Tapered Block Bootstrap

Author(s): Efstathios Paparoditis and Dimitris N. Politis

Source: *Biometrika*, Vol. 88, No. 4 (Dec., 2001), pp. 1105-1119
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2673704
Accessed: 23-09-2017 23:22 UTC

Biometrika (2001), **88**, 4, *pp.* 1105–1119

Printed in Great Britain

Tapered block bootstrap

By EFSTATHIOS PAPARODITIS

Department of Mathematics and Statistics, University of Cyprus, Nicosia, Cyprus
stathisp@ucy.ac.cy

AND DIMITRIS N. POLITIS

Department of Mathematics, University of California, San Diego, La Jolla, California 92093-0112, U.S.A.
dpolitis@ucsd.edu

Summary

We introduce and study tapered block bootstrap methodology that yields an improvement over the well-known block bootstrap for time series of Künsch (1989). The asymptotic validity and the favourable bias properties of the tapered block bootstrap are shown. The important practical issues of optimally choosing the window shape and the block size are addressed in detail, while some finite-sample simulations are presented validating the good performance of the tapered block bootstrap.

Some key words: Bias; Confidence interval; Resampling; Spectral density estimation; Subsampling; Time series; Variance estimation.

1. Introduction

Suppose $X_1, \ldots, X_N$ are observations from the strictly stationary real-valued sequence $\{X_n, n \in \mathbb{Z}\}$ having mean $\mu = E(X_t)$ and autocovariance sequence

$$R(s) = E\{(X_t - \mu)(X_{t+|s|} - \mu)\}.$$

Both μ and $R(.)$ are unknown, and the objective is to obtain interval estimates for μ in a nonparametric fashion. To achieve this goal, an approximation is required for the sampling distribution of $\bar{X}_N = N^{-1}\sum_{t=1}^N X_t$. Since $\bar{X}_N$ is typically asymptotically normal, estimating the variance $\sigma_N^2 = \text{var}(N^{1/2}\bar{X}_N) = R(0) + 2\sum_{s=1}^N (1 - s/N)R(s)$, or its asymptotic limit $\sigma_\infty^2 = \sum_{s=-\infty}^\infty R(s)$, becomes crucial.

Recall that sufficient conditions for validity of a Central Limit Theorem for $\bar{X}_N$ are given by $E(|X_1|^{2+\delta}) < \infty$ and $\sum_{k=1}^\infty \alpha_X^{\delta/(2+\delta)}(k) < \infty$, for some $\delta > 0$. The strong mixing coefficients are defined as usual by $\alpha_X(k) = \sup_{A,B}|\text{pr}(A \cap B) - \text{pr}(A)\,\text{pr}(B)|$, where $A \in \mathscr{F}_{-\infty}^0$ and $B \in \mathscr{F}_k^\infty$ are events in the σ-algebras generated by $\{X_n, n \leqslant 0\}$ and $\{X_n, n \geqslant k\}$ respectively; see for example Rosenblatt (1985, pp. 62, 73).

To address this dependent situation, Künsch (1989) proposed the block boostrap procedure. Let $\hat{\sigma}_{b,\text{BB}}^2$ denote the block bootstrap estimator of $\text{var}(N^{1/2}\bar{X}_N)$. Consistency of $\hat{\sigma}_{b,\text{BB}}^2$ is immediate considering that, under regularity conditions,

$$\text{bias}(\hat{\sigma}_{b,\text{BB}}^2) = E(\hat{\sigma}_{b,\text{BB}}^2) - \sigma_\infty^2 = -\frac{1}{b}\sum_{k=-\infty}^{\infty} |k|R(k) + O(b/N) + o(1/b), \tag{1}$$

$$\mathrm{var}(\hat{\sigma}^2_{b,\mathrm{BB}}) = \frac{4}{3}\frac{b}{N}\sigma^4_\infty + o(b/N); \tag{2}$$

see Künsch (1989). It is now apparent that the mean squared error of $\hat{\sigma}^2_{b,\mathrm{BB}}$ is minimised by choosing b proportional to $N^{1/3}$, in which case this mean squared error is of order $O(N^{-2/3})$. The same rate of convergence characterises the circular bootstrap of Politis & Romano (1992b) and the stationary bootstrap of Politis & Romano (1994), as well as the subsampling variance estimator of Carlstein (1986); see for example Lahiri (1999) or Politis et al. (1999, Ch. 10).

Although it does not seem possible to improve upon the order of magnitude $O(b/N)$ of $\mathrm{var}(\hat{\sigma}^2_{b,\mathrm{BB}})$, the order $O(1/b)$ of the bias of $\hat{\sigma}^2_{b,\mathrm{BB}}$ is unnecessarily large. It is for considerations of bias that, in studying the higher-order accuracy of the block bootstrap estimator for the sampling distribution of the sample mean, Götze & Künsch (1996) resort to studentising using a different variance estimator not related to the block bootstrap; see also Davison & Hall (1993). Realising this deficiency, Carlstein et al. (1998) proposed a bootstrap that links the bootstrap blocks so that the bias is reduced. Nevertheless, this matched-block method, implementation of which is cumbersome in practice, is in any case not successful in reducing the bias of the bootstrap variance estimator by an order of magnitude except in the special case when the series $\{X_t\}$ is Markov; see Carlstein et al. (1998, p. 306).

In § 2, we propose and study the tapered block bootstrap which yields a variance estimator with bias of order $O(1/b^2)$ and variance of order $O(b/N)$ in a general class of strong-mixing series $\{X_t\}$; we also show the asymptotic validity and the favourable bias properties of the tapered block bootstrap. Practical issues, including the important issue of optimal block size choice, are discussed in § 3, where some finite-sample simulation results are also given; all proofs are deferred to the Appendix.

2. THE TAPERED BLOCK BOOTSTRAP

We start by centring the data, letting

$$Y_t = X_t - \bar{X}_N \quad (t = 1, 2, \ldots, N);$$

note that a more complete notation for Y_t would be $Y_{t,N}$, but no confusion arises with the simpler notation Y_t. We also introduce a sequence of data-tapering windows $w_n(.)$ for $n = 1, 2, \ldots$; the weights $w_n(t)$ are values in $[0, 1]$, with $w_n(t) = 0$ for $t \notin \{1, 2, \ldots, n\}$. From the above, it is immediate that $\|w_n\|_1 \leqslant n$ and $\|w_n\|_2 \leqslant n^{\frac{1}{2}}$, where $\|w_n\|_1 = \sum_{t=1}^n |w_n(t)|$ and $\|w_n\|_2 = \{\sum_{t=1}^n w_n^2(t)\}^{\frac{1}{2}}$. The idea behind the multiplicative application of a tapering window to data is to give reduced weight to data near the endpoints of the window.

The notion of tapering for time series has been well studied especially in connection with spectral estimation; see for example Welch (1967), Brillinger (1975, Ch. 5), Priestley (1981, Ch. 7), Dahlhaus (1985, 1990) and Künsch (1989). It is customary to construct the sequence of data-tapering windows $w_n(.)$ by means of dilations of a single function $w : \mathbb{R} \to [0, 1]$, so that

$$w_n(t) = w\left(\frac{t - 0{\cdot}5}{n}\right). \tag{3}$$

We will generally assume that the function $w(.)$ satisfies the following assumptions.

Assumption 1. We have $w(t) \in [0, 1]$ for all $t \in \mathbb{R}$, $w(t) = 0$ if $t \notin [0, 1]$, and $w(t) > 0$ for t in a neighbourhood of $\frac{1}{2}$.

Assumption 2. The function $w(t)$ is symmetric about $t = 0{\cdot}5$ and nondecreasing for $t \in [0, \frac{1}{2}]$.

The tapered block bootstrap algorithm is now defined as follows.

Step 1. First choose a positive integer b less than N, and let $i_0, i_1, \ldots, i_{k-1}$ be drawn independently and identically distributed with distribution uniform on the set $\{1, 2, \ldots, Q\}$, where $Q = N - b + 1$; here we take $k = [N/b]$, where $[.]$ denotes the integer part, although different choices for k are also possible.

Step 2. For $m = 0, 1, \ldots, k - 1$, let

$$Y^*_{mb+j} := w_b(j) \frac{b^{\frac{1}{2}}}{\| w_b \|_2} Y_{i_m + j - 1} \quad (j = 1, 2, \ldots, b).$$

Thereby, we construct a bootstrap pseudo-series $Y_1^*, Y_2^*, \ldots, Y_l^*$, where $l = kb$. The above procedure defines a probability measure, conditional on the data $X_1, \ldots, X_N$, that will be denoted by pr*; expectation and variance with respect to pr* are denoted by E^* and var* respectively.

Step 3. Finally, construct the bootstrap sample mean $\bar{Y}_l^* = l^{-1} \sum_{i=1}^{l} Y_i^*$.

The tapered block bootstrap estimator of the asymptotic variance of the sample mean is $\hat{\sigma}^2_{b,\mathrm{TBB}} := \mathrm{var}^*(l^{\frac{1}{2}} \bar{Y}_l^*)$; similarly, the tapered block bootstrap estimator of the sampling distribution $\mathrm{pr}\{N^{\frac{1}{2}}(\bar{X}_N - \mu) \leqslant x\}$ is given by either $\mathrm{pr}^*\{l^{\frac{1}{2}} \bar{Y}_l^* \leqslant x\}$ or $\mathrm{pr}^*\{l^{\frac{1}{2}}(\bar{Y}_l^* - \mu^*) \leqslant x\}$, where μ^* is defined in equation (6).

Note that, if we let $w(t) = 1_{[0,1]}(t)$ be the indicator function of the interval $[0, 1]$, then the tapered block bootstrap boils down to the block bootstrap of Künsch (1989). Nevertheless, as will be shown shortly, the desired bias-correction advantage ensues only if we take w to be continuous, and in particular if the following assumption holds.

Assumption 3. The self-convolution $w * w(t)$ is twice continuously differentiable at the point $t = 0$, where $w * w(t) = \int_{-1}^{1} w(x) w(x + |t|)\, dx$.

Also note that the tapered block bootstrap pseudo-series at 'time' $mb + j$ is given by

$$Y^*_{mb+j} = w_b(j) \frac{b^{\frac{1}{2}}}{\| w_b \|_2} Y_{i_m + j - 1},$$

and not by the naive downweighting scheme $w_b(j) Y_{i_m + j - 1}$. The 'inflation' factor $b^{\frac{1}{2}} / \| w_b \|_2$ is necessary to compensate for the decrease of the variance of the Y_i^*'s effected by the shrinking caused by the window w_b. That this is the correct inflation factor is demonstrated by our theoretical results that follow.

THEOREM 1. *Suppose that equation* (3) *and Assumptions* 1–3 *hold and that* $\sum_{k=1}^{\infty} k^2 |R(k)| < \infty$. *If* $b \to \infty$ *as* $N \to \infty$ *but with* $b = o(N^{1/3})$, *then*

$$E(\hat{\sigma}^2_{b,\mathrm{TBB}}) = \sigma^2_{\infty} + \Gamma / b^2 + o(1/b^2),$$

where

$$\Gamma = \frac{(w * w)''(0)}{2 (w * w)(0)} \sum_{k=-\infty}^{\infty} k^2 R(k).$$

Theorem 1 shows that the tapered block bootstrap method achieves its goal of producing pseudo-series that yield a less biased estimator of the asymptotic variance of the sample mean. The bias of $\hat{\sigma}^2_{b,\mathrm{TBB}}$ is of order $O(1/b^2)$, whereas the untapered block bootstrap results in an estimator of σ^2_∞ of bias $O(1/b)$. The next theorem shows that the order of magnitude of the variance of $\sigma^2_{b,\mathrm{TBB}}$ is $O(b/N)$, as in the untapered case.

THEOREM 2. *Suppose that equation* (3) *and Assumptions* 1–2 *hold and that* $E(|X_t|^{6+\delta})<\infty$, *for some* $\delta>0$, *and* $\sum_{k=1}^{\infty} k^2\alpha_X(k)^{\delta/(6+\delta)}<\infty$. *If* $b\to\infty$ *as* $N\to\infty$ *but with* $b=o(N)$, *then*

$$\operatorname{var}(\hat{\sigma}^2_{b,\mathrm{TBB}})=\Delta\frac{b}{N}+o(b/N),$$

where

$$\Delta=2\sigma^4_\infty\int_{-1}^{1}\frac{(w*w)^2(x)}{(w*w)^2(0)}\,dx.$$

Under the conditions of Theorems 1 and 2 it follows that the mean squared error of estimator $\hat{\sigma}^2_{b,\mathrm{TBB}}$ is of order $O(1/b^4)+O(b/N)$. To minimise this, one should pick b proportional to $N^{1/5}$, in which case the mean squared error of $\hat{\sigma}^2_{b,\mathrm{TBB}}$ is $O(N^{-4/5})$, which compares favourably to the best rate of $O(N^{-2/3})$ achieved by the block bootstrap. Optimal selection of the block size b is discussed in § 3.

We now show the asymptotic validity of the tapered block bootstrap estimator of the sampling distribution of the sample mean $\bar{X}_N$.

THEOREM 3. *Under the combined conditions of Theorems* 1 *and* 2, *we have*

$$\sup_x|\mathrm{pr}^*\{l^{\frac{1}{2}}(\bar{Y}_l^*-\mu^*)\leqslant x\}-\mathrm{pr}\{N^{\frac{1}{2}}(\bar{X}_N-\mu)\leqslant x\}|\to 0, \tag{4}$$

$$\sup_x|\mathrm{pr}^*\{l^{\frac{1}{2}}\bar{Y}_l^*\leqslant x\}-\mathrm{pr}\{N^{\frac{1}{2}}(\bar{X}_N-\mu)\leqslant x\}|\to 0, \tag{5}$$

where the above two convergences hold in probability, and

$$\mu^*:=E^*(\bar{Y}_l^*)=\frac{b^{\frac{1}{2}}}{\|w_b\|_2}\left\{\frac{1}{bQ}\sum_{i=0}^{N-b}\sum_{j=1}^{b}w_b(j)Y_{i+j}\right\}, \tag{6}$$

with $Q=N-b+1$.

Note that equations (4) and (5) tacitly assume that $\sigma^2_\infty\neq 0$; if it so happens that $\sigma^2_\infty=0$, then approximations (4) and (5) will still hold true if we replace the sup-distance with a distance associated with weak convergence such as Prohorov's distance.

Although it is easier to use approximation (5), it is intuitively clear that approximation (4) will be more accurate, because the target distribution of $N^{\frac{1}{2}}(\bar{X}_N-\mu)$ has exactly zero mean; it makes sense, therefore, to have a distribution estimator shifted by an amount that makes it have exactly, as opposed to approximately, zero mean, as first pointed out by Lahiri (1991). In this connection, note that Theorem 3 establishes the asymptotic validity of the tapered block bootstrap for estimation of the sampling distribution of $\bar{X}_N$; however, it does not indicate the rate of convergence of the tapered block bootstrap distribution estimator. It is conjectured that the tapered block bootstrap shares with the untapered block bootstrap the property of higher-order accuracy after studentisation/standardisation and under some additional regularity conditions; see for example Lahiri

(1991) or Götze & Künsch (1996). The proof of this claim requires the development of valid Edgeworth expansions for the tapered block bootstrap distribution. Here, we merely give an informal justification of the superiority of the unstudentised tapered block bootstrap distribution estimator over its block bootstrap counterpart.

Under some moment and mixing conditions, the Berry–Esseen bound for stationary sequences (Tikhomirov, 1980) gives

$$\mathrm{pr}\{N^{\frac{1}{2}}(\bar{X}_N - \mu) \leqslant x\} = \Phi(x/\sigma_\infty) + O(N^{-\frac{1}{2}}), \tag{7}$$

where Φ denotes the standard normal distribution function. Similarly, in the tapered block bootstrap world, a Berry–Esseen bound for the average of the k independent and identically distributed components comprising the bootstrap sample mean $\bar{Y}_l^*$ gives

$$\mathrm{pr}^*\{l^{\frac{1}{2}}(\bar{Y}_l^* - \mu^*) \leqslant x\} = \Phi(x/\hat{\sigma}_{b,\mathrm{TBB}}) + O_p(k^{-\frac{1}{2}});$$

see the proof of Theorem 3. Note that, under some additional regularity conditions, the $O_p(k^{-\frac{1}{2}})$ in the above may be decreased; nevertheless, this is not important for the current discussion since the error present in the term $\Phi(x/\hat{\sigma}_{b,\mathrm{TBB}})$ is in any case at least of order $O_p(k^{-\frac{1}{2}})$. If we recall that $k \sim N/b$ and take b proportional to $N^{1/5}$, see Remark 1 above, an application of the delta-method gives

$$\mathrm{pr}^*\{l^{\frac{1}{2}}(\bar{Y}_l^* - \mu^*) \leqslant x\} = \Phi(x/\sigma_\infty) + O_p\left(\frac{1}{N^{2/5}}\right). \tag{8}$$

Now let $\mathrm{pr}^\diamond\{l^{\frac{1}{2}}(\bar{Y}_l^\diamond - \mu^\diamond) \leqslant x\}$ denote the untapered block bootstrap distribution estimator, obtained by taking $w(t) = 1_{[0,1]}(t)$ and $\mu^\diamond = E^\diamond(\bar{Y}_l^\diamond)$. A similar Berry–Esseen argument, together with taking b proportional to $N^{1/3}$, which is optimal in the untapered case, gives

$$\mathrm{pr}^\diamond\{l^{\frac{1}{2}}(\bar{Y}_l^\diamond - \mu^\diamond) \leqslant x\} = \Phi(x/\sigma_\infty) + O_p\left(\frac{1}{N^{1/3}}\right). \tag{9}$$

Putting equations (7), (8) and (9) together we have that

$$\sup_x |\mathrm{pr}^*\{l^{\frac{1}{2}}(\bar{Y}_l^* - \mu^*) \leqslant x\} - \mathrm{pr}\{N^{\frac{1}{2}}(\bar{X}_N - \mu) \leqslant x\}| = O_p\left(\frac{1}{N^{2/5}}\right),$$

whereas

$$\sup_x |\mathrm{pr}^\diamond\{l^{\frac{1}{2}}(\bar{Y}_l^\diamond - \mu^\diamond) \leqslant x\} - \mathrm{pr}\{N^{\frac{1}{2}}(\bar{X}_N - \mu) \leqslant x\}| = O_p\left(\frac{1}{N^{1/3}}\right).$$

Thus, there is clear evidence in favour of the unstudentised tapered block bootstrap approximation over its block bootstrap counterpart. Note however that, in the above and throughout the paper, the optimal choice for b has the minimisation of the mean squared error of variance estimation as its target. The notion of optimal distribution estimation may, in general, lead to a different optimal choice for b; see for example Hall et al. (1995). We will not pursue this approach further here.

Remark 1. Equation (3) and Assumptions 1–3 have the implication that values towards the block endpoints are downweighted in the tapered block bootstrap procedure. By comparison to the matched-block bootstrap of Carlstein et al. (1998), we see that the tapered block bootstrap effects an effortless matching of the blocks that are to be joined by shrinking the block endpoints to zero. This viewpoint explains our choice of working with the central series $\{Y_t\}$, which is approximately of mean zero, as opposed to the original $\{X_t\}$ series which has mean μ. The tapered block bootstrap shrinkage idea could also be implemented using the uncentred $\{X_t\}$ data, but in this case the block endpoints should be shrunk towards $\bar{X}_N$, instead of zero, and the resulting procedure becomes less transparent.

Remark 2. The applicability of the tapered block bootstrap extends to statistics that are not exactly linear, as in the case of the sample mean, but are approximately linear; for example, the tapered block bootstrap applies to the classes of statistics defined in Examples 2.2 and 2.4 of Künsch (1989), i.e. to functions of linear statistics and M-estimators, for which Künsch's (1989) tapered block jackknife is valid. Nevertheless, it should be stressed that our tapered block bootstrap bears no relation to the random-weighted bootstrap procedure suggested in equation (2.12) of Künsch (1989) whose properties are still unknown. Nevertheless, our tapered block bootstrap is intimately related to Künsch's (1989) tapered block jackknife in the same way that Efron's bootstrap is intimately related to Tukey's jackknife. Note that, under Assumptions 1–3, Künsch's (1989) tapered block jackknife estimator of the variance of the sample mean achieves the desired rate of bias, namely $O(1/b^2)$. It should also be mentioned that estimators with bias of order $O(1/b^2)$ have been known in the literature since the early paper of Parzen (1957); see Politis & Romano (1995) for a literature review, as well as a different approach to the problem of bias reduction. Nevertheless, the tapered block bootstrap goes one step further as it manages to yield an estimator of the whole sampling distribution and not just its variance.

3. PRACTICAL CONCERNS AND FINITE-SAMPLE PERFORMANCE

3·1. *Choosing the tapering window*

Many different choices for tapering windows have been discussed in the literature; see Welch (1967), Brillinger (1975, Ch. 5), Priestley (1981, Ch. 7), Dahlhaus (1985) and Künsch (1989). Zhurbenko (1986, p. 123) mentions some optimality properties of the Kolmogorov–Zhurbenko window but the construction of this window is very cumbersome; in addition, Dahlhaus (1985) shows that the simpler Tukey–Hanning window has quite comparable performance.

From Theorems 1 and 2 it follows that

$$\text{MSE}(\hat{\sigma}^2_{b,\text{TBB}}) \sim \frac{\Gamma^2}{b^4} + \Delta\frac{b}{N}.$$

Hence, the large-sample $\text{MSE}(\hat{\sigma}^2_{b,\text{TBB}})$ is minimised if we choose

$$b_{\text{opt}} = \left(\frac{4\Gamma^2}{\Delta}\right)^{1/5} N^{1/5}, \tag{10}$$

giving

$$\text{MSE}_{\text{opt}} \sim \left(\Gamma^{2/5}\Delta^{4/5}\frac{5}{4^{4/5}}\right) N^{-4/5}. \tag{11}$$

The quantities Γ and Δ depend on the window w and the covariance structure $R(k)$.

The window that minimises the quantity $|\Gamma|\Delta^2$, treating the covariance structure $R(k)$ as fixed, is the one that minimises the quantity

$$|\tilde{w}''(0)|\ \|\tilde{w}\|_2^4, \tag{12}$$

where $\tilde{w}(t) = (w * w)(t)/(w * w)(0)$. With this objective in mind, we focus on a simple class of trapezoidal functions whose typical member is $w_c^{\text{trap}}(t)$, defined by

$$w_c^{\text{trap}}(t) = \begin{cases} t/c, & \text{if } t \in [0, c], \\ 1, & \text{if } t \in [c, 1-c], \\ (1-t)/c, & \text{if } t \in [1-c, 1]; \end{cases}$$

in the above, c is some fixed constant in $(0, \frac{1}{2}]$. After some calculations it follows that we need to choose $c = 0{\cdot}43$ in order to minimise the quantity (12). For such a choice we then compute $(\tilde{w}_{0\cdot43}^{\text{trap}})''(0) = -10{\cdot}9$ and $\|\tilde{w}_{0\cdot43}^{\text{trap}}\|_2^2 = 0{\cdot}5495$. Note that Welch (1967) suggested the window $w_{0\cdot5}^{\text{trap}}$, corresponding to $w_{0\cdot5}^{\text{trap}} * w_{0\cdot5}^{\text{trap}}$ being the well-known Parzen window; however, $w_{0\cdot43}^{\text{trap}}$ has superior mean squared error performance.

3·2. *Choosing the block size b in practice*

Recall that $b_{\text{opt}} = (4\Gamma^2/\Delta)^{1/5} N^{1/5}$, where

$$\Gamma = \sum_{k=-\infty}^{\infty} k^2 R(k) \frac{(w*w)''(0)}{2(w*w)(0)}, \quad \Delta = 2\sigma_\infty^4 \int_{-1}^{1} \frac{(w*w)^2(x)}{(w*w)^2(0)}\, dx.$$

However, the quantities Γ and Δ involve the unknown parameters $\sum_{k=-\infty}^{\infty} k^2 R(k)$ and $\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} R(k)$. To achieve accurate estimation of these two infinite sums, we use the 'flat-top' lag-window of Politis & Romano (1995); that is, we estimate $\sum_{k=-\infty}^{\infty} k^2 R(k)$ by $\sum_{k=-M}^{M} \lambda(k/M) k^2 \hat{R}(k)$, where

$$\hat{R}(k) = N^{-1} \sum_{i=1}^{N-|k|} (X_i - \bar{X}_N)(X_{i+|k|} - \bar{X}_N),$$

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

Similarly, we estimate $\sum_{k=-\infty}^{\infty} R(k)$ by $\sum_{k=-M}^{M} \lambda(k/M)\hat{R}(k)$. Plugging these into Γ and Δ, we obtain estimators $\hat{\Gamma}$ and $\hat{\Delta}$, as well as our estimated block size

$$\hat{b}_{\text{opt}} = \left(\frac{4\hat{\Gamma}^2}{\hat{\Delta}}\right)^{1/5} N^{1/5}. \tag{13}$$

Smoothing with the 'flat-top' lag-window $\lambda(t)$ 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 suggestion $\hat{b}_{\text{opt}}$, we give the following result.

THEOREM 4. *Assume the conditions of Theorem* 2, *as well as Assumption* 3.
(i) *If* $\sum_{k=-\infty}^{\infty} k^4 R(k) < \infty$, *then taking M proportional to* $N^{1/5}$ *we have*

$$\hat{b}_{\text{opt}} = b_{\text{opt}}\{1 + O_p(N^{-2/5})\}.$$

(ii) *If* $R(k)$ *has an exponential decay, then taking* $M \sim A \log N$, *for some appropriate* *constant A, we have*

$$\hat{b}_{\text{opt}} = b_{\text{opt}} \left\{ 1 + O_p \left(\frac{\log^{1/2} N}{N^{1/2}} \right) \right\}. \tag{14}$$

(iii) *If $R(k) = 0$ for $|k| >$ some integer q, then taking $M = 2q$ we have*

$$\hat{b}_{\text{opt}} = b_{\text{opt}} \left\{ 1 + O_p \left(\frac{1}{N^{1/2}} \right) \right\}.$$

Note that, even in the worst case scenario of part (i), the $O_p(N^{-2/5})$ rate in our block size estimator is a significant improvement over the $O_p(N^{-2/7})$ achieved by Bühlmann & Künsch (1999) in their block size estimator for the block bootstrap. However, over and above the improved asymptotics, another important reason for using the 'flat-top' lag-window is that its bandwidth M can be chosen in practice by a simple inspection of the correlogram, i.e. the plot of $\hat{R}(k)$ versus k. In particular, Politis & Romano (1995) suggest looking for the smallest integer $\hat{m}$ after which the correlogram appears negligible, and taking $M = 2\hat{m}$.

Note that $R(k)$ decays exponentially in many interesting cases, in particular the case of stationary ARMA models (Brockwell & Davis, 1991, p. 91) on which we now focus. First note that the 'recipe' $M = 2\hat{m}$, where $\hat{m}$ is obtained by a correlogram inspection, does not contradict the recommendation $M \sim A \log N$ offered in Theorem 4(ii). On the contrary, the $M = 2\hat{m}$ recipe should be viewed as an empirical way of pinpointing the constant A in $M \sim A \log N$. To see this, consider the simple AR(1) model, $X_t = \rho X_{t-1} + Z_t$, for $t \in \mathbb{Z}$, where $\{Z_t\}$ are independent and identically distributed $N(0, \sigma^2)$ and $|\rho| < 1$. It is well known that in this case the autocovariance $R(k)$ satisfies $R(k) = R(0)\rho^k$, for $k \geqslant 0$. Let the autocorrelations be defined as $\rho_X(k) := R(k)/R(0)$; then the estimated autocorrelations are given by $\hat{\rho}_X(k) := \hat{R}(k)/\hat{R}(0) = \rho^k + O_p(N^{-1/2})$. To say that $\hat{R}(k) \simeq 0$ for $k > \hat{m}$, means that $\hat{\rho}_X(\hat{m}+1)$ is not significantly different from zero; this in turn means that $-c_1 N^{-1/2} < \hat{\rho}_X(\hat{m}+1) < c_2 N^{-1/2}$ for some positive constants c_1, c_2. It therefore follows that, with probability tending to one,

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

for some positive constants A_1 and A_2.

Perhaps the most attractive feature of the $M = 2\hat{m}$ recipe is its adaptivity to different correlation structures. Arguments similar to the above show that, if $R(k)$ has a polynomial, as opposed to exponential, decay, then $\hat{m}$ itself grows at a polynomial rate, as advisable in that case; see Theorem 4(i). In addition, if $R(k) = 0$ for $|k| > q$, but $R(q) \neq 0$, then it is easy to see that $\hat{m} \to q$ in probability; this corresponds to the interesting case of an MA(q) model which is consistent with the set-up of Theorem 4(iii). Thus, all cases of Theorem 4 are well addressed, and the recipe $M = 2\hat{m}$ is an omnibus rule-of-thumb that automatically gives good bandwidth choices without having to prespecify the correlation structure.

3·3. *Finite-sample performance*

We focus on the sample mean case, using the window $w_{0\cdot43}^{\text{trap}}$ discussed in § 3·1. This window choice implies

$$\Gamma = -5{\cdot}45 \sum_{k=-\infty}^{\infty} k^2 R(k), \quad \Delta = 1{\cdot}1\sigma_\infty^4.$$

The simulation covered many different time series models; since the results were qualitatively similar in all cases, we present only the results for the two nonlinear time series models below.

Model 1. In the nonlinear autoregressive model, NAR,

$$X_t = 0{\cdot}6 \sin(X_{t-1}) + Z_t,$$

for $t \in \mathbb{Z}$, where $\{Z_t\}$ are independent and identically distributed $N(0, 1)$.

Model 2. In the exponential autoregressive model, EXPAR,

$$X_t = \{0{\cdot}8 - 1{\cdot}1 \exp(-50X_{t-1}^2)\}X_{t-1} + 0{\cdot}1Z_t,$$

for $t \in \mathbb{Z}$, where $\{Z_t\}$ are independent and identically distributed $N(0, 1)$.

Model 1 is a simple nonlinear autoregression, while Model 2 has been previously used by Auestad & Tjøstheim (1990).

The simulations were implemented in the following way, 1000 time series being generated using each choice of model and sample size N. For each series, the tapered block bootstrap and block bootstrap were performed for a variety of block sizes; for each block size, we generated 1000 tapered block bootstrap, and/or block bootstrap, pseudo-series from which estimation of σ_∞^2 was possible; in addition, the construction of 95% equal-tailed confidence intervals for μ was achieved using the tapered block bootstrap, and/or block bootstrap, distribution as in Theorem 3, and in particular using the centred approximation (4). The empirical mean squared error of estimation of σ_∞^2, as well as the empirical coverage of the nominally 95% confidence intervals, were assessed over the 1000 true series, and are presented in Figs 1 and 2. Note that the pointwise standard errors associated with the empirical coverages shown in Fig. 1 is about 0·01. Regarding Fig. 2, the pointwise standard errors vary with block size; however, it suffices to note that, except near the points where curves meet, the deviations of one curve from the other in Fig. 2 are typically of the order of several standard errors.

It is apparent that a significant improvement is offered by the tapered block bootstrap, over the block bootstrap, in both aspects. For example, for the same block size, it is almost invariably better to use the tapered block bootstrap in terms of both coverage level and variance estimation. Furthermore, Figs 1 and 2 show that the tapered block bootstrap at or near its empirical block size performs much better than the untapered block bootstrap at or near its empirical optimal block size; for example, the reduction in mean squared error of variance estimation is of the order of 30%.

Of course, the true optimal block size will not be known in practice, and we now investigate the performance of our suggestion (13) for $\hat{b}_{\text{opt}}$. To provide a fair comparison of the two methods, we use a similar procedure based on plug-in estimates with flat-top kernels to estimate the optimal block size for the block bootstrap.

We report two types of finding, namely the performance of estimator $\hat{b}_{\text{opt}}$, that is seeking to substantiate the claims of Theorem 4, and, even more importantly, the comparative performance of the two variance estimators in conjunction with our estimated optimal block sizes.

The simulation was based on the MA(2) model, $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$, for $t \in \mathbb{Z}$, where $\{Z_t\}$ are independent and identically distributed $N(0, 1)$. Despite the simplicity of the MA(2) model, note that, by letting θ_1 and θ_2 take different values from the set $\{-1, -0{\cdot}6, -0{\cdot}3, 0{\cdot}1, 0{\cdot}4, 0{\cdot}7, 1\}$, a wide range of covariance shapes is obtained, from near

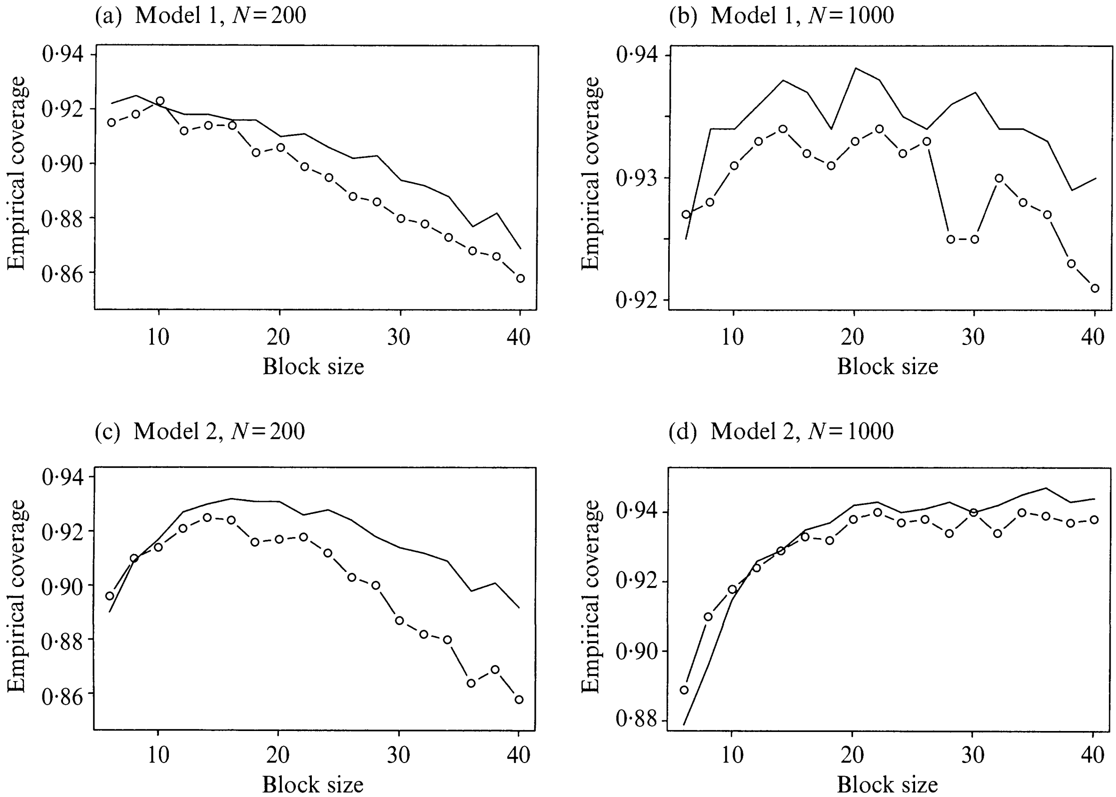

Fig. 1. Empirical coverage, as a function of the block size b, of 95% equal-tailed confidence intervals obtained from the centred approximation (4) for the nonlinear autoregressive model, Model 1, and the exponential autoregressive model, Model 2, for two sample sizes $N=200$ and $N=1000$; —— denotes tapered block bootstrap, and —○— denotes bootstrap.

independence, $\theta_1=0{\cdot}1$ and $\theta_2=0{\cdot}1$, to strong positive dependence, $\theta_1=1$ and $\theta_2=1$, or strong negative dependence, $\theta_1=0{\cdot}1$ and $\theta_2=-1$, or $\theta_1=-1$ and $\theta_2=0{\cdot}1$.

For each of the 49 model parameter combinations, 1200 time series of length N were generated, with $N=200$. To interpret the entries of Table 1 properly, note that the true value of σ_∞^2 varies immensely with changes in θ_1 and θ_2; for example, $\sigma_\infty^2=1{\cdot}44$ in the nearly independent case of $\theta_1=0{\cdot}1$ and $\theta_2=0{\cdot}1$, $\sigma_\infty^2=9$ in the strong positive dependent case of $\theta_1=1$ and $\theta_2=1$, and $\sigma_\infty^2=0{\cdot}01$ in the strong negative dependent cases of $\theta_1=0{\cdot}1$ and $\theta_2=-1$, or $\theta_1=-0{\cdot}3$ and $\theta_2=-0{\cdot}6$, or $\theta_1=-0{\cdot}6$ and $\theta_2=-0{\cdot}3$, or $\theta_1=-1$ and $\theta_2=0{\cdot}1$.

Table 1 shows the empirical mean squared errors of $\hat{\sigma}^2_{b_{\text{opt}},\text{TBB}}$ divided by the true value of σ_∞^4; this is equivalent to looking at the mean squared error of $\hat{\sigma}^2_{b_{\text{opt}},\text{TBB}}/\sigma_\infty^2$. Inspection of Table 1 pinpoints the cases of strong negative dependence as the 'difficult' ones with respect to this 'relative' mean squared error measure, relative to σ_∞^4; this is hardly surprising since we are in effect dividing by something close to zero.

Another way of obtaining a relative mean squared error measure is to look at the ratio of the mean squared error of the tapered block bootstrap over that of the block bootstrap. To achieve a fair comparison, the block sizes of both tapered block bootstrap and block boostrap were estimated using the ideas of § 3·2, that is using the flat-top kernels; Table 2 shows this ratio. Although some entries in Table 2 are bigger than one, the overwhelming

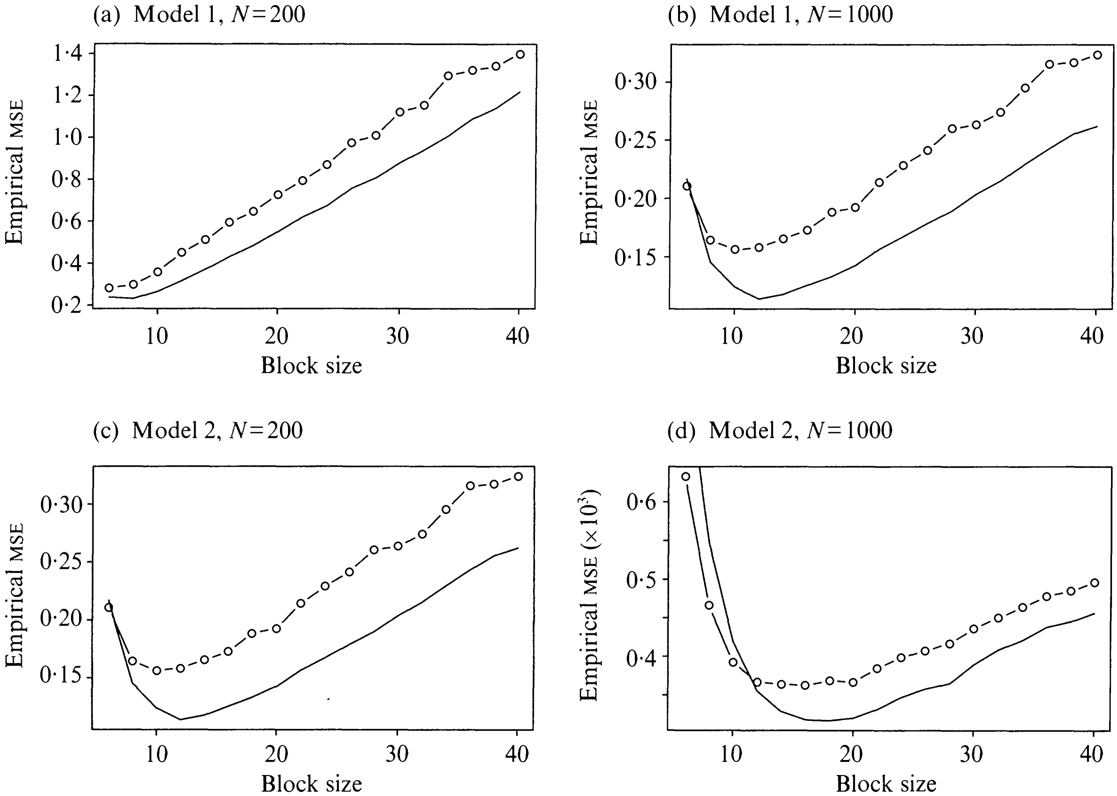

Fig. 2. Empirical mean squared errors, as a function of the block size b, of the tapered block bootstrap and block bootstrap estimators of σ_∞^2 for the nonlinear autoregressive model, Model 1, and the exponential autoregressive model, Model 2, for two sample sizes $N = 200$ and $N = 1000$; —— denotes tapered block bootstrap, and —○— denotes block bootstrap.

Table 1. *Ratio of empirical mean squared error for the tapered block bootstrap divided by* σ_∞^4 *with estimated block size, for different cases of* MA(2) *models and* $N = 200$

θ_1	$\theta_2 = -1·0$	$\theta_2 = -0·6$	$\theta_2 = -0·3$	$\theta_2 = 0·1$	$\theta_2 = 0·4$	$\theta_2 = 0·7$	$\theta_2 = 1·0$
−1·0	0·157	0·243	0·582	268·449	0·675	0·086	0·088
−0·6	0·527	3·361	10·268	0·362	0·067	0·078	0·074
−0·3	1·428	20·324	0·269	0·059	0·065	0·063	0·071
0·1	242·668	0·243	0·105	0·053	0·064	0·065	0·062
0·4	0·975	0·125	0·068	0·064	0·065	0·066	0·071
0·7	0·230	0·100	0·052	0·066	0·066	0·072	0·070
1·0	0·174	0·080	0·049	0·054	0·068	0·065	0·067

majority of the entries are less than one and often very close to zero, thereby giving empirical evidence of the tapered block bootstrap's superiority over the block bootstrap. In particular, the average, across models, gain from using the tapered version exceeds 30%, as the average of the entries of Table 2 is about 0·69. Note also that the tapered block bootstrap vastly outperforms the block bootstrap in the aforementioned 'difficult' cases of strong negative dependence.

Table 3 shows the finite-sample performance of $\hat{b}_{\text{opt}}$. Except for the 'difficult' cases of

Table 2. *Ratio of empirical mean squared error for the tapered block bootstrap divided by the corresponding empirical mean squared error for the untapered block bootstrap with estimated block size, for different cases of* MA(2) *models and* $N = 200$

θ_1	$\theta_2 = -1{\cdot}0$	$\theta_2 = -0{\cdot}6$	$\theta_2 = -0{\cdot}3$	$\theta_2 = 0{\cdot}1$	$\theta_2 = 0{\cdot}4$	$\theta_2 = 0{\cdot}7$	$\theta_2 = 1{\cdot}0$
−1·0	0·801	0·514	0·012	<0·000	0·154	0·187	0·130
−0·6	0·001	0·003	0·002	1·922	0·175	0·949	1·196
−0·3	0·003	0·004	0·052	0·606	1·245	0·971	0·961
0·1	0·001	0·057	0·861	1·084	0·944	0·929	0·915
0·4	0·002	0·795	1·285	0·893	0·962	0·964	0·956
0·7	0·206	0·936	1·260	0·895	0·995	0·977	0·971
1·0	0·760	1·110	1·131	0·833	0·892	0·925	0·937

strong negative dependence, the bias of $\hat{b}_{\text{opt}}$ is quite small. The nonnegligible bias in those 'difficult' cases is to be expected: if σ_∞^2 were exactly zero, then the true value b_{opt} would be infinity, and estimators of infinity are typically negatively biased!

Table 3. *The empirical mean of the tapered block bootstrap block size estimator* $\hat{b}_{\text{opt}}$, *with the standard deviation of* $\hat{b}_{\text{opt}}$ *in parentheses and the true value of* b_{opt} *in square brackets, for different cases of* MA(2) *models and* $N = 200$

θ_1	$\theta_2 = -1{\cdot}0$	$\theta_2 = -0{\cdot}6$	$\theta_2 = -0{\cdot}3$	$\theta_2 = 0{\cdot}1$	$\theta_2 = 0{\cdot}4$	$\theta_2 = 0{\cdot}7$	$\theta_2 = 1{\cdot}0$
−1·0	18·10	24·42	37·59	37·80	14·30	12·80	12·41
	[17] (7·76)	[23] (14·44)	[33] (29·24)	[76] (47·35)	[11] (5·51)	[14] (2·82)	[13] (1·92)
−0·6	30·77	48·41	47·26	11·32	9·73	10·92	10·86
	[26] (28·29)	[51] (40·13)	[103] (33·19)	[10] (5·19)	[10] (2·34)	[11] (1·32)	[11] (1·19)
−0·3	44·61	47·57	26·89	7·04	8·92	9·91	9·91
	[45] (35·97)	[87] (32·98)	[23] (21·10)	[4] (2·69)	[10] (1·89)	[10] (1·29)	[10] (1·27)
0·1	49·43	28·75	12·26	6·68	8·00	8·66	8·99
	[105] (48·75)	[24] (21·88)	[13] (4·11)	[7] (2·23)	[9] (1·92)	[9] (1·77)	[9] (1·63)
0·4	43·08	16·84	9·42	6·80	7·42	8·10	8·33
	[36] (46·30)	[16] (5·07)	[9] (3·64)	[7] (2·25)	[8] (2·09)	[9] (1·99)	[9] (1·89)
0·7	25·99	12·54	7·75	6·57	7·13	7·73	7·79
	[22] (16·98)	[12] (4·27)	[7] (3·14)	[6] (2·24)	[8] (2·17)	[8] (2·01)	[9] (2·01)
1·0	17·57	10·24	7·27	6·53	6·96	7·30	7·57
	[17] (6·85)	[10] (3·85)	[5] (2·92)	[7] (2·24)	[7] (2·20)	[8] (2·07)	[8] (2·02)

The MA(2) simulation experiment was repeated with sample size $N = 800$. The reduction of the mean squared error of the tapered block bootstrap resulting from the increased sample size was quite dramatic, confirming the fast rate of convergence of $\hat{\sigma}^2_{\hat{b}_{\text{opt}},\text{TBB}}$. Both bias and standard deviation of $\hat{b}_{\text{opt}}$ were significantly decreased, roughly by 50%, as compared to the $N = 200$ case, thus substantiating the claim of Theorem 4(iii).

ACKNOWLEDGEMENT

Many thanks are due to the editor and a referee for their detailed comments.

Appendix

Proofs

Proof of Theorems 1 *and* 2. The easiest way to prove Theorems 1 and 2 is to relate $\hat{\sigma}^2_{b,\text{TBB}}$ to Künsch's (1989) tapered jackknife; recall that Künsch's σ^2_{jack} is an estimator of σ^2_∞/N, and therefore the comparison will be between $\hat{\sigma}^2_{b,\text{TBB}}$ and $N\sigma^2_{\text{jack}}$. The following lemma extends Künsch's (1989) Theorem 3.4 to the case of general windows.

Lemma A1. *Assume equation* (3) *and Assumptions* 1–2. *Then*

$$\hat{\sigma}^2_{b,\text{TBB}} \equiv \text{var}^*(l^{1/2}\bar{Y}^*_l) = N\sigma^2_{\text{jack}},$$

where σ^2_{jack} *is defined in Künsch* (1989, p. 1220).

Proof of Lemma A1. We compute the first two tapered block bootstrap moments. Recall that

$$\mu^* = E^*(\bar{Y}^*_l) = \frac{b^{1/2}}{\|w_b\|_2}\left\{\frac{1}{bQ}\sum_{i=0}^{N-b}\sum_{j=1}^{b} w_b(j)Y_{i+j}\right\},$$

where $Q = N - b + 1$. Similarly, we have,

$$\text{var}^*(l^{1/2}\bar{Y}^*_l) = \frac{1}{Q}\sum_{i=0}^{N-b}\frac{1}{b}\left\{\sum_{j=1}^{b} w_b(j)\frac{b^{1/2}}{\|w_b\|_2}Y_{i+j} - \frac{b^{1/2}}{\|w_b\|_2}C\right\}^2,$$

where

$$C = \frac{1}{Q}\sum_{i=0}^{N-b}\sum_{j=1}^{b} w_b(j)Y_{i+j}.$$

Thus,

$$\text{var}^*(l^{1/2}\bar{Y}^*_l) = \frac{1}{Q}\frac{1}{\|w_b\|_2^2}\sum_{i=0}^{N-b}\left\{\sum_{j=1}^{b} w_b(j)Y_{i+j} - C\right\}^2,$$

Note that, since $w_b(j) = 0$ if $j \notin \{1, 2, \ldots, b\}$, we can write C as

$$C = \sum_{t=1}^{N}\sum_{j=0}^{N-b} w_b(t-j)\frac{Y_t}{Q}.$$

Hence,

$$\begin{aligned}
\text{var}^*(l^{1/2}\bar{Y}^*_l) &= \frac{1}{Q}\frac{1}{\|w_b\|_2^2}\sum_{i=0}^{N-b}\left\{\sum_{j=1}^{b} w_b(j)(X_{i+j} - \bar{X}_N) - \sum_{t=1}^{N}\sum_{j=0}^{N-b} w_b(t-j)\frac{X_t - \bar{X}_N}{Q}\right\}^2 \\
&= \frac{1}{Q}\frac{1}{\|w_b\|_2^2}\sum_{i=0}^{N-b}\left\{\sum_{j=1}^{b} w_b(j)X_{i+j} - \|w_b\|_1\bar{X}_N + \|w_b\|_1\bar{X}_N - \sum_{t=1}^{N}\sum_{j=0}^{N-b} w_b(t-j)\frac{X_t}{Q}\right\}^2 \\
&= \frac{1}{Q}\frac{1}{\|w_b\|_2^2}\sum_{i=0}^{N-b}\left\{\sum_{j=1}^{b} w_b(j)X_{i+j} - \sum_{t=1}^{N}\sum_{j=0}^{N-b} w_b(t-j)\frac{X_t}{Q}\right\}^2 = N\sigma^2_{\text{jack}},
\end{aligned}$$

as given by Künsch (1989). □

Theorems 1 and 2 follow respectively from Theorem 3.2(ii) and Theorem 3.3 of Künsch (1989) in conjunction with our Lemma A1. □

Proof of Theorem 3. First note that the assumed conditions imply that $E(|X_1|^{2+\delta}) < \infty$ and $\sum_{k=1}^{\infty}\alpha_X^{\delta/(2+\delta)}(k) < \infty$, which, as stated in § 1, are sufficient for the Central Limit Theorem:

$$\sup_x |\text{pr}\{N^{1/2}(\bar{X}_N - \mu) \leqslant x\} - \Phi(x/\sigma_\infty)| \to 0.$$

Therefore, to prove the theorem, we just need to show that the two tapered block bootstrap distributions, in (4) and in (5), are appropriately close to $\Phi(x/\sigma_\infty)$.

Recall that $\bar{Y}_l^*$ is an average of k independent and identically distributed components, that are taken randomly with replacement from the set

$$\left\{\frac{1}{b}\sum_{j=1}^{b} w_b(j)\frac{b^{1/2}}{\|w_b\|_2} Y_{i+j} : i=0,1,\ldots,N-b\right\}.$$

We can estimate the common third bootstrap moment of those independent and identically distributed components by

$$\frac{1}{Q}\sum_{i=0}^{N-b}\left\{\frac{1}{b}\sum_{j=1}^{b} w_b(j)\frac{b^{1/2}}{\|w_b\|_2} Y_j\right\}^3 = O\left\{\frac{1}{Q}\sum_{i=0}^{N-b}\left(\frac{1}{b}\sum_{j=1}^{b} Y_j\right)^3\right\}.$$

In the above we used the facts that $w_b(j)\leqslant 1$ and $b^{1/2}/\|w_b\|_2 = O(1)$; the latter follows because equation (3) implies that $\|w_b\|_2^2/b \to \|w\|_2^2$, with $\|w\|_2^2 = \int_0^1 w^2(t)\,dt > 0$ by Assumption 1.

However, it is well known that, under the assumed conditions,

$$\frac{1}{Q}\sum_{i=0}^{N-b}\left(\frac{1}{b}\sum_{j=1}^{b} Y_j\right)^3 = O_p(1);$$

see Liu & Singh (1992) or Politis & Romano (1992a). Therefore, the third bootstrap moment of the independent and identically distributed components is bounded in probability; hence, in a set with probability tending to one, a Berry–Esseen bound applies for the average of the k components comprising the bootstrap sample mean $\bar{Y}_l^*$, giving

$$\mathrm{pr}^*\left\{l^{1/2}\frac{\bar{Y}_l^* - E^*(\bar{Y}_l^*)}{\hat{\sigma}_{b,\mathrm{TBB}}} \leqslant x\right\} = \Phi(x) + O_p(k^{-1/2}).$$

Note that, by Theorems 1 and 2, $\hat{\sigma}_{b,\mathrm{TBB}} \to \sigma_\infty$ in probability; the observation that $k \sim N/b \to \infty$ completes the proof of (4). To show (5), note that, by (6),

$$E^*(\bar{Y}_l^*) = \frac{C}{b^{1/2}\|w_b\|_2}.$$

We now wish to bound the quantity C that was defined in the proof of Lemma A1:

$$\begin{aligned}
C &= \sum_{t=1}^{N}\sum_{j=0}^{N-b} w_b(t-j)\frac{Y_t}{Q} \\
&= O(b)\frac{Y_1}{Q} + O(b)\frac{Y_2}{Q} + \ldots O(b)\frac{Y_b}{Q} + \sum_{t=b+1}^{N-b}\sum_{j=0}^{N-b} w_b(t-j)\frac{Y_t}{Q} \\
&\quad + O(b)\frac{Y_{N-b+1}}{Q} + O(b)\frac{Y_{N-b+2}}{Q} + \ldots O(b)\frac{Y_N}{Q} \\
&= \left(\frac{b}{N}\right)\sum_{t=1}^{b} Y_t + \|w_b\|_1 \frac{1}{Q}\sum_{t=b+1}^{N-b} Y_t + O\left(\frac{b}{N}\right)\sum_{t=b+1}^{N} Y_t.
\end{aligned}$$

Since by construction $N^{-1}\sum_{t=1}^{N} Y_t = 0$, it follows that

$$C = O_p\left(\frac{b^2}{N}\right) + \|w_b\|_1 O_p\left(\frac{b}{N}\right) + O_p\left(\frac{b^2}{N}\right) = O_p\left(\frac{b^2}{N}\right)$$

since $\|w_b\|_1 \leqslant b$.

Thus,

$$E^*(\bar{Y}_l^*) = \frac{C}{b^{1/2}\|w_b\|_2} = O_p\left(\frac{b}{N}\right),$$

and (5) follows from (4), Slutsky's theorem and the condition $b = o(N^{1/3})$. □

Proof of Theorem 4. 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

$$\sum_{k=-M}^{M} \lambda(k/M)\hat{R}(k) = \sum_{k=-\infty}^{\infty} R(k) + O_p(\log^{1/2} N/N^{1/2});$$

see Politis & Romano (1995). In the same vein, we can similarly show that

$$\sum_{k=-M}^{M} \lambda(k/M)k^2\hat{R}(k) = \sum_{k=-\infty}^{\infty} k^2R(k) + O_p(\log^{1/2} N/N^{1/2}).$$

An application of the delta method completes the proof. □

References

Auestad, B. & Tjøstheim, D. (1990). Identification of nonlinear time series: first order characterization and order determination. *Biometrika* **77**, 669–87.

Brillinger, D. R. (1975). *Time Series, Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

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

Bühlmann, P. & Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Comp. 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**, 1171–9.

Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T. & Künsch, H. (1998). Matched-block bootstrap for dependent data. *Bernoulli* **4**, 305–28.

Dahlhaus, R. (1985). On a spectral density estimate obtained by averaging periodograms. *J. Appl. Prob.* **22**, 598–610.

Dahlhaus, R. (1990). Nonparametric high resolution spectral estimation. *Prob. Theory Rel. Fields* **85**, 147–80.

Davison, A. & Hall, P. (1993). On studentising and blocking methods for implementing the bootstrap for dependent data. *Aust. J. Statist.* **35**, 215–24.

Götze, F. & Künsch, H. (1996). Second order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* **24**, 1914–33.

Hall., P., Horowitz, J. L. & Jing, B. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–74.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217–41.

Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statist. Prob. Lett.* **11**, 335–41.

Lahiri, S. N. (1999). 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 *Exploring the Limits of Bootstrap*, Ed. R. LePage and L. Billard, pp. 225–48. New York: John Wiley.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28**, 329–48.

Politis, D. N. & Romano, J. P. (1992a). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* **20**, 1985–2007.

Politis, D. N. & Romano, J. P. (1992b). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*, Ed. R. LePage and L. Billard, pp. 263–70. New York: John Wiley.

Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *J. Am. Statist. Assoc.* **89**, 1303–13.

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

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

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

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

Tikhomirov, A. N. (1980). On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory Prob. Applic.* **25**, 790–809.

Welch, P. D. (1967). The use of the Fast Fourier Transform for estimation of spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–3.

Zhurbenko, I. G. (1986). *The Spectral Analysis of Time Series*. Amsterdam: North-Holland.

[*Received May* 2000. *Revised January* 2001]