Stochastic Processes and their Applications 45 (1993) 155-167
North-Holland

On the sample variance of linear statistics derived from mixing sequences

Dimitris N. Politis and Joseph P. Romano

Department of Statistics, Purdue University, West Lafayette, IN, USA, and Department of Statistics, Stanford University, Stanford, CA, USA

Received 15 November 1990
Revised 9 September 1991 and 20 January 1992

In a sample $X_1, \ldots, X_N$, independently and identically distributed with distribution F, a linear statistic $\bar{T} = (1/N)\sum_{i=1}^{N} T_i$ can be defined, where $T_i = \phi(X_i)$, and $\phi(\cdot)$ is some function. For this statistic, a 'natural' nonparametric variance estimator is the sample variance $(1/N)\sum_{i=1}^{N}(T_i - \bar{T})^2$, the denominator $N-1$ often being used instead of N.

However, if the sample is stationary but weakly dependent, the same estimator would not work, since it fails to take into account the covariances among the T_i's. Moreover, in many time series problems, the objective is to estimate a parameter of the Mth dimensional marginal, and not just of the first-dimensional marginal distribution. Thus, the linear statistic in this case must be of the form $T(X_1, \ldots, X_N) = (1/(N-M+1))\sum_{i=1}^{N-M+1} T_i$, where $T_i = \phi_M(X_i, \ldots, X_{i+M-1})$, and $\phi_M(\cdot)$ is now a function of a whole block of observations.

In the present report, we formulate the nonparametric variance estimator corresponding to a sample variance of the linear statistic $T(X_1, \ldots, X_N)$. The proposed estimator depends on a design parameter b that tends to infinity as the sample size N increases. The optimal rate at which b should tend to infinity is found that minimizes the asymptotic order of the Mean Squared Error in estimation. Special emphasis is given to the case where M tends to infinity as well as N, in which case a general version of the linear statistic is introduced that estimates a parameter of the whole (infinite-dimensional) joint distribution of the sequence $\{X_n, n \in \mathbb{Z}\}$.

mixing sequences * linear statistics * nonparametric variance estimation

1. Introduction

Let $\{X_n, n \in \mathbb{Z}\}$ be a strictly stationary and weakly dependent multivariate time series, where X_1 takes values in $\mathbb{R}^d$. The degree of dependence is quantified by the various mixing coefficients (cf. Roussas and Ioannides, 1987). We will particularly make use of Rosenblatt's α-mixing or strong-mixing coefficient which is defined as follows:

$$\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|, \tag{1}$$

where $A \in \mathscr{F}_{-\infty}^{0}$, $B \in \mathscr{F}_k^{\infty}$ are events in the σ-algebras generated by $\{X_n, n \leq 0\}$ and $\{X_n, n \geq k\}$ respectively.

Correspondence to: Drs. D.N. Politis and J.P. Romano, Department of Statistics, Purdue University, West Lafayette, IN 47907, USA.

Suppose $\mu \in \mathbb{R}^D$ is a parameter of the whole (infinite-dimensional) joint distribution of sequence $\{X_n, n \in \mathbb{Z}\}$. As a matter of course, parameters of a finite-dimensional marginal are also covered in this general framework. We will now describe a way to obtain consistent estimators of μ in the form of an average of functions defined on the observations, i.e., in the form of a general linear statistic.

For each $N = 1, 2, \ldots$ let $B_{i,M,L}$ be the block of M consecutive observations starting from $(i-1)L+1$, i.e., the subseries $X_{(i-1)L+1}, \ldots, X_{(i-1)L+M}$, where M, L are integer functions of N. Note that $B_{i,M,L}$ for $i = \ldots, -1, 0, +1, \ldots$ can be gotten from $\{X_n, n \in \mathbb{Z}\}$ by a 'window' of width M which is 'moving' at lags L at a time.

Now define $T_{i,M,L} = \phi_M(B_{i,M,L})$, where $\phi_M : \mathbb{R}^{dM} \to \mathbb{R}^D$ is some function. So for fixed N, the $T_{i,M,L}$ for $i \in \mathbb{Z}$ constitute a strictly stationary sequence. In practice we would observe a segment $X_1, \ldots, X_N$ from the time series $\{X_n\}$, which would permit us to compute $T_{i,M,L}$ for $i = 1, \ldots, Q$ only, where $Q = [(N-M)/L] + 1$ and $[\cdot]$ is the integer part function. We can think of the $T_{i,M,L}$'s as a triangular array whose Nth row consists of $T_{i,M,L}$, $i = 1, \ldots, Q$.

Finally we are in a position to define the *general linear statistic*:

$$\bar{T}_N = \frac{1}{Q} \sum_{i=1}^{Q} T_{i,M,L}. \tag{2}$$

Under broad regularity conditions $\bar{T}_N$ is a consistent estimator of μ. Loosely stated, these regularity conditions consist of a weak dependence structure (allowing the variance of $\bar{T}_N$ to tend to zero as $N \to \infty$), and a condition of unbiasedness or asymptotic unbiasedness of $T_{1,M,L}$, i.e., $ET_{1,M,L} = \mu$, or $ET_{1,M,L} \to \mu$ as $M \to \infty$. We will mention here some examples of time series statistics that can fit in this framework. For the examples assume X_n is univariate, i.e., $d = 1$.

(I) The sample mean: $\bar{X}_N = (1/N) \sum_{i=1}^{N} X_i$. Just take $M = L = 1$ and ϕ_M to be the identity function. Here the parameter of interest is $\mu = EX_1$.

(II) The (unbiased) sample autocovariance at lag s: $(1/(N-s)) \sum_{i=1}^{N-s} X_i X_{i+s}$. Take $L = 1$, $M = s+1$, $\phi_M(x_1, \ldots, x_M) = x_1 x_M$, and $\mu = EX_1 X_{1+s}$.

(III) The lag-window spectral density estimator, where we take

$$\phi_M(B_{i,M,L}) = \frac{1}{2\pi M} \left| \sum_{t=L(i-1)+1}^{L(i-1)+M} W_t^{(M)} X_t \, \mathrm{e}^{-\mathrm{j}tw} \right|^2, \tag{3}$$

i.e., $T_{i,M,L}(w)$ is the periodogram of block $B_{i,M,L}$ of data 'tapered' by the function $W_t^{(M)}$, and evaluated at the point $w \in [0, 2\pi]$. (Note that the symbol j denotes the unit of imaginary numbers $\sqrt{-1}$, in order to avoid confusion with i, the block count.) In this example, $\mu = f(w)$, the spectral density function evaluated at the point w.

For multivariate time series we can similarly use our formulation of the general linear statistic to define the sample cross-covariance and cross-spectrum estimators.

In this paper we address the question of the nonparametric estimation of the asymptotic variance-covariance matrix of $\sqrt{Q}\,\bar{T}_N$, as $N \to \infty$. Having such a variance estimate is required in order to get confidence regions for μ via the Central Limit Theorem.

To fix ideas, let us momentarily go back to the sample mean case, as in example (I). Let $X_1, \ldots, X_N$ be observations from the univariate stationary sequence $\{X_n\}$, with mean $\mu = EX_1$. Then

$$\sigma_N^2 \equiv \operatorname{Var}(\sqrt{N}\,\bar{X}_N) = \operatorname{Var}(X_1) + 2\sum_{i=1}^{N}\left(1 - \frac{i}{N}\right)\operatorname{Cov}(X_1, X_{1+i}). \tag{4}$$

Assuming a sufficiently weak dependence structure such that $\sum_{i=1}^{\infty} |\operatorname{Cov}(X_1, X_{1+i})| < \infty$ (namely that the variance of $\bar{X}_N$ is of order $1/N$, as it is under independence), it is seen that σ_∞^2 exists. The question then is how to estimate σ_N^2 or σ_∞^2.

One way is to estimate the covariances $\operatorname{Cov}(X_1, X_{1+i})$ and to plug them in the formula above. However, this simple procedure is not consistent. The problem that arises is that for lags close to N, the covariance estimates are unreliable because they are based on a progressively smaller sample size. This is a common problem in the literature concerning spectral estimation (cf. Grenander and Rosenblatt, 1957; Hannan, 1970; Brillinger, 1975; Priestley, 1981), where 'tapering' the estimated covariances is proposed, before plugging into the summation formula. Since estimating the variance of the sample mean is essentially estimating the spectral density at frequency zero, this proposal would work here as well.

The problem itself however seems to suggest another way to look at this situation. From the sample of size N we can only hope to estimate well $\operatorname{Cov}(X_1, X_{1+i})$ for $i = 1, \ldots, b$, where $b \ll N$. It follows that we can only hope to estimate well σ_b^2, and not σ_N^2. But there is a natural way to estimate σ_b^2 from $X_1, \ldots, X_N$, namely to look at the sample variability of $(1/\sqrt{b})\sum_{t=i}^{i+b-1} X_t$, for $i = 1, \ldots, N-b+1$. This idea leads to the natural estimate

$$\hat{\sigma}_{b/N}^2 = \frac{1}{N-b+1}\sum_{i=1}^{N-b+1}\left(\frac{1}{\sqrt{b}}\sum_{t=i}^{i+b-1} X_t - \sqrt{b}\,\bar{X}_N\right)^2, \tag{5}$$

which can be viewed as a 'sample variance' in this case.

It is easy to see that if b is fixed, then $\hat{\sigma}_{b/N}^2 \to \sigma_b^2 \equiv \operatorname{Var}((1/\sqrt{b})\sum_{t=1}^{b} X_t)$, as $N \to \infty$, with probability one. Similarly, if $b \to \infty$ as well but at a smaller rate than N (allowing $N - b \to \infty$), it is seen that $\hat{\sigma}_{b/N}^2 \to \sigma_\infty^2$, as $N \to \infty$, with probability one.

The question remains: Why use $\hat{\sigma}_{b/N}^2$ as an estimate of the variance of $\sqrt{N}\,\bar{X}_N$? The justification is that σ_b^2 is a closer approximation to σ_N^2 than $\sigma_1^2 = \operatorname{Var}(X_1)$ is. The approximation becomes better and better as b and N tend to infinity, because from $\sigma_b^2 \to \sigma_\infty^2$ and $\sigma_N^2 \to \sigma_\infty^2$, it is seen that $|\sigma_N^2 - \sigma_b^2| \to 0$.

The same idea extends immediately to the context of a general linear statistic. As before, from the sample of the $T_{i,M,L}$'s, for $i = 1, \ldots, Q$, we can only hope to estimate well the variance of $(1/\sqrt{b})\sum_{i=k}^{k+b-1} T_{i,M,L}$, instead of the variance of $(1/\sqrt{Q})\sum_{i=1}^{Q} T_{i,M,L}$. This leads to the natural nonparametric estimate

$$\hat{V}_{b/N} \equiv \frac{1}{Q-b+1}\sum_{i=1}^{Q-b+1}\left(\frac{1}{\sqrt{b}}\sum_{j=i}^{i+b-1} T_{j,M,L} - \sqrt{b}\,\bar{T}_N\right)^2, \tag{6}$$

which can be termed a *blocked sample variance* and will be the subject of our subsequent investigation.

In the papers by Carlstein (1986a, 1986b, 1988, 1989), a quite similar variance estimator was introduced for a general (not necessarily linear) statistic based on subseries values, and consistency and asymptotic normality were proved in the case of estimating a parameter of a finite-dimensional marginal of the stationary process. By taking advantage of the special structure associated with a general *linear* statistic, we are able to further obtain results on the bias and the variance of the variance estimator as well. However, the central contribution of the present paper is to allow for the possibility of working with an estimator consistent for a parameter of the whole (infinite-dimensional) distribution of the process. This is achieved by letting the parameter M tend to infinity with increasing sample size. An additional feature is the use of *overlapping* subseries used in the variance estimator. Using overlapping instead of adjacent nonoverlapping subseries does not reduce the order of magnitude of the variance of $\hat{V}_{b/N}$, but it typically reduces it by a constant factor which in the sample mean example is 33%.

It is important to also note that the sample variance estimate $\hat{\sigma}_b^2$ coincides with the 'stationary' or 'moving blocks' jackknife estimate of variance introduced in Künsch (1989) and Liu and Singh (1988). Analogously, the variance estimate $\hat{V}_{b/N}$ coincides with the 'blocks of blocks' jackknife estimate that was studied in Politis and Romano (1989, 1990).

The blocked sample variance will be formally defined in the next section, and conditions under which it is a consistent estimator of the asymptotic variance of $\sqrt{Q}\ \bar{T}_N$ will be given. In Section 3, the asymptotic order of the Mean Squared Error of the blocked sample variance will be calculated, and the optimal (from the point of view of Mean Squared Error) rate at which the design parameter b should tend to infinity will be identified.

2. Consistency of the sample variance estimator

Let us now introduce some basic assumptions in connection with the set-up of linear statistics defined in the introduction. These assumptions will be used in showing consistency of the sample variance estimator. In what follows, all order notations and asymptotic statements will hold for the sample size N tending to infinity.

(A0) $\{X_n, n \in \mathbb{Z}\}$ is strictly stationary and α-mixing, i.e., $\alpha_X(k) \to 0$ as $k \to \infty$, where $\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$, and $A \in \mathscr{F}_{-\infty}^{0}$, $B \in \mathscr{F}_k^{\infty}$ are events in the σ-algebras generated by $\{X_n, n \leqslant 0\}$ and $\{X_n, n \geqslant k\}$ respectively.

(A1) $E|T_{1,M,L}^{(n)}|^{2p+\delta} < C$, for all $n = 1, \ldots, D$, for all M, where p is an integer with $p > 2$, and $0 < \delta \leqslant 2$, $C > 0$ are some constants. (Note that the law of $T_{1,M,L}$ does not depend on L since it is obtained from the first block of observations.)

(A2) $ET_{1,M,L} = \mu + O(Q^{-1/2})$, where μ is a parameter of the infinite-dimensional joint distribution of the X_n's.

(A3) $\sqrt{Q}\,(\bar{T}_N - E\bar{T}_N) \xrightarrow{\mathscr{L}} N(0, \sum_\infty)$, the multivariate normal distribution with a positive definite covariance matrix $\sum_\infty = (\sigma_{i,j,\infty})$, where $\sigma_{i,i,\infty} > 0$, $i = 1, \ldots, D$.

In stating the assumptions we have denoted $T^{(n)}_{i,M,L}$ to be the nth coordinate of $T_{i,M,L}$ (recall that the parameter μ takes values in $\mathbb{R}^D$, and thus so does its estimator $\bar{T}_N$).

The asymptotic normal distribution of assumption (A3) can be used to yield approximate confidence regions for $E\bar{T}_N$, which would be asymptotically valid confidence regions for μ as well, provided a stronger form of assumption (A2) holds, namely that $ET_{1,M,L} = \mu + o(Q^{-1/2})$. Alternatively, by using an asymptotic expansion for the bias of the form $ET_{1,M,L} = \mu + \mu_1 + o(Q^{-1/2})$, we can get confidence regions for μ based on the asymptotic normal distribution, provided now that we can estimate μ_1.

However, the asymptotic variance matrix must also be estimated in order for the Central Limit Theorem to be used. To this effect we introduce the *blocked* sample variance matrix $\hat{V}_{b/N}$, whose entry at row n_1 and column n_2 estimates the covariance between $(1/\sqrt{b})\sum_{j=1}^{b} T^{(n_1)}_{j,M,L}$ and $(1/\sqrt{b})\sum_{j=1}^{b} T^{(n_2)}_{j,M,L}$, and is given by

$$\hat{V}^{n_1,n_2}_{b/N} = \frac{1}{Q-b+1} \sum_{i=1}^{Q-b+1} \left(\frac{1}{\sqrt{b}} \sum_{j=i}^{i+b-1} T^{(n_1)}_{j,M,L} - \sqrt{b}\, \bar{T}^{(n_1)}_N \right) \times \left(\frac{1}{\sqrt{b}} \sum_{j=i}^{i+b-1} T^{(n_2)}_{j,M,L} - \sqrt{b}\, \bar{T}^{(n_2)}_N \right). \tag{7}$$

The following theorem gives conditions ensuring the consistency of the blocked sample variance.

Theorem 1. *Under assumptions* (A0), (A1), (A2), (A3) *and if*

(i) $M = o(N)$ *and* $L \sim aM$, *for some* $a \in (0, 1]$;

(ii) $b \to \infty$ *and* $b = o(Q)$;

(iii) $\sum_{k=1}^{\infty} k^{p-1}(\alpha_X(k))^{\delta/(2p+\delta)} < \infty$;

then, for any $n_1, n_2 \in \{1, \ldots, D\}$, *we have*

$$\hat{V}^{n_1,n_2}_{b/N} \xrightarrow{P} \sigma_{n_1,n_2,\infty}. \qquad \square \tag{8}$$

The proof of Theorem 1 amounts to controlling the order of magnitude of the Mean Squared Error of $\hat{V}^{n_1,n_2}_{b/N}$. It will not be given in detail here, since in the next section the problem of calculating the asymptotic order of the Mean Squared Error and the optimal rate at which b should tend to infinity will be explicitly addressed. As it turns out, $E\hat{V}^{n_1,n_2}_{b/N} = V^{n_1,n_2}_{b/N} + O(b/Q)$ and $\mathrm{Var}(\hat{V}^{n_1,n_2}_{b/N}) = O(b/Q)$, where $V^{n_1,n_2}_{b/N} \equiv \mathrm{Cov}((1/\sqrt{b})\sum_{i=1}^{b} T^{(n_1)}_{i,M,L}, (1/\sqrt{b})\sum_{i=1}^{b} T^{(n_2)}_{i,M,L})$, and $V^{n_1,n_2}_{b/N} \to \sigma_{n_1,n_2,\infty}$ as $b \to \infty$, by the Central Limit Theorem of assumption (A3).

It is important to observe that if μ is a parameter of a finite-dimensional distribution, say the autocovariance at lag s, then both M and L can be taken to be fixed numbers, say $M = s+1$ and $L = 1$. For a parameter of the infinite-dimensional joint distribution however, M should tend to infinity with the sample size N. It turns out that in this case, due to the high dependence of the summands $T_{i,M,L}$, the variance of the general linear statistic $\bar{T}_N$ is of order $O(M/N)$, regardless of whether $L = o(M)$, or $L \sim aM$.

To elaborate, note that

$$\operatorname{Var}(\bar{T}_N) = \frac{1}{Q}\left[\operatorname{Var}(T_{1,M,L}) + 2\sum_{i=1}^{Q}\left(1-\frac{i}{Q}\right)\operatorname{Cov}_{T,M,L}(i)\right],$$

where $\operatorname{Cov}_{T,M,L}(k) \equiv \operatorname{Cov}(T_{1,M,L}, T_{1+k,M,L})$. Now the summation $\sum_{i=1}^{Q}(1-(i/Q))\operatorname{Cov}_{T,M,L}(i)$ can be broken into two sums, namely $\sum_{i=1}^{[M/L]}$ and $\sum_{i=[M/L]+1}^{Q}$, the first one being of order $O(M/L)$ and the second being uniformly bounded (see also the proof of Lemma 2 and equation (22)). Since $Q \sim N/L$, it follows that $\operatorname{Var}(\bar{T}_N) = O(M/N)$.

Nevertheless, the covariance estimate $\hat{V}_{b/N}^{n_1,n_2}$ is consistent only if the variance of $\bar{T}_N$ is of order $O(1/Q)$. Our assumption (A3) and condition $L \sim aM$ of Theorem 1 are there to ensure that this is true.

It should also be noted that condition (iii) is not hard to fulfill; in particular, it is satisfied if the following holds:

(iv) $\alpha_X(k) = O(k^{-\lambda})$, where $\lambda > p(2p+\delta)/\delta$.

Note that (iv) is one of the most relaxed conditions on the mixing rate for the Central Limit Theorem of assumption (A3) to hold in the first place. So, by assuming (iv) for some sufficiently large λ, we can omit (A3) from the assumptions of Theorem 1, since then (A3) follows from a theorem of Tikhomirov (1980). However we need the existence of a common asymptotic variance. So if we formulate the weaker assumption

(A3′) For n_1, n_2 taking values in $\{1,\ldots,D\}$, and for $Q \to \infty$ as $N \to \infty$, $\lim_{N\to\infty}\operatorname{Cov}((1/\sqrt{Q})\sum_{i=1}^{Q} T_{i,M,L}^{(n_1)}, (1/\sqrt{Q})\sum_{i=1}^{Q} T_{i,M,L}^{(n_2)})$ exists and equals $\sigma_{n_1,n_2,\infty}$,

the following corollary of Theorem 1 is immediate.

Corollary 1. *Under assumptions* (A0), (A1), (A2), (A3′) *and conditions* (i), (ii), *and* (iv), *we have*

$$\hat{V}_{b/N}^{n_1,n_2} \xrightarrow{\text{P}} \sigma_{n_1,n_2,\infty}. \qquad \square \tag{9}$$

3. Mean squared error of variance estimation

In this section, the asymptotic order of the Mean Squared Error of the blocked sample variance will be calculated, and the optimal (from the point of view of Mean Squared Error) rate at which b should tend to infinity will be found.

We will concentrate on μ and $\bar{T}_N$ being univariate, i.e., $D=1$, in order to avoid cumbersome notations. The same arguments and results apply to the general case as well, if we focus on any single element of the blocked sample variance matrix. We will investigate the asymptotic second order properties of the blocked sample variance estimator

$$\hat{V}_{b/N}=\frac{1}{Q-b+1}\sum_{i=1}^{Q-b+1}\left(\frac{1}{\sqrt{b}}\sum_{j=i}^{i+b-1}T_{j,M,L}-\sqrt{b}\,\bar{T}_N\right)^2, \tag{10}$$

which, under the hypotheses of Theorem 1, is a consistent estimator of σ_∞^2 ($\sigma_{1,1,\infty}$ in the notation of the previous section).

In starting, let us define some quantities that are closely related to $\hat{V}_{b/N}$, though easier to work with. So for $l=1,2,\ldots,h$, where h is an integer that depends on the sample size N, let

$$\hat{V}_{b/N}^{l/h}\equiv\frac{1}{q_l}\sum_{i=1}^{q_l}\left(\frac{1}{\sqrt{b}}\sum_{j=(i-1)h+l}^{(i-1)h+b+l-1}T_{j,M,L}-\sqrt{b}\,\bar{T}_N\right)^2, \tag{11}$$

where $q_l=[(Q-b-l+1)/h]+1$. It is easy to see that $\hat{V}_{b/N}=(1/(Q-b+1))\times\sum_{l=1}^{h}q_l\hat{V}_{b/N}^{l/h}$, and because all the q_l's are of the same asymptotic order as $q\equiv[(Q-b)/h]+1=q_1$, it is immediate that $\hat{V}_{b/N}\sim(1/h)\sum_{l=1}^{h}\hat{V}_{b/N}^{l/h}$.

We will investigate the statistical properties of $\hat{V}_{b/N}$ through the properties of the $\hat{V}_{b/N}^{l/h}$'s. In particular, it would suffice to examine the statistical properties of $\hat{V}_{b/N}^{1/h}$, since the $\hat{V}_{b/N}^{l/h}$'s are asymptotically identically distributed for $l=1,2,\ldots,h$. This is established in the following Theorem 2 and Lemma 1, concerning the calculation of the first two moments of $\hat{V}_{b/N}^{l/h}$, and its asymptotic normality respectively.

Theorem 2. *Under the assumptions and conditions of Theorem* 1, *and the additional requirement $h\sim a_h b$, for some $a_h\in(0,1]$, we have*

$$E\hat{V}_{b/N}^{l/h}=V_{b/N}+\mathrm{O}(b/Q), \tag{12}$$

$$\mathrm{Var}(\hat{V}_{b/N}^{l/h})=\mathrm{O}(b/Q), \tag{13}$$

for any $l\in\{1,2,\ldots,h\}$, where $V_{b/N}\equiv\mathrm{Var}((1/\sqrt{b})\sum_{i=1}^{b}T_{i,M,L})$.

Lemma 1. *Under the assumptions and conditions of Corollary* 1, *and the additional requirement $h\sim a_h b$, for some $a_h\in(0,1]$, the estimator $\hat{V}_{b/N}^{l/h}$ is asymptotically normal, namely*

$$(\hat{V}_{b/N}^{l/h}-V_{b/N})/\sqrt{\mathrm{Var}(\hat{V}_{b/N}^{l/h})}\overset{\mathscr{L}}{\Rightarrow}\mathrm{N}(0,1), \tag{14}$$

for any $l\in\{1,2,\ldots,h\}$.

To calculate the bias of $\hat{V}_{b/N}^{1/h}$, which is defined as

$$\mathrm{Bias}(\hat{V}_{b/N}^{1/h})\equiv E\hat{V}_{b/N}^{1/h}-V_{Q/N}=(V_{b/N}-V_{Q/N})+\mathrm{O}(b/Q), \tag{15}$$

it remains to estimate $V_{b/N} - V_{Q/N}$. Note that $V_{Q/N} \equiv \mathrm{Var}((1/\sqrt{Q})\sum_{i=1}^{Q} T_{i,M,L})$, and $\lim_{N\to\infty} V_{Q/N} = \sigma_\infty^2$, as in assumption (A3′). This will be the subject of the following lemma.

Lemma 2. *Under the assumptions and conditions of Corollary* 1, $V_{Q/N} - V_{b/N} = \mathrm{O}(1/b)$.

Using the estimate from Lemma 2 and Theorem 2, and bearing in mind that $\mathrm{Bias}(\hat{V}_{b/N}) = E\hat{V}_{b/N} - V_{Q/N} = \mathrm{Bias}(\hat{V}_{b/N}^{1/h})$, Corollary 2 is offered.

Corollary 2. *Under the assumptions and conditions of Corollary* 1 *we have*

$$\mathrm{Bias}(\hat{V}_{b/N}) = \mathrm{Bias}(\hat{V}_{b/N}^{1/h}) = \mathrm{O}(1/b) + \mathrm{O}(b/Q). \qquad \square$$

Note that under the additional condition $b = \mathrm{o}(\sqrt{Q})$, as for example in Theorem 3.2 of Künsch (1989), then $\mathrm{Bias}(\hat{V}_{b/N}) = \mathrm{O}(1/b)$. This additional condition is quite reasonable as will be obvious from Corollary 3 given below.

It is also interesting that if h is of the same order of magnitude as b, the bias and variance of $\hat{V}_{b/N}^{1/h}$ are of the same asymptotic order as the bias and variance of $\hat{V}_{b/N}$, due to the very strong dependence among the $\hat{V}_{b/N}^{l/h}$'s.

Lemma 3. *Under the assumptions and conditions of Corollary* 1, $\mathrm{Var}(\hat{V}_{b/N}) = \mathrm{O}(b/Q)$.

The estimate offered in Lemma 3 can not generally be improved. This can be verified by a theorem of Künsch (1989) stating that in the sample mean example (where $M = L = 1$, $Q \sim N$, and $\phi_M(\cdot)$ is the identity function) we have $\mathrm{Var}(\hat{V}_{b/N}) \sim (4b/(3Q))\sigma_\infty^4$. As a Corollary of Lemma 3 and Corollary 2, the choice of the block size b is suggested in order to minimize the Mean Squared Error of variance estimation. If we define the Mean Squared Error to be $\mathrm{MSE}(\hat{V}_{b/N}) \equiv E(\hat{V}_{b/N} - V_{Q/N})^2$, then we have

Corollary 3. *Under the assumptions and conditions of Corollary* 1, *the choice* $b \sim a_b Q^{1/3}$, *for some constant* $a_b > 0$, *minimizes the asymptotic order of the Mean Squared Error, yielding* $\mathrm{MSE}(\hat{V}_{b/N}) = \mathrm{O}(Q^{-2/3})$. $\square$

The constant a_b could be calculated if the specifics of the particular estimation problem are given, thus enabling us to explicitly calculate the proportionality constants in the order of magnitude results of Corollary 2 and Lemma 3. For example, if the linear statistic under consideration is the sample mean, $\mathrm{Bias}(\hat{V}_{b/N}) \sim (-2/b)\sum_{k=1}^{\infty} k\,\mathrm{Cov}(X_0, X_k)$, and $\mathrm{Var}(\hat{V}_{b/N}) \sim (4b/(3Q))\sigma_\infty^4$ (cf. Künsch, 1989). In that case, an asymptotically optimal choice would be to let $a_b = 2^{1/3}3^{1/3}|\sum_{k=1}^{\infty} k\,\mathrm{Cov}(X_0, X_k)|^{2/3}/\sigma_\infty^{4/3}$, where of course all unknown quantities must be consistently estimated.

Bearing in mind that estimation of σ_∞^2 is our original problem at hand, one can see that the problem of choosing b in practice poses difficulties, even in the simplest example (case of the sample mean). One possible way out would be to calculate

$\hat{V}_{b/N}$ for a suitable range of b values. In this way we would have $\hat{V}_{b/N}$ expressed as a function of b that should be simultaneously solved (for $\hat{V}_{b/N}$ and b) with the relationship guaranteeing asymptotic MSE optimality, i.e., $\hat{V}_{b/N} = \hat{R}\sqrt{6N/b^3}$, where $\hat{R}$ is a consistent estimate of $|\sum_{k=1}^{\infty} k\,\mathrm{Cov}(X_0, X_k)|$. Other practical guidelines for choosing b are given in Politis and Romano (1989).

Some comments are also in order regarding the order of magnitude of the Mean Squared Error as given in Corollary 3. For concreteness let us once again consider the sample mean example. Recall that in the case of an *independent* sample we would take $b=1$ in the definition of the blocked sample variance and we would recover the ordinary sample variance which has a Mean Squared Error of order $O(N^{-1})$ associated with it. However, in the presence of weak dependence quantified by a mixing condition, the Mean Squared Error of the blocked sample variance jumps to being of order $O(N^{-2/3})$, a very abrupt change if we consider that the covariance terms in equation (4) could be arbitrarily close to zero.

It is interesting to consider the case of m-dependence, i.e., the case where $\alpha_X(k)=0$, for all $k>m$, which is a situation intermediate between independence (which is the same as 0-dependence) and strong mixing. Looking at the proof of Lemma 2 it is seen that $V_{Q/N} - V_{b/N}$ would still be of order $O(1/b)$, thus yielding a minimum Mean Squared Error again of order $O(N^{-2/3})$. Thus, in this particular situation, it seems preferable to use formula (4) directly, replacing the unknown covariances with sample estimates, and taking advantage of the knowledge that $\mathrm{Cov}(X_1, X_{k+1})=0$, for all $k>m$.

To conclude, observe that in the sample mean case the blocked sample variance estimate is *identical* to a spectral estimate of the lag-window type with $W_t^{(M)} \equiv 1$ (cf. formula (3)) evaluated at frequency zero (cf. Künsch 1989). In this respect, the choice $b \sim a_b N^{1/3}$ and the resulting Mean Squared Error of order $O(N^{-2/3})$ are well known to be optimal. To further reduce the Mean Squared Error, (by means of reducing the bias), the use of a tapering window $W_t^{(b)}$ is suggested in the spectral estimation literature. This technique was also successfully applied in the tapered blocks jackknife method of Künsch (1989). In Priestley (1981) many different tapering windows are presented that lead to the bias of the spectral estimate being of $O(1/b^2)$, resulting in an optimal choice $b \sim a_b N^{1/5}$ and a Mean Squared Error of order $O(N^{-4/5})$. The prototype of such a window is $W_t^{(b)} = w((t-0.5)/b)$, where $w:(0,1) \to (0,1)$ is a function symmetric about $\frac{1}{2}$ and nondecreasing on $(0,\frac{1}{2})$. Taking $w(t)=2t$ for $t \leqslant \frac{1}{2}$ corresponds to a spectral estimate that is equivalent to a periodogram smoothed using Parzen's kernel.

By analogy to the extensively studied sample mean case, we can define the *tapered blocked sample variance* $\hat{V}_{b/N}^{(W)}$ which, in the setting of the general linear statistic $\bar{T}_N$, is an estimator of the variance of $\sqrt{Q}\,\bar{T}_N$,

$$\hat{V}_{b/N}^{(W)} \equiv \frac{1}{Q-b+1} \sum_{i=1}^{Q-b+1} \left(\frac{1}{\sqrt{b}} \sum_{j=i}^{i+b-1} W_{j-i+1}^{(b)}\, T_{j,M,L} - \sqrt{b}\,\bar{T}_N \right)^2. \tag{16}$$

It is plausible that with proper choice of the tapering window the estimator $\hat{V}_{b/N}^{(W)}$

will achieve a smaller Mean Squared Error than $\hat{V}_{b/N}$, by possessing a bias of smaller order.

4. Technical proofs

Proof of Theorem 2. First note that conditions (i), (ii) imply $q \to \infty$ and $h \to \infty$, as well as $Q \to \infty$. In addition, they imply that $N/M \sim aQ$.

We will carry out the proof for the case $l=1$, the other cases being similar. Let us denote $\tilde{B}_i = (1/\sqrt{b}) \sum_{j=(i-1)h+1}^{(i-1)h+b} T_{j,M,L}$. Then,

$$\hat{V}_{b/N}^{1/h} = \frac{1}{q} \sum_{i=1}^{q} (\tilde{B}_i - \sqrt{b}\, \bar{T}_N)^2$$
$$= \frac{1}{q} \sum_{i=1}^{q} \left\{ \tilde{B}_i - E\tilde{B}_i - \sqrt{b} \left(\bar{T}_N - \frac{1}{\sqrt{b}} E\tilde{B}_i \right) \right\}^2 = A_N - 2C_N + D_N,$$

where

$$A_N = \frac{1}{q} \sum_{i=1}^{q} (\tilde{B}_i - E\tilde{B}_i)^2,$$

$$C_N = \frac{1}{q} \sum_{i=1}^{q} \sqrt{b} \left(\bar{T}_N - \frac{1}{\sqrt{b}} E\tilde{B}_i \right) (\tilde{B}_i - E\tilde{B}_i),$$

$$D_N = \frac{1}{q} \sum_{i=1}^{q} b \left(\bar{T}_N - \frac{1}{\sqrt{b}} E\tilde{B}_i \right)^2.$$

The proof will proceed by showing that A_N is the dominant part of $\hat{V}_{b/N}^{1/h}$, both in terms of expected value and variance. Recall that the X_n's are α_X mixing and the $\tilde{B}_i$ are functions of finite blocks of them. Hence the $\tilde{B}_i$ are $\alpha_{\tilde{B},M,L}$ mixing with

$$\alpha_{\tilde{B},M,L}(n) \leq \alpha_X([(n-1)h - (b-1)]L - M), \tag{17}$$

for $n \geq n_0 = [M/(hL) + (b-1)/h] + 1$. Since from our conditions $M = O(L)$ and $b = O(h)$, it is ensured that there will be a smallest n_0 such that (17) will hold regardless of the value of N. Hence, for all practical purposes, the $\tilde{B}_i$'s, $i \in \mathbb{Z}$, can be treated as governed by the *same* mixing coefficient, namely the right-hand side of formula (17).

First note that $EA_N = \mathrm{Var}(\tilde{B}_1) = V_{b/N}$, and

$$\mathrm{Var}(A_N) = \frac{1}{q} \mathrm{Var}(\tilde{B}_1 - E\tilde{B}_1)^2 + \frac{2}{q^2} \sum_{i=1}^{q-1} (q-i)$$
$$\times \mathrm{Cov}\{(\tilde{B}_1 - E\tilde{B}_1)^2, (\tilde{B}_{i+1} - E\tilde{B}_{i+1})^2\}. \tag{18}$$

But a well-known theorem of Ibragimov (cf. Roussas and Ioannides, 1987) gives

$$\mathrm{Cov}\{(\tilde{B}_1 - E\tilde{B}_1)^2, (\tilde{B}_{i+1} - E\tilde{B}_{i+1})^2\}$$
$$\leq 10(E|\tilde{B}_1 - E\tilde{B}_1|^{2p})^{2/p} (\alpha_{\tilde{B},M,L}(i))^{(p-2)/p}. \tag{19}$$

Also, from a theorem of Yokoyama (1980) and conditions (iii) and (A1), the following moment inequality holds:

$$E|\tilde{B}_1 - E\tilde{B}_1|^{2p} \leq K_X(E|T_{1,M,L}|^{2p+\delta})^{2p/(2p+\delta)}, \tag{20}$$

where K_X depends only on a_X and p. Combining the above with assumption (A1) yields

$$\mathrm{Var}(A_N) = \mathrm{O}\left(\frac{1}{q} + \frac{20}{q^2}\sum_{i=1}^{q-1}(q-i)(\alpha_{\tilde{B},M,L}(i))^{(p-2)/p}\right). \tag{21}$$

Now, from condition (iii) it also follows that $\sum_{k=1}^{\infty}(\alpha_X(k))^{(p-2)/p} < \infty$, since in assumption (A1) it is assumed that $p \geq 3$. Thus, by the discussion relating $\alpha_{\tilde{B},M,L}$ with α_X, it follows that $\mathrm{Var}(A_N) = \mathrm{O}(1/q) = \mathrm{O}(b/Q)$, since $q \sim Q/h$, and h is of the same order of magnitude as b.

To complete the proof, it is not hard to see that $ED_N = \mathrm{O}(b/Q)$, $ED_N^2 = \mathrm{O}(b^2/Q^2)$, $EC_N = \mathrm{O}(\sqrt{b/(qQ)}) = \mathrm{O}(b/Q)$, and $EC_N^2 = \mathrm{O}(b/qQ) = \mathrm{O}(b^2/Q^2)$. To elaborate, let us focus on D_N, since C_N can be handled in a similar way.

$$D_N = b(\bar{T}^N - \mu)^2 + b\left(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_1\right)^2 + 2b(\bar{T}_N - \mu)\left(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_1\right).$$

Since assumption (A2) implies $\mu = (1/\sqrt{b})E\tilde{B}_1 = \mathrm{O}(Q^{-1/2})$, and $E\bar{T}_N - \mu = \mathrm{O}(Q^{-1/2})$ as well, we gather that $ED_N = bE(\bar{T}_N - \mu)^2 + \mathrm{O}(b/Q)$. Finally, $E(\bar{T}_N - \mu)^2 = \mathrm{Var}(\bar{T}_N) + (E\bar{T}_N - \mu)^2 = \mathrm{O}(1/Q)$, and the result $ED_N = \mathrm{O}(b/Q)$ is obtained.

Similarly, look at

$$\frac{D_N^2}{b^2} = (\bar{T}_N - \mu)^4 + \left(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_1\right)^4 + 6(\bar{T}_N - \mu)^2\left(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_1\right)^2$$
$$+ 4(\bar{T}_N - \mu)\left(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_1\right)^3 + 4(\bar{T}_N - \mu)^3\left(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_1\right).$$

Again by using Yokoyama's theorem and assumption A_2 we have $E(\bar{T}_N - \mu)^4 = \mathrm{O}(Q^{-2})$ and $E(\bar{T}_N - \mu)^3 = \mathrm{O}(Q^{-3/2})$, from which the result $ED_N^2 = \mathrm{O}(b^2/Q^2)$ is proved. □

Proof of Lemma 1. Again the proof will be carried out for the case $l = 1$ only, the other cases being similar. To prove the asymptotic normality, consider the decomposition $\hat{V}_{b/N}^{1/h} = A_N - 2C_N + D_N$, which is used in the proof of Theorem 2. From condition (iv) and Tikhomirov's theorem it follows that

$$(A_N - V_{b/N})\sqrt{\mathrm{Var}(A_N)} \stackrel{\mathcal{L}}{\Rightarrow} \mathrm{N}(0, 1),$$

because $EA_N = V_{b/N}$. Looking again at the proof of Theorem 2, it is easy to see that $\sqrt{Q/b}\, C_N \xrightarrow{P} 0$ and $\sqrt{Q/b}\, D_N \xrightarrow{P} 0$. Since both $\mathrm{Var}(A_N)$ and $\mathrm{Var}(\hat{V}_{b/N}^{1/h})$ are of the same asymptotic order, namely b/Q, an application of Slutsky's theorem completes the proof. □

Proof of Lemma 2. From Ibragimov's theorem and assumption (A1) it follows that

$$|\mathrm{Cov}_{T,M,L}(k)| = \mathrm{O}(\alpha_{T_{M,L}}(k))^{(2(p-1)+\delta)/(2p+\delta)}, \tag{22}$$

where $\mathrm{Cov}_{T,M,L}(k) \equiv \mathrm{Cov}(T_{0,M,L}, T_{k,M,L})$. As in the proof of Theorem 2, it is seen that, viewed as a sequence in $i = 1, 2, \ldots$ for any N, the $T_{i,M,L}$'s are $\alpha_{T_{M,L}}$ mixing with

$$\alpha_{T_{M,L}}(k) \leqslant \alpha_X(kL - M),$$

for $k \geqslant m+1$, where $m = [1/a]$.

Also, from condition (iv), $\alpha_X(k) = \mathrm{O}(k^{-\lambda})$; therefore, $\alpha_{T_{M,L}}(k) = \mathrm{O}(kL-M)^{-\lambda} = \mathrm{O}(M^{-\lambda}(ak-1)$, for $k \geqslant m+1$.

Hence

$$|\mathrm{Cov}_{T,M,L}(k)| = \mathrm{O}(M^{-\nu}(ak-1)^{-\nu}), \tag{23}$$

for $k \geqslant m+1$, where $\nu = \lambda(2(p-1)+\delta)/(2p+\delta)$. Since by condition (iv) we have $\nu > (p/\delta)(2p-2+\delta) \geqslant 9$, it is immediately seen that $\sum_{k=0}^{\infty} k|\mathrm{Cov}_{T,M,L}(k)|$ is bounded by a constant independent of M and L.

Now

$$\begin{aligned}
\tfrac{1}{2}(V_{Q/N} - V_{b/N}) &= \sum_{k=1}^{b} k\left(\frac{1}{b} - \frac{1}{Q}\right) \mathrm{Cov}_{T,M,L}(k) + \sum_{k=b+1}^{Q} \left(1 - \frac{k}{Q}\right) \mathrm{Cov}_{T,M,L}(k) \\
&= \sum_{k=1}^{m} k\left(\frac{1}{b} - \frac{1}{Q}\right) \mathrm{Cov}_{T,M,L}(k) + \sum_{k=m+1}^{b} k\left(\frac{1}{b} - \frac{1}{Q}\right) \mathrm{Cov}_{T,M,L}(k) \\
&\quad + \sum_{k=b+1}^{Q} \left(1 - \frac{k}{Q}\right) \mathrm{Cov}_{T,M,L}(k).
\end{aligned}$$

Since m is a fixed constant, the first term in the above expression is $\mathrm{O}(1/b)$. Thus

$$\tfrac{1}{2}(V_{Q/N} - V_{b/N}) = \mathrm{O}(1/b) + C_b - C_Q,$$

where

$$C_b = \sum_{k=m+1}^{b} \frac{k}{b} \mathrm{Cov}_{T,M,L}(k) + \sum_{k=b+1}^{Q} \mathrm{Cov}_{T,M,L}(k),$$

$$C_Q = \sum_{k=m+1}^{Q} \frac{k}{Q} \mathrm{Cov}_{T,M,L}(k).$$

Finally, since $|\sum_{k=b+1}^{Q} \mathrm{Cov}_{T,M,L}(k)| \leqslant \sum_{k=b+1}^{Q} (k/b)|\mathrm{Cov}_{T,M,L}(k)|$, and using the fact that the sum $\sum_{k=0}^{\infty} k|\mathrm{Cov}_{T,M,L}(k)|$ is bounded, we have that $C_b = \mathrm{O}(1/b)$, and $C_Q = \mathrm{O}(1/Q) = \mathrm{o}(1/b)$. □

Proof of Lemma 3. As mentioned in Section 3, we have that $\hat{V}_{b/N} \sim (1/h)\sum_{l=1}^{h} \hat{V}_{b/N}^{l/h}$. Now if we let $\xi_k = \hat{V}_{b/N}^{k/h}$ we have

$$\mathrm{Var}\left(\frac{1}{h}\sum_{k=1}^{h} \xi_k\right) = \frac{1}{h^2}\sum_{i=1}^{h}\sum_{j=1}^{h} \mathrm{Cov}(\xi_i, \xi_j) = \frac{1}{h^2}\mathrm{O}\left(\frac{h^2 b}{Q}\right) = \mathrm{O}(b/Q),$$

where it was used that by Theorem 2 and the Cauchy-Schwarz inequality $\mathrm{Cov}(\xi_i, \xi_j) = \mathrm{O}(b/Q)$, for any i and j. Hence the lemma is proved. □

Acknowledgements

Many thanks are due to the anonymous referees for their detailed comments that helped ameliorate the presentation of this work.

References

[1] D.R. Brillinger, Time Series, Data Analysis and Theory (Holt, Rinehart and Winston, New York, 1975).

[2] E. Carlstein, The use of subseries values for estimating the variance of a general statistic from a stationary sequence, Ann. Statist. 14 (1986a) 1171-1179.

[3] E. Carlstein, Asymptotic normality for a general statistic from a stationary sequence, Ann. Probab. 14 (1986b) 1371-1379.

[4] E. Carlstein, Law of large numbers for the subseries values of a statistic from a stationary sequence, Statist. 19 (1988) 295-299.

[5] E. Carlstein, Asymptotic multivariate normality for the subseries values of a general statistic from a stationary sequence with applications to nonparametric confidence intervals, Probab. Math. Statist. 10 (1989) 191-200.

[6] U. Grenander and M. Rosenblatt, Statistical Analysis of Stationary Time Series (Dover, New York, 1957).

[7] E.J. Hannan, Multiple Time Series (Wiley, New York, 1970).

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

[9] R.Y. Liu and K. Singh, Moving blocks jackknife and bootstrap capture weak dependence, unpublished manuscript, Dept. of Statist. Rutgers Univ. (New Brunswick, NJ, 1988).

[10] D.N. Politis and J.P. Romano, A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, Ann. Statist. 20 (1992).

[11] D.N. Politis and J.P. Romano, A nonparametric resampling procedure for multivariate confidence regions in time series analysis, in: C. Page and R. Le Page, eds., Proc. of INTERFACE '90, 22nd Symp. on the Interface of Comput. Sci. and Statist., 17-19 May 1990 (East Lansing, MI, 1990) pp. 98-103.

[12] M.B. Priestley, Spectral Analysis and Time Series (Academic Press, New York, 1981).

[13] G.G. Roussas and D. Ioannides, Moment inequalities for mixing sequences of random variables, Stoch. Anal. Appl. 5(1) (1987) 61-120.

[14] A.N. Tikhomirov, On the convergence rate in the Central Limit Theorem for weakly dependent random variables, Theory Probab. Appl. XXV (1980) 790-809.

[15] R. Yokoyama, Moment bounds for stationary mixing sequences, Z. Wahrsch. Verw. Gebiete 52 (1980) 45-57.