof of Theorem 3.2. □

ACKNOWLEDGMENT

We are indebted to Dr. Andrew Patton of the London School of Economics for compiling a Matlab computer code for practical implementation of the block selection algorithm presented here; the code is publicly available from the website: http://fmg.lse.ac.uk/~patton/code.html.

REFERENCES

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society Supplement* 8:27–41.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika* 37:1–16.

Berkowitz, J., Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Rev.* 19:1–54.

Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. New York: Springer.

Bühlmann, P., Künsch, H. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* 31:295–310.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.* 14:1171–1179.

Götze, F., Künsch, H. (1996). Second order correctness of the blockwise bootstrap for stationary observations, *Ann. Statist.* 24:1914–1933.

Hall, P., Horowitz, J., Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* 82:561–574.

Ibragimov, I. A., Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

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

Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statist. Probab. Lett.* 11:335–341.

Lahiri, S. N. (1999a). On second-order properties of the stationary bootstrap method for studentized statistics. In: Ghosh, S., ed. *Asymptotics, Nonparametrics, and Time Series*. Statistics: Textbooks and Monographs No. 158, New York: Marcel Dekker, pp. 683–711.

Lahiri, S. N. (1999b). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* 27:386–404.

Liu, R. Y., Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In: LePage, R., Billard, L., eds. *Exploring the Limits of Bootstrap*. New York: John Wiley, pp. 225–248.

Paparoditis, E., Politis, D. N. (2001). Tapered block bootstrap. *Biometrika* 88(4):1105–1119.

Paparoditis, E., Politis, D. N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *Econometrics J.* 5(1):131–148.

Politis, D. N. (2003). Adaptive bandwidth choice. *J. Nonparametric Statist.* 15(4–5): 517–533.

Politis, D. N., Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In: LePage, R., Billard, L., eds. *Exploring the Limits of Bootstrap*. New York: John Wiley, pp. 263–270.

Politis, D. N., Romano, J. P. (1993). On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes and Applications* 45:155–167.

Politis, D. N., Romano, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* 89:1303–1313.

Politis, D. N., Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* 16:67–103.

Politis, D. N., Romano, J. P., Wolf, M. (1999). *Subsampling*. New York: Springer.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*. Boston: Birkhäuser.

Schmeiser, B. (1982). Batch size effects in the analysis of simulation output. *Operations Res.* 30:556–568.

Schmeiser, B. (1990). Simulation experiments. In: Heyman, D.P., Sobel, M.J., eds. *Stochastic Models*. North-Holland, pp. 295–330.

White, H. (2000). A reality check for data snooping. *Econometrica* 69:1097–1127.