The Annals of Statistics
1992, Vol. 20, No. 4, 1985–2007

A GENERAL RESAMPLING SCHEME FOR TRIANGULAR ARRAYS OF α-MIXING RANDOM VARIABLES WITH APPLICATION TO THE PROBLEM OF SPECTRAL DENSITY ESTIMATION[1]

BY DIMITRIS N. POLITIS AND JOSEPH P. ROMANO[2]

Purdue University and Stanford University

In 1989 Künsch introduced a modified bootstrap and jackknife for a statistic which is used to estimate a parameter of the m-dimensional joint distribution of stationary and α-mixing observations. The modification amounts to resampling whole blocks of consecutive observations, or deleting whole blocks one at a time. Liu and Singh independently proposed (in 1988) the same technique for observations that are m-dependent. However, many time-series statistics, notably estimators of the spectral density function, involve parameters of the whole (infinite-dimensional) joint distribution and, hence, do not fit in this framework. In this report we generalize the "moving blocks" resampling scheme of Künsch and Liu and Singh; a still modified version of the nonparametric bootstrap and jackknife is seen to be valid for general linear statistics that are asymptotically normal and consistent for a parameter of the whole joint distribution. We then apply this result to the problem of estimation of the spectral density.

1. Introduction. The bootstrap and jackknife [Efron (1979, 1982)] have proven to be powerful tools for approximating the sampling distribution and variance of complicated statistics defined on a sequence of independent identically distributed (i.i.d.) random variables [Bickel and Freedman (1981) and Singh (1981)]. It also has found application in problems where the assumption of i.i.d. random variables is violated, but always by means of reducing the problem to an approximate i.i.d. setting by focusing on the "residuals" of some general regression. Such examples include linear regression [Freedman (1981) and Liu (1988)], autoregressive time series [Efron and Tibshirani (1986) and Swanepoel and van Wyk (1986)], nonparametric regression and nonparametric kernel spectral estimation [Härdle and Bowman (1988) and Franke and Härdle (1992)]. In all these situations it is the residuals that are being resampled, not the original observations.

Recently, Künsch (1989) and Liu and Singh (1988) independently introduced a nonparametric version of the bootstrap and jackknife that is valid for weakly dependent stationary observations. This technique amounts to resampling or deleting one by one whole blocks of observations. Künsch's exposition was influenced by an earlier related work of Carlstein (1986).

Received December 1989; revised January 1992.
[1]Research supported in part by NSF Grant DMS-89-05874.
[2]Also supported by NSF Grant DMS-89-57217.
AMS 1980 *subject classifications*. Primary 62M10; secondary 62G05.
Key words and phrases. Time series, spectral density, weak dependence, nonparametric estimation, resampling methods, bootstrap, jackknife.

However, in both Künsch (1989) and Liu and Singh (1988), attention is concentrated on estimators of parameters of the m-dimensional distribution of the observations, with m fixed. Nonetheless, in many time-series problems the objective is to estimate a parameter of the whole (infinite-dimensional) joint distribution. A prime example is the problem of estimating the spectral density function. We will extend the technique of Künsch and Liu and Singh in this direction. In order to do this, we naturally will be brought into the setting of a triangular array of mixing random variables.

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 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. The special case of m-dependence holds if $\alpha_X(k) = 0$, $\forall\, k > m$, where m is some fixed integer.

Suppose μ is a parameter of the whole (infinite-dimensional) joint distribution of sequence $\{X_n, n \in \mathbb{Z}\}$. The objective is to obtain confidence intervals for μ based on a stretch of observations from time series $\{X_n\}$. Attention will focus on estimators of μ that can be put in the form of an average of functions defined on the observations. For each $N = 1, 2, \dots$, let $B_{i,M,L}$ be the block of M consecutive observations starting from $(i-1)L+1$, that is, the subseries $X_{(i-1)L+1}, \dots, X_{(i-1)L+M}$, where M and L are integer functions of N. Note that $B_{i,M,L}$ for $i = \dots, -1, 0, +1, \dots$ can be obtained 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 $\phi_M\colon \mathbb{R}^{dM} \to \mathbb{R}$. So, for fixed N, the $T_{i,M,L}$ for $i \in \mathbb{Z}$ constitute a strictly stationary sequence. In practice a segment $X_1, \dots, X_N$ from the time series $\{X_n\}$ would be observed, which would permit us to compute $T_{i,M,L}$ for $i = 1, \dots, Q$ 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 $T_{i,M,L}$, $i = 1, \dots, Q$. The general linear statistic is now defined by

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

A linear statistic of this form for the special case $M = m$ and $L = 1$ was also discussed in Künsch (1989), as an estimator of a parameter of the m-dimensional joint distribution of sequence $\{X_n, n \in \mathbb{Z}\}$, where m is a fixed integer. We now mention some examples of time series statistics that can fit in the framework of the general linear statistic. For the following examples, assume X_n is univariate, that is, $d = 1$.

(I) The sample mean: $\bar{X} = (1/N)\Sigma_{i=1}^{N} X_i$. Just take $M = L = 1$ and ϕ_M to be the identity function.

(II) The (unbiased) sample autocovariance at lag s:

$$\frac{1}{N-s}\sum_{i=1}^{N-s} X_i X_{i+s}.$$

Take $L = 1$, $M = s + 1$ and $\phi_M(x_1, \dots, x_M) = x_1 x_M$.

(III) The lag-window spectral density estimator (cf. Section 5).

For multivariate time series, we similarly can use the above formulation of the general linear statistic to define the sample cross-covariance and cross-spectrum estimators (cf. Section 6).

We propose a "blocks of blocks" resampling scheme as follows: Given the observations $X_1, \dots, X_N$, we concentrate on row N of the triangular array, that is, on $T_{i,M,L}$, $i = 1, \dots, Q$. Define $\mathscr{B}_j$ to be the block of b consecutive $T_{i,M,L}$'s starting from $T_{(j-1)h+1,M,L}$, that is, the block $\mathscr{B}_j = (T_{(j-1)h+1,M,L}, \dots, T_{(j-1)h+b,M,L})$. This should be compared with the definition of our $T_{i,M,L}$ from the original observations. The $\mathscr{B}_j$ blocks can be obtained from the $T_{i,M,L}$, $i = 1, \dots, Q$ by means of a window of width b moving at lags h at a time. Note that there are $q = [(Q - b)/h] + 1$ such $\mathscr{B}_j$, $j = 1, \dots, q$.

Sampling with replacement from the set $\{\mathscr{B}_1, \dots, \mathscr{B}_q\}$, or deleting one of the $\mathscr{B}_j$ at a time from the Nth row of the triangular array, defines our blocks of blocks bootstrap and jackknife procedures. In Sections 3 and 4 we show that under regularity conditions these procedures yield consistent estimates of the sampling distribution and the variance of $\bar{T}_N$. Künsch [(1989), page 1218] used the blocks of blocks idea in the special case $M = m$ and $L = h = 1$ and indicated how by a simple blocking transformation the problem can be translated into one with $m = 1$, in which case the "moving blocks" procedure of Künsch (1989) and Liu and Singh (1988) immediately applies. It is to be noted that while examples (I) and (II) can be handled by the moving blocks procedure, our example (III) *cannot*.

The reason is that taking M fixed is done under the condition that, for *fixed* M, $\bar{T}_N \to_p \mu$ as $N \to \infty$, where μ is the parameter of interest. However, this requires unbiasedness, that is, $ET_{1,M,L} = \mu$. It turns out that, as in the case of the spectral density, this is not a valid assumption. A more plausible assumption is just to require asymptotic unbiasedness, that is, $ET_{1,M,L} \to \mu$ as $M \to \infty$.

As a matter of fact, if we are not willing to make analytical corrections for the bias, it will be apparent that it is necessary to let $M \to \infty$ at a properly fast rate. In the spectral estimation problem this amounts to *undersmoothing* the periodogram. It will be further shown that undersmoothing is not required for variance estimation, but only for the proper centering of confidence intervals obtained either by the blocks of blocks methodology or by the central limit theorem.

2. Basic assumptions. In the sequel all limits and order notations will be taken as $N \to \infty$ unless otherwise stated. We will make frequent use of the following assumptions:

(A_0) $\{X_n, n \in \mathbb{Z}\}$ is strictly stationary and α-mixing, that is, $\alpha_X(k) \to 0$ as $k \to \infty$.

(A_1) $E|T_{1,M,L}|^{2p+\delta} < C$, for all M, where p is an integer with $p > 2$ and $0 < \delta \leq 2$ and $C > 0$ are some constants. (Note that by definition $T_{1,M,L}$ does not depend on L since it is obtained from the first block of observations.)

(A_2) $ET_{1,M,L} = \mu + o(Q^{-1/2})$, where μ is a parameter of the infinite-dimensional joint distribution of the X_n's.

(A_3) $\sqrt{Q}(\overline{T}_N - E\overline{T}_N) \Rightarrow_{\mathscr{L}} N(0, \sigma_\infty^2)$, with $0 < \sigma_\infty^2 < \infty$.

Assumption (A_2) is basically required so that the asymptotic order of the bias of $\overline{T}_N$ is smaller than the asymptotic order of its standard deviation. It should be pointed out that assumptions (A_2) and (A_3) together permit us to treat confidence intervals for $E\overline{T}_N$ as confidence intervals for μ asymptotically, since they imply (by Slutsky's theorem) that

$$\sqrt{Q}\left(\overline{T}_N - \mu\right) \Rightarrow_{\mathscr{L}} N\left(0, \sigma_\infty^2\right). \tag{3}$$

The asymptotic normal distribution of (3) can be used to yield approximate confidence intervals for μ. However, the variance σ_∞^2 must somehow be estimated because, more often than not, a closed-form calculation is not feasible. In addition, a *different* estimate of the sampling distribution of $\sqrt{Q}(\overline{T}_N - \mu)$ might be a still better approximation than (3), thus giving confidence intervals that are more accurate. It is these two roles that the jackknife and bootstrap are usually called to play.

The following lemma ensures that the $T_{i,M,L}$ are weakly dependent in each row if the original series is weakly dependent.

LEMMA 1. *The following hold for each N fixed*:

(a) *If the X_n are m-dependent, then the $T_{i,M,L}$, $i \in \mathbb{Z}$, are m'-dependent, with $m' = [(M + m)/L]$.*

(b) *If the X_n are α-mixing with mixing coefficient $\alpha_X(k)$, then the $T_{i,M,L}$, $i \in \mathbb{Z}$, are also α-mixing with mixing coefficient $\alpha_{T_{M,L}}(k) \leq \alpha_X(kL - M)$, for $k \geq [M/L] + 1$.*

If we have $aM \leq L$, *for some constant* $a > 0$, *then the following also holds*:

(c) *If the* X_n *are* α*-mixing with mixing coefficient* $\alpha_X(k)$, *then the* $T_{i,M,L}$, $i \in \mathbb{Z}$, *are also* α*-mixing with mixing coefficient* $\alpha_{T_{M,L}}(k) \leq \alpha_X(kL - M)$, *for* $k \geq [1/a] + 1$.

If in addition to $aM \leq L$ *we have* $M \to \infty$ *as* $N \to \infty$, *then the following holds*:

(d) *For any fixed* $k \geq [1/a] + 1$, *we have* $\lim_{N\to\infty} \alpha_{T_{M,L}}(k) = 0$.

PROOF. (i) For $k \geq m' = [(M + m)/L]$, $T_{i,M,L}$ and $T_{i+k,M,L}$ are functions of $B_{i,M,L}$ and $B_{i+k,M,L}$, respectively, and hence are independent.

(ii) Similarly, looking at $T_{i,M,L}$ and $T_{i+k,M,L}$ for $k \geq [M/L] + 1$, there is a block of $kL - M$ observations separating $B_{i,M,L}$ and $B_{i+k,M,L}$. Hence $\alpha_{T_{M,L}}(k) \leq \alpha_X(kL - M)$. □

Parts (c) and (d) of the lemma are trivial consequences of parts (a) and (b), but they will be most useful since they give bounds for the mixing coefficient of the $T_{i,M,L}$'s that hold *regardless* of the value of N. It is interesting to observe that part (d) implies that the $T_{i,M,L}$'s are asymptotically m-dependent, with $m = [1/a]$.

In the next two sections, we will extend the asymptotic results of Künsch (1989) and Liu and Singh (1988) in the case of the triangular array of the $T_{i,M,L}$, where M is allowed to tend to infinity as $N \to \infty$.

3. Blocks of blocks jackknife. By focusing attention on row N of the triangular array of the $T_{i,M,L}$, $i = 1,\dots,Q$, define $\mathscr{B}_j$ to be the block of b consecutive $T_{i,M,L}$'s starting from $T_{(j-1)h+1,M,L}$; that is, $\mathscr{B}_j = (T_{(j-1)h+1,M,L}, \dots, T_{(j-1)h+b,M,L})$. The $\mathscr{B}_j$ blocks can be obtained from the $T_{i,M,L}$, $i = 1,\dots,Q$, by means of a window of width b moving at lags h at a time. Note that there are $q = [(Q - b)/h] + 1$ such $\mathscr{B}_j$, $j = 1,\dots,q$. The $\mathscr{B}_j$'s depend on M and L as well, although we do not explicitly put it in the notation. The block size b and the lag h will be assumed to be integer functions of N. Let $\overline{T}_{N,-j}$ be the average of the remaining $T_{i,M,L}$'s after deleting block $\mathscr{B}_j$. Then, define the so-called pseudovalues J_j, $j = 1,\dots,q$, by $J_j = (1/b)(Q\overline{T}_N - (Q - b)\overline{T}_{N,-j})$.

The blocks of blocks jackknife estimate of the variance of $\sqrt{Q}\,\overline{T}_N$ now is defined as

$$\hat{V}_{\text{JACK}}\left(\sqrt{Q}\,\overline{T}_N\right) = \frac{b}{q}\sum_{j=1}^{q}\left(J_j - \overline{T}_N\right)^2. \tag{4}$$

The following theorem gives conditions ensuring the consistency of the jackknife estimate of variance.

THEOREM 1. *Under assumptions* (A_0)–(A_3) *and if*:

(i) $M = o(N)$ *and* $L \sim aM$, *for some* $a \in (0, 1]$,
(ii) $b \to \infty$ *and* $h = O(b)$ *and* $b = O(h)$,
(iii) $b = o(Q)$,
(iv) $\sum_{k=1}^{\infty} k^{p-1}(\alpha_X(k))^{\delta/(2p+\delta)} < \infty$,

then

$$\hat{V}_{\mathrm{JACK}}\left(\sqrt{Q}\,\overline{T}_N\right) \to_p \sigma_\infty^2. \tag{5}$$

PROOF. Note that conditions (i)–(iii) imply $q \to \infty$ and $h \to \infty$, as well as $Q \to \infty$. In addition, they imply that $N/M \sim aQ$. Let

$$\tilde{B}_i = \frac{1}{\sqrt{b}} \sum_{j=(i-1)h+1}^{(i-1)h+b} T_{j,M,L}.$$

Then,

$$\begin{aligned}
\hat{V}_{\mathrm{JACK}}\left(\sqrt{Q}\,\overline{T}_N\right) &= \frac{b}{q}\sum_{i=1}^{q}\left(\frac{1}{b}\sum_{j=(i-1)h+1}^{(i-1)h+b} T_{j,M,L} - \overline{T}_N\right)^2 \\
&= \frac{1}{q}\sum_{i=1}^{q}\left(\tilde{B}_i - \sqrt{b}\,\overline{T}_N\right)^2 \\
&= \frac{1}{q}\sum_{i=1}^{q}\left\{\tilde{B}_i - E\tilde{B}_i - \sqrt{b}\left(\overline{T}_N - \frac{1}{\sqrt{b}}E\tilde{B}_i\right)\right\}^2 \\
&= A_N - 2C_N + D_N,
\end{aligned}$$

where

$$\begin{aligned}
A_N &= \frac{1}{q}\sum_{i=1}^{q}\left(\tilde{B}_i - E\tilde{B}_i\right)^2, \\
C_N &= \frac{1}{q}\sum_{i=1}^{q}\sqrt{b}\left(\overline{T}_N - \frac{1}{\sqrt{b}}E\tilde{B}_i\right)\left(\tilde{B}_i - E\tilde{B}_i\right), \\
D_N &= \frac{1}{q}\sum_{i=1}^{q} b\left(\overline{T}_N - \frac{1}{\sqrt{b}}E\tilde{B}_i\right)^2.
\end{aligned}$$

Now from the central limit theorem [assumptions (A_2) and (A_3)] we have $\overline{T}_N = \mu + O_p(Q^{-1/2})$, and since

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

by invoking condition (iii) we get $b(\overline{T}_N - \mu)^2 \to_p 0$.

From (A_2) and (iii) we get

$$b(\overline{T}_N - \mu)\left(\mu - \frac{1}{\sqrt{b}} E\tilde{B}_i\right) = o_p(bQ^{-1}) \to_p 0,$$

$$b\left(\mu - \frac{1}{\sqrt{b}} E\tilde{B}_i\right)^2 = o(bQ^{-1}) \to 0.$$

Gathering these three results yields $D_N \to_p 0$.

Now by reasoning as in Lemma 1, 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

$$(6) \qquad \alpha_{\tilde{B}, M, L}(n) \le \alpha_X([(n-1)h - (b-1)]L - M)$$

for $n \ge n_0 = [M/hL + (b-1)/h] + 1$. From conditions (i) and (ii) [namely, $M = O(L)$, $b = O(h)$], it is assured that there will be a smallest n_0 such that (6) will hold regardless of the value of N. Hence, for all practical purposes, all rows of the triangular array (i.e., for each N the sequence $T_{i, M, L}$, $i \in \mathbb{Z}$) can be treated as governed by the *same* mixing coefficient, namely, the right-hand side of (6).

Now,

$$\operatorname{Var} A_N = \frac{1}{q} \operatorname{Var}\left(\tilde{B}_1 - E\tilde{B}_1\right)^2 + \frac{2}{q^2} \sum_{i=1}^{q-1} (q-i) \operatorname{Cov}\left\{\left(\tilde{B}_1 - E\tilde{B}_1\right)^2, \left(\tilde{B}_{i+1} - E\tilde{B}_{i+1}\right)^2\right\}.$$

However,

$$\operatorname{Cov}\left\{\left(\tilde{B}_1 - E\tilde{B}_1\right)^2, \left(\tilde{B}_{i+1} - E\tilde{B}_{i+1}\right)^2\right\} \le 10\left(E\left|\tilde{B}_1 - E\tilde{B}_1\right|^{2p}\right)^{2/p} \left(\alpha_{\tilde{B}, M, L}(i)\right)^{(p-2)/p}$$

[cf. Roussas and Ioannides (1987), page 109].

Also, under condition (iv) and (A_1) the following moment inequality [Yokoyama (1980) and Roussas (1988)] holds:

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

where K_X depends only on α_X and p. Combining the above with assumption (A_1) yields

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

Now, from condition (iv) it also follows that $\sum_{k=1}^{\infty} (\alpha_X(k))^{(p-2)/p} < \infty$, since in assumption (A_1) it is assumed that $p \ge 3$. Thus, by the discussion relating $\alpha_{\tilde{B}, M, L}$ with α_X, it follows that $\operatorname{Var} A_N = O(1/q) \to 0$. Hence, by Chebyshev's inequality we get $A_N \to_p \sigma_\infty^2$ since by assumption (A_3) we have

$$EA_N = \operatorname{Var} \tilde{B}_1 = \operatorname{Var}\left(\sqrt{b}\, \frac{1}{b} \sum_{j=1}^{b} T_{j,M,L}\right) \to_p \sigma_\infty^2.$$

Finally, look at C_N. By the α-mixing property of the $\tilde{B}_i$ and by the same argument that showed $\operatorname{Var} A_N \to 0$, it can be shown that $\operatorname{Var}\{(1/q)\Sigma_{i=1}^{q}(\tilde{B}_i - E\tilde{B}_i)\} \to 0$ and hence $(1/q)\Sigma_{i=1}^{q}(\tilde{B}_i - E\tilde{B}_i) \to_p 0$. Also

$$\sqrt{b}\left(\overline{T}_N - (1/\sqrt{b}) E\tilde{B}_i\right) = \sqrt{b}\, O_p(Q^{-1/2}) = o_p(1) \quad \text{by condition (ii)}.$$

Therefore, $C_N \to_p 0$ and the proof is complete. □

REMARK 1. From the proof of Theorem 1 it is seen that assumption (A_2) is not needed in its full force. In particular, the following assumption can be substituted in place of assumption (A_2), and Theorem 1 would still be valid.

(A'_2) $ET_{1,M,L} = \mu + O(Q^{-1/2})$.

This is of quite some interest, since in problems where a trade-off of bias and variance of an estimator exists (cf. the spectral density example in Section 5), an optimal estimator from the point of view of mean squared error (MSE) would satisfy (A'_2) but not (A_2).

REMARK 2. It also can be shown [cf. Politis (1990)] that the requirement $b = O(h)$ can be dropped from condition (ii) of Theorem 1. In particular, the variance of $\hat{V}_{\text{JACK}}$ is of order $O(b/Q)$, regardless of choice of h. Hence, the following corollary of Theorem 1 is true.

COROLLARY 1. *Under assumptions* (A_0), (A_1), (A'_2) *and* (A_3) *and if*:

(i) $M = o(N)$ *and* $L \sim aM$, *for some* $a \in (0, 1]$,
(ii′) $b \to \infty$ *and* $h = O(b)$,
(iii) $b = o(Q)$,
(iv) $\Sigma_{k=1}^{\infty} k^{p-1}(\alpha_X(k))^{\delta/(2p+\delta)} < \infty$,

then

$$\hat{V}_{\text{JACK}}\left(\sqrt{Q}\, \overline{T}_N\right) \to_p \sigma_\infty^2. \tag{7}$$

A very important implication of Corollary 1 is that h can be taken to be a *fixed* constant and is not required to tend to infinity. Intuitively, the choice of h influences the constant factor in $\operatorname{Var}(\hat{V}_{\text{JACK}}) = O(b/Q)$, and it is advisable to let $h = 1$. For the special case of example (I), where $\overline{T}_N$ is the sample mean, it has been shown [cf. Künsch (1989) and Brillinger (1975)] that letting $h = 1$ corresponds to a 33% reduction of $\lim(Q/b)\operatorname{Var}(\hat{V}_{\text{JACK}})$ over letting $h = b$.

Another important observation is that if μ is a parameter of a finite-dimensional marginal distribution of sequence $\{X_n\}$, say the m-dimensional marginal, then M could be taken to be a *fixed* constant equal to m, provided the estimator $\overline{T}_N$ is designed so that it satisfies assumption (A'_2). In the autocovariance example (II), the choice $M = s + 1$ makes $\overline{T}_N$ exactly unbiased for μ.

Moreover, if M is fixed, then the choice $L = 1$ is also permissible [cf. condition (i)]. However, in the general case where $M \to \infty$, condition $L \sim aM$ is necessary for the consistency of $\hat{V}_{\mathrm{JACK}}$ (see also the discussion after Theorem 3 in Section 5).

Finally, observe that condition (iv) is satisfied if any one of the following three holds:

(v) The X_n's are m-dependent, that is, $\alpha_X(k) = 0$, $\forall\ k > m$.
(vi) $\alpha_X(\cdot)$ has an exponential decay.
(vii) $\alpha_X(k) = O(k^{-\lambda})$, where $\lambda > p(2p + \delta)/\delta$.

Examples of Gaussian processes satisfying one of the above mixing conditions are given in Ibragimov and Rozanov (1978).

Note that if we assume the mixing condition (vii) for some sufficiently large λ, assumption (A_3) can be omitted from Corollary 1, since it is a consequence of Tikhomirov's (1980) results on the central limit theorem [see also Rosenblatt (1984)]. However, the existence of a common asymptotic variance for all rows of the array is needed. So let us formulate the following assumption:

(A'_3) If $Q \to \infty$ as $N \to \infty$, $\lim_{N \to \infty} \mathrm{Var}((1/\sqrt{Q})\Sigma_{i=1}^{Q} T_{i,M,L})$ exists and equals $\sigma_\infty^2 > 0$.

Corollary 2. *Under assumptions* (A_0), (A_1), (A'_2) *and* (A'_3) *and conditions* (i), (ii′), (iii) *and* (vii), *we have*

$$\hat{V}_{\mathrm{JACK}}\left(\sqrt{Q}\,\overline{T}_N\right) \to_p \sigma_\infty^2.$$

Proof. We just need to verify assumption (A_3). From Tikhomirov (1980) this follows if $\lambda > 6$, since $E|T_{i,M,L}|^3 < \infty$. However, this is satisfied because $p(2p + \delta)/\delta > 6$. □

4. Blocks of blocks bootstrap. The blocks of blocks bootstrap is defined as follows: Sampling with replacement from the set $\{\mathscr{B}_1, \dots, \mathscr{B}_q\}$ defines (conditionally on the original observations $X_1, \dots, X_N$) a probability measure denoted by P^*. Let $Y_1, \dots, Y_k$ be i.i.d. samples from P^*. Obviously each Y_i is a block of size b which is denoted by $Y_i = (y_{i1}, \dots, y_{ib})$. Concatenate the y_{ij} in one big block of size $l = kb$ which will be called $T_1^*, \dots, T_l^*$, where $T_i^* = y_{rv}$, for $r = [i/b]$, $v = i - br$.

A natural assumption is as follows:

(A_4) Q and l are of the same asymptotic order, that is, $l = O(Q)$ and $Q = O(l)$.

The blocks of blocks bootstrap approximation to the sampling distribution of $\overline{T}_N$ is provided by the following theorem.

THEOREM 2. *Under assumptions* (A_0), (A_1), (A_2), (A_3) *and* (A_4) *and if*:

(i) $M = o(N)$ *and* $L \sim aM$, *for some* $a \in (0, 1]$,
(ii′) $b \to \infty$ *and* $h = O(b)$,
(iii′) $b = o(Q^{1/2})$,
(iv) $\sum_{k=1}^{\infty} k^{p-1}(\alpha_X(k))^{\delta/(2p+\delta)} < \infty$,

then

$$\sup_x \left| P^*\left\{\sqrt{l}\left(\overline{T}_l^* - \overline{T}_N\right) \le x\right\} - P\left\{\sqrt{Q}\left(\overline{T}_N - \mu\right) \le x\right\}\right| \to_p 0. \tag{8}$$

PROOF. First note that (A_4) along with $b = o(Q^{1/2})$ implies that $k \to \infty$. Now, by (A_2) and (A_3), it follows that

$$\sup_x \left| P\left\{\sqrt{Q}\left(\overline{T}_N - \mu\right) \le x\right\} - \Phi(x/\sigma_\infty)\right| \to 0.$$

So, to prove the theorem, it suffices to prove the following three statements:

$$\sup_x \left| P^*\left\{\frac{\overline{T}_l^* - E^*\overline{T}_l^*}{\mathrm{Var}^*\ \overline{T}_l^*} \le x\right\} - \Phi(x)\right| \to_p 0, \tag{I}$$

$$l\ \mathrm{Var}^*\ \overline{T}_l^* \to_p \sigma_\infty^2, \tag{II}$$

$$E^*\overline{T}_l^* = \overline{T}_N + o_p(l^{-1/2}), \tag{III}$$

where E^* and Var^* stand for E and Var under the bootstrap probability P^*.

Define $\tilde{Y}_i = (1/\sqrt{b})\sum_{j=1}^{b} y_{ij}$. Then $\sqrt{l}(\overline{T}_l^* - E^*\overline{T}_l^*) = \sqrt{k}\sum_{i=1}^{k}(\tilde{Y}_i - E^*\tilde{Y}_i)$. Now the $\tilde{Y}_i$ are i.i.d. under P^*. Therefore by the Berry–Esseen theorem it is seen that (I) holds provided $E^*|\tilde{Y}_1 - E^*\tilde{Y}_1|^3$ is bounded in probability. However,

$$E^*\left|\tilde{Y}_1 - E^*\tilde{Y}_1\right|^3 = \frac{1}{q}\sum_{i=1}^{q}\left|\tilde{B}_i - \frac{1}{q}\sum_{j=1}^{q}\tilde{B}_j\right|^3,$$

which by the triangle inequality is bounded above by

$$\frac{1}{q}\left\{\left(\sum_{i=1}^{q}|\tilde{B}_i|^3\right)^{1/3} + q^{-2/3}\left|\sum_{j=1}^{q}\tilde{B}_j\right|\right\}^3.$$

This quantity converges in probability to $\{(E|\tilde{B}_1|^3)^{1/3} + |E\tilde{B}_1|\}^3$, because a weak law of large numbers holds for the α-mixing sequences $\tilde{B}_i$ and $|\tilde{B}_i|^3$, similarly as in the proof of Theorem 1. So (I) is proved.

Now look at

$$l\ \mathrm{Var}^*\left(\overline{T}_l^*\right) = l\ \mathrm{Var}^*\left\{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{1}{b}\sum_{j=1}^{b} y_{ij}\right)\right\} = l\ \mathrm{Var}^*\left\{\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\sqrt{b}}\tilde{Y}_i\right\} = \mathrm{Var}^*\ \tilde{Y}_1$$

because the $\tilde{Y}_i$ are i.i.d. and $l = kb$. However,

$$\mathrm{Var}^*\ \tilde{Y}_1 = \frac{1}{q}\sum_{i=1}^{q}\left(\tilde{B}_i - E^*\tilde{Y}_1\right)^2,$$

where

$$E^*\tilde{Y}_1 = \frac{1}{q}\sum_{i=1}^{q}\tilde{B}_i = \frac{1}{q}\sum_{i=1}^{q}\sum_{j=(i-1)h+1}^{(i-1)h+b}\frac{1}{\sqrt{b}}T_{j,M,L}.$$

In this double sum, each $T_{j,M,L}$ for $b \le j \le Q - b$ is represented exactly $[b/h + 1]$ times, because of the overlapping $\mathscr{B}_j$ blocks. Taking into account that each $T_{j,M,L} = O_p(1)$ (by Chebyshev's inequality) yields

$$E^*\tilde{Y}_1 = \frac{1}{q\sqrt{b}}\left[\frac{b}{h} + 1\right]\sum_{i=1}^{Q}T_{i,M,L} - \frac{1}{q\sqrt{b}}O_p\left(2b\left[\frac{b}{h}\right]\right).$$

Recalling that $q = [(Q - b)/h] + 1$ and that $b = o(Q)$ yields

$$\left|E^*\tilde{Y}_1 - \sqrt{b}\,\overline{T}_N\right| = O_p\left(\frac{b^{3/2}}{hq}\right) = O_p\left(\frac{b^{3/2}}{Q}\right).$$

Therefore

$$\operatorname{Var}^*\tilde{Y}_1 = \frac{1}{q}\sum_{i=1}^{q}\left(\tilde{B}_i - \sqrt{b}\,\overline{T}_N + O_p\left(\frac{b^{3/2}}{Q}\right)\right)^2;$$

in view of the fact that from Corollary 1 we have

$$\hat{V}_{\mathrm{JACK}}\left(\sqrt{Q}\,\overline{T}_N\right) = \frac{1}{q}\sum_{i=1}^{q}\left(\tilde{B}_i - \sqrt{b}\,\overline{T}_N\right)^2 \to_p \sigma_\infty^2$$

and condition (iii′), we have proved (II).

Finally, look at

$$E^*\overline{T}_l^* = E^*\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\sqrt{b}}\tilde{Y}_i = \frac{1}{\sqrt{b}}E^*\tilde{Y}_1 = \overline{T}_N + O_p\left(\frac{b}{Q}\right)$$

by our previous discussion. By using now condition (iii′) [namely, $b = o(\sqrt{Q})$] and in view of the fact that l and Q are assumed of the same order, (III) is proved and so is the theorem. □

REMARK 3. Assumption (A_4) was only used at the very last part of the proof. It is easily seen that Theorem 2 is still true if we substitute (A_4) with the more technical assumption that $k \to \infty$ and $l = o(Q^2/b^2)$.

REMARK 4. It is easy to infer that under the assumptions of Theorem 2 the following result is also true:

$$\text{(9)} \qquad \sup_x\left|P^*\left\{\sqrt{l}\left(\overline{T}_l^* - E^*\overline{T}_l^*\right) \le x\right\} - P\left\{\sqrt{Q}\left(\overline{T}_N - \mu\right) \le x\right\}\right| \to_p 0.$$

Practical use of this result involves an increased computational effort, in order to calculate $E^*\overline{T}_l^*$. Nevertheless, there is an additional by-product of this extra effort, namely, that in this form [equation (9)] the blocks of blocks bootstrap estimate of sampling distribution is potentially a more accurate distribution estimate than the one offered by the central limit theorem. This claim was proven [cf. Lahiri (1990)] in the special case of the sample mean, in which the blocks of blocks and the moving blocks methods coincide.

The sample mean result, with little modification, can be applied to show that under some additional regularity conditions (including an exponential mixing rate and a Cramér-type condition on the distributions), the blocks of blocks bootstrap approximation of (9) is accurate to more than first order, in the case where μ is a parameter of a finite, say, m-dimensional, distribution of the $\{X_n\}$ process. In that case, M is taken equal to m, L is taken equal to 1, and all rows of the triangular array of the $T_{i,M,L}$, $i \in \mathbb{Z}$, are identical and therefore can be treated as just one new stationary sequence $\{T_i,\ i \in \mathbb{Z}\}$ [see also Künsch (1989)].

The above discussion is more precisely formulated in Lemma 2. Assuming that the sequence $\{X_n,\ n \in \mathbb{Z}\}$ is defined on the probability space $(\Omega, \mathscr{A}, P)$, denote $\mathscr{D}_n$, $n \in \mathbb{Z}$, a sequence of sub-σ-fields of $\mathscr{A}$, and $\mathscr{D}_{n_1}^{n_2}$ the σ-field generated by $\mathscr{D}_{n_1}, \dots, \mathscr{D}_{n_2}$.

LEMMA 2. *Suppose μ is a parameter of the m-dimensional joint distribution of the sequence $\{X_n,\ n \in \mathbb{Z}\}$, with m finite. Also assume that $ET_1 = \mu$ and $E|T_1|^4 < \infty$, where $T_i \equiv T_{i,M,L}$, for $M = m$, $L = 1$. Under assumptions* (A_0), (A'_3) *and* (A_4), *conditions* (ii′), (iii′) *and* (vi) *and the following three additional conditions*:

(a_1) *$\exists\ d > 0$ such that, for all $k, n \in \mathbb{N}$, with $n > 1/d$, there exists a $\mathscr{D}_{k-n}^{k+n}$-measurable random variable $Z_{k,n}$, for which $E|T_k - Z_{k,n}| \le d^{-1}e^{-dn}$ and $E|Z_{k,n_k}|^{12} I(|Z_{k,n_k}| < k^{1/4}) < d^{-1}$, where n_k is a sequence of real numbers satisfying $\log k = o(n_k)$ and $n_k = O(\log k)^{1+d^{-1}}$, as $k \to \infty$*;

(a_2) *$\exists\ d > 0$ such that, for all $k, n \in \mathbb{N}$, with $k > n > 1/d$, and for all $t > d$,*

$$E\big|E\big(\exp[jt(T_{k-n} + T_{k-n+1} + \cdots + T_{k+n})]\,\big|\,\mathscr{D}_i, i \ne k\big)\big| \le e^{-d},$$

where j is used to denote the imaginary unit $\sqrt{-1}$;

(a_3) *$\exists\ d > 0$ such that, for all $k, n_1, n_2 \in \mathbb{N}$ and $A \in \mathscr{D}_{n_1-n_2}^{n_1+n_2}$,*

$$E\big|P(A|\mathscr{D}_i, i \ne n_1) - P(A|\mathscr{D}_i, 0 < |n_1 - i| \le k + n_2)\big| \le d^{-1}e^{-dk};$$

the following is true:

$$\sup_x \left| P^*\left\{\sqrt{l}\,\frac{\overline{T}_l^* - E^*\overline{T}_l^*}{\sqrt{\operatorname{Var}^*(\sqrt{l}\,\overline{T}_l^*)}} \le x\right\} - P\left\{\sqrt{Q}\,\frac{\overline{T}_N - \mu}{\sigma_\infty} \le x\right\}\right| = o_p(Q^{-1/2}). \tag{10}$$

The proof is an immediate application of Lahiri's (1990) result to the stationary sequence $\{T_i,\ i \in \mathbb{Z}\}$. Equation (10) can be strengthened to hold with probability 1 if it is additionally assumed that $b = o(Q^{1/4})$ [which is not desirable because it can be shown, cf. Politis (1990), that the MSE of the blocks of blocks jackknife or bootstrap estimate of variance is asymptotically minimized by taking $b \sim A_b Q^{1/3}$, with a resulting MSE of order $Q^{-2/3}$], or by assuming a stronger moment condition [cf. Lahiri (1990)].

However, the applicability of Lemma 2 is limited by the fact that conditions (a_1)–(a_3) are quite difficult to verify [cf. Götze and Hipp (1983)] in specific settings. Furthermore, a corresponding result for the general case, where $m = \infty$, seems intractable at this point. It should be noted that analogs of both Lemma 2 and Theorem 2 are valid in a multivariate setting as well, that is, where $\overline{T}_N$ takes values in $\mathbb{R}^D$ [cf. Lahiri (1990) and Politis and Romano (1992)].

Returning to the general case, it is easily seen from the proof of Theorem 2 that assumption (A_2) was used only to ensure that $P\{\sqrt{Q}(\overline{T}_N - \mu) \le x\} \to \Phi(x/\sigma_\infty)$. If it is replaced by the weaker (A'_2), the following result remains true.

COROLLARY 3. *Under assumptions* (A_0), (A_1), (A'_2), (A_3) *and* (A_4) *and if*:

(i) $M = o(N)$ *and* $L \sim aM$, *for some* $a \in (0, 1]$,
(ii′) $b \to \infty$ *and* $h = O(b)$,
(iii′) $b = o(Q^{1/2})$,
(iv) $\sum_{k=1}^{\infty} k^{p-1}(\alpha_X(k))^{\delta/(2p+\delta)} < \infty$,

then

$$(11) \quad \sup_x \left| P^*\left\{\sqrt{l}\left(\overline{T}_l^* - E^*\overline{T}_l^*\right) \le x\right\} - P\left\{\sqrt{Q}\left(\overline{T}_N - E\overline{T}_N\right) \le x\right\}\right| \to_p 0$$

as well as

$$(12) \quad \sup_x \left| P^*\left\{\sqrt{l}\left(\overline{T}_l^* - \overline{T}_N\right) \le x\right\} - P\left\{\sqrt{Q}\left(\overline{T}_N - E\overline{T}_N\right) \le x\right\}\right| \to_p 0.$$

REMARK 5. The result of Corollary 3 allows for the construction of bootstrap confidence intervals for $E\overline{T}_N$, in the case where $\overline{T}_N$ is an estimator with bias of the same order as its standard deviation. In that case, even the central limit theorem of assumption (A_3) can only provide asymptotic confidence intervals for $E\overline{T}_N$ and not for μ. To obtain confidence intervals for μ using either the blocks of blocks bootstrap or the central limit theorem, an adjustment for the bias must be made via an expansion of the form $E\overline{T}_N = \mu + \mu_1 + o(Q^{-1/2})$, provided μ_1 can itself be estimated with $o_p(Q^{-1/2})$ accuracy.

REMARK 6. Using the δ-method, it is immediately seen that both the jackknife and the bootstrap remain valid for statistics of the form $g((1/Q)\sum_{i=1}^{Q} T_{i,M,L})$ as long as the function g has a nonzero derivative at μ. This is the statement in Bickel and Freedman (1981) that "the bootstrap commutes with smooth functions." An example of such a function that we will use later on is $g(x) = \log x$, for $x > 0$.

REMARK 7. Note that in the construction of the jackknife and bootstrap we are effectively resampling whole *blocks* (of size b) *of blocks* (of size M) of the original observations. Equivalently, this can be thought of as resampling *bigger* blocks of size $(b-1)L+M$ of the observations X_n, but then applying to them your estimating procedure based on a window of width M moving at lags hL at a time. In Künsch (1989) and in Liu and Singh (1988), because they consider $L=1$, $h=1$ and $M=1$ or at best $M=m$, a fixed number, this distinction does not appear. It should be pointed out, though, that this equivalence holds only up to some asymptotically negligible "edge" effects (at the places where the bigger blocks are joined together in forming a resampled sequence). However, in a finite sample situation these edge effects could be important, and it is advisable to eliminate them. The blocks of blocks construction does eliminate the edge effects, as well as making the whole procedure more transparent, since by separating the whole construction into two stages,

1. blocking to get a consistent estimator,
2. blocking to get a valid bootstrap or jackknife procedure,

we have greater freedom to fine-tune our design parameters. Choosing the design parameters in the estimation and in the bootstrap procedures independently is, of course, of great practical value and is also an attribute of the procedures in Härdle and Bowman (1988) and Franke and Härdle (1992).

As before, we also have the following.

COROLLARY 4. *In the hypotheses of Theorem* 2 *and Corollary* 3, *condition* (vii) *can be substituted for condition* (iv), *and assumption* (A'_3) *can be substituted for assumption* (A_3), *and the respective results of Theorem* 2 *and Corollary* 3 *will hold true.*

5. Approximate confidence intervals for the spectral density. The concrete application which in fact was the motivation for our previous abstract discussion is the following. In this section assume that the time series X_t is univariate. Let

$$T_{i,M,L}(w) = \frac{1}{2\pi M}\left|\sum_{t=L(i-1)+1}^{L(i-1)+M} W_t X_t e^{-jtw}\right|^2,$$

that is, $T_{i,M,L}$ is the periodogram of block B_i of data, "tapered" by the function W_t, evaluated at some frequency $w \in [-\pi,\pi]$. Note that the symbol j denotes the imaginary unit $\sqrt{-1}$, to avoid confusion with i, the block count. Also define $\overline{T}_N = (1/Q)\sum_{i=1}^{Q} T_{i,M,L}(w)$ as before. It can be shown [Zhurbenko (1980)] that $\overline{T}_N$ is a consistent estimator of the spectral density function $f(w)$ which is defined by $f(w) = (1/2\pi)\sum_{s=-\infty}^{\infty} R(s)e^{-jsw}$, where $R(s) = EX_t X_{t+|s|}$ is the autocovariance (for simplicity, assume $EX_t = 0$).

Estimators of this type are called lag-window spectral estimators and were previously considered by Bartlett (1946, 1950), Welch (1967), Brillinger (1975), Zhurbenko (1979, 1980) and Thomson and Chave (1988). It is interesting to point out the close connection of the lag-window estimators with the more commonly used kernel smoothed estimators that were pioneered by Daniell (1946), Grenander and Rosenblatt (1957), Blackman and Tukey (1959) and Parzen (1961). This connection actually led Bartlett to introduce the triangular kernel for smoothing that now bears his name. To be specific let us take $W_t = 1$, $\forall\ t$ (no tapering). Then it can be shown [cf. Priestley (1981)] that if $\hat{R}(s) = (1/N)\sum_{k=1}^{N-s} X_k X_{k+s}$ is the usual sample autocovariance,

$$\overline{T}_N \simeq \hat{f}(w) = \frac{1}{2\pi} \sum_{s=-M}^{M} \left(1 - \frac{|s|}{M}\right) \hat{R}(s) e^{-jws},$$

where $\hat{f}(w)$ is the kernel smoothed (with Bartlett's kernel) estimator. By computing a tapered periodogram for each block (with an appropriate choice of data window W_t), we can obtain a correspondence with other kernel estimators (not just Bartlett's).

For concreteness we will continue our discussion in this case (with $W_t = 1$), although everything can be generalized to the case that we have tapering. Under regularity conditions [namely, that $f(w)$ is Lipschitz and the fourth-order spectral density is bounded, which can be ensured by a condition on the mixing coefficient], it can be shown [cf. Zhurbenko (1979, 1980)] that, for $M = o(N)$ as $N \to \infty$, the estimator $\overline{T}_N$ is asymptotically normal, with asymptotic variance given by

$$\operatorname{Var}(\overline{T}_N) \sim c \frac{M}{N} f^2(w)(1 + \eta(w)), \tag{13}$$

where $\eta(w) = 0$ if $w \neq 0 \pmod{\pi}$ and $\eta = 1$ if $w = 0, \pm\pi$. If $L \sim aM$, the constant c can be explicitly calculated as $c = a + 2a\sum_{k=1}^{[1/a]}(1 - ka)^2$ [cf. Welch (1967)]; if $L = o(M)$, we have $c = 2/3$. Thus the existence of an asymptotic variance [cf. assumption (A'_3)] is ensured.

It is to be noted that a sufficient condition for the spectral densities up to order ρ to exist and be bounded is $\alpha_X(k) = O(k^{-\lambda})$, where $\lambda > (\rho + 1)((v - 2)/v)$, and $E|X_1|^{\rho v} < \infty$, for some $v > 2$ [cf. Zhurbenko (1980)]. Hence, by taking $v = 2 + \varepsilon$, with ε small enough, the asymptotic formula (13) holds under condition (vii) and if $E|X_1|^{10} < \infty$.

Let us now check the conditions that will enable us to apply our theorems in this setting. Using Yokoyama's (1980) theorem and condition (iv), it is easy to see that assumption (A_1) will hold, provided $E|X_1|^{2r+\varepsilon} < \infty$, for $r = 2p + \delta$, and some $\varepsilon > 0$.

Now, to check assumption (A_2), it is well known [cf. Hannan (1970)] that if $f \in \mathscr{C}^1$ or if it is just Lipschitz near w, then the bias of the periodogram of a series of M observations is $O(\log M/M)$. But a sufficient condition for $f \in \mathscr{C}^1$ is that $\sum_s |s|\,|R(s)| < \infty$, which is ensured by condition (vii) and a moment assumption.

Therefore, $ET_{1,M,L}(w) = f(w) + O((\log M)/M)$, and in order to have $O((\log M)/M) = o(\sqrt{M/N})$, it is required that $N/M^3 \to 0$. If, as usual, we put $M \sim AN^{\beta}$, then it is required that $\beta > 1/3$. This is what we referred to earlier on as *undersmoothing*, since for the particular case of the Bartlett estimator the choice $\beta = 1/3$ makes the asymptotic order of the squared bias equal to the order of the variance, thus minimizing (asymptotically) the MSE. Note that if, with appropriate tapering, we succeed in bringing the bias down to $O(M^{-q})$, then to get assumption (A_2) to hold we would need $\beta > 1/(1+2q)$, which is undersmoothing in the general case.

By the above observations and using Corollaries 2 and 4 the following result is established.

THEOREM 3. *Suppose $M \sim AN^{\beta}$, $\beta > 1/3$. Under assumptions* (A_0) *and* (A_4) *and conditions* (i), (ii′), (iii′) *and* (vii), *and under the additional condition*:

(viii) $E|X_1|^{2r+\varepsilon} < \infty$, *for some* $\varepsilon > 0$, *and for* $r = 2p + \delta$, *where* $p > 2$ *is some integer, and* $\delta \in (0, 2]$ *is some real constant*,

the following are true (*where* c_0 *is some positive constant*):

$$\sqrt{Q}\left(\overline{T}_N - f(w)\right) \Rightarrow_{\mathscr{L}} N\big(0, c_0 f^2(w)(1+\eta(w))\big); \tag{14}$$

$$\sup_x \left| P^*\left\{\sqrt{l}\left(\overline{T}_l^* - E\overline{T}_l^*\right) \le x\right\} - P\left\{\sqrt{Q}\left(\overline{T}_N - f(w)\right) \le x\right\}\right| \to_p 0; \tag{15}$$

$$\sup_x \left| P^*\left\{\sqrt{l}\left(\overline{T}_l^* - \overline{T}_N\right) \le x\right\} - P\left\{\sqrt{Q}\left(\overline{T}_N - f(w)\right) \le x\right\}\right| \to_p 0; \tag{16}$$

$$\hat{V}_{\text{JACK}}\left(\sqrt{Q}\,\overline{T}_N\right) \to_p c_0 f^2(w)(1+\eta(w)). \tag{17}$$

In addition, if $M \sim AN^{1/3}$, then $E\overline{T}_N$ should be substituted instead of $f(w)$ in the left-hand sides of equations (14)–(16).

The case where $M \sim AN^{1/3}$ deserves special attention, because it corresponds to a $\overline{T}_N$ estimator that has asymptotically minimum mean squared error. As in Remark 5 (after Corollary 3), in that case (14)–(16) would only provide confidence intervals for $E\overline{T}_N$ and not for $f(w)$. However, an expansion for the bias of $\overline{T}_N$ is readily available, by the approximate equality of $\overline{T}_N$ and Bartlett's $\hat{f}(w)$ [cf. Priestley (1981)], namely, $f(w) = E\overline{T}_N + \mu_1 + o(Q^{-1/2})$, where $\mu_1 = (1/M)\Sigma_{s=-\infty}^{s=\infty}|s|R(s)e^{-jsw}$ is the "generalized" first derivative of $f(w)$, [cf. Parzen (1961)]. By using an estimator $\hat{\mu}_1$, such that $\hat{\mu}_1 - \mu_1 = o_p(Q^{-1/2})$, (14)–(16), with $f(w) - \hat{\mu}_1$ put in place of $f(w)$ in their respective left-hand sides, would yield asymptotically valid confidence intervals for $f(w)$. This sort of analytic bias correction in the confidence intervals is also present in the work of Härdle and Bowman (1988) and Franke and Härdle (1992).

It is also important to consider the two cases, $L \sim aM$ and $L = o(M)$, in order to explain the assumption of condition $L \sim aM$ in all our asymptotic results. From (13), it is apparent that the estimator $\overline{T}_N$ [with $L = o(M)$] has a smaller asymptotic variance than the estimator $\overline{T}_N$ (with $L \sim aM$). However, note that in either case the variance of $\overline{T}_N$ is asymptotically proportional to M/N and that the gain in taking $L = o(M)$ over taking $L \sim aM$ is minor, since by letting a be small enough, we can bring the constant c arbitrarily close to $2/3$.

Moreover, it easy to see that the blocks of blocks bootstrap and jackknife estimates of the variance of $\overline{T}_N$ turn out to be proportional to $1/Q$, which is of the same order as M/N *if and only if* $L \sim aM$ (recall that $Q \sim N/L$). Hence, the blocks of blocks bootstrap and jackknife estimates are simply inconsistent (by an order of magnitude!) in the case $L = o(M)$, M tending to infinity. The reason is that the whole blocking idea is crucially based on having *weakly* dependent data [see also Künsch (1989), page 1225]. The $T_{i,M,L}$ "pseudodata" are weakly dependent if $L \sim aM$, and strongly dependent if $L = o(M)$.

The result of Theorem 3 can be taken one step further by an approximate studentization, since typically [cf. Babu and Singh (1983), Hall (1988) and DiCiccio and Romano (1988)] this leads to a faster rate of convergence. Since we know that $\overline{T}_N \to_p f(w)$, an application of Slutsky's theorem shows that, if $f(w) \neq 0$, (16) also implies

$$\sup_x \left| P^*\left\{ \sqrt{l}\, \frac{\overline{T}_l^* - \overline{T}_N}{\overline{T}_N} \le x \right\} - P\left\{ \sqrt{Q}\, \frac{\overline{T}_N - f(w)}{f(w)} \le x \right\} \right| \to_p 0. \tag{18}$$

We will prefer to use this version for setting confidence intervals because under the hypotheses of Theorem 3 the quantity $\sqrt{Q}\,(\overline{T}_N - f(w))/f(w)$ is asymptotically pivotal since $\sqrt{Q}\,(\overline{T}_N - f(w))/f(w) \Rightarrow_{\mathscr{L}} N(0, c_0)$, where c_0 does not depend on $f(w)$.

Incidentally, there is a close connection between this studentization and the logarithmic transformation. The logarithmic transformation is suggested (by the δ-method) in order to stabilize the variance in models such as this, where the standard deviation is approximately proportional to the mean. In particular, after some algebra and using the fact that $\log(1 + x/N) - x/N \to 0$, it can be shown that (16) is equivalent to

$$\begin{aligned} \sup_x \Big| & P^*\left\{ \sqrt{l}\left(\log \overline{T}_l^* - \log \overline{T}_N\right) \le x \right\} \\ & - P\left\{ \sqrt{Q}\left(\log \overline{T}_N - \log f(w)\right) \le x \right\} \Big| \to_p 0. \end{aligned} \tag{19}$$

This latter representation is more convenient to work with in practice, because it brings us in a location-parameter setting.

6. Approximate confidence intervals for the cross-spectral density. Assume now that the time series $\{X_t\}$ is bivariate, that is, $X_t = (X_{t,1}, X_{t,2})$, with $EX_t = 0$. Let

$$T_{i,M,L}(w) = \frac{1}{2\pi M}\left(\sum_{t=L(i-1)+1}^{L(i-1)+M} W_t X_{t,1} e^{-jtw} \right)\left(\sum_{t=L(i-1)+1}^{L(i-1)+M} W_t X_{t,2} e^{+jtw} \right),$$

that is, $T_{i,M,L}$ is the cross-periodogram of block B_i of data, tapered by the function W_t, evaluated at some frequency $w \in [-\pi, \pi]$. Again, take $W_t = 1$ for concreteness. Also, define $\overline{T}_N = (1/Q)\sum_{i=1}^{Q} T_{i,M,L}(w)$ as before. The cross-spectral density function $f(w)$ is defined by $f(w) = (1/2\pi)\sum_{s=-\infty}^{\infty} R(s)e^{-jsw}$, where now $R(s) = EX_{t,1}X_{t+s,2}$ is the cross-covariance. However, it should be noted that, in general, the cross-spectral density is a complex function. Let us denote by $f_R(w)$ and $f_I(w)$ its real and imaginary parts and by $\overline{T}_N^R$, $\overline{T}_N^I$ the real and imaginary parts of $\overline{T}_N$.

Similarly as in the previous section it can be proved that $|\overline{T}_N|$, $\overline{T}_N^R$ and $\overline{T}_N^I$ have bias of order $O(\log M/M)$, under the mixing condition (vii). Hence, by ensuring that $M \to \infty$ and $M = o(N)$, we see that $|\overline{T}_N|$, $\overline{T}_N^R$ and $\overline{T}_N^I$ are consistent estimators of $|f(w)|$, $f_R(w)$ and $f_I(w)$, respectively.

The situation then would appear to resemble the previously discussed case of the (individual) spectral density, only in more complicated form. For example, under the strong assumption that X_t is Gaussian (and other regularity assumptions), it can be proved [Jenkins (1963) and Zhurbenko (1986)] that, for some constant c,

$$\operatorname{Var}(|\overline{T}_N|) \sim c\frac{M}{N}|f(w)|^2\left(1 + \frac{f_1(w)\,f_2(w)}{|f(w)|^2}\right)(1 + \eta(w)), \tag{20}$$

where $f_1(w)$ and $f_2(w)$ are the individual spectral densities of $X_{t,1}$ and $X_{t,2}$, respectively. Similar formulae [involving $f(w)$, $f_1(w)$ and $f_2(w)$] also exist for the asymptotic variances of $\overline{T}_N^R$ and $\overline{T}_N^I$.

In view of the complicated form of the asymptotic variances, it seems that the bootstrap and jackknife represent a most practical way to assess the statistical accuracy of the cross-spectral density statistics $|\overline{T}_N|$, $\overline{T}_N^R$ and $\overline{T}_N^I$. We can formulate the following theorem, which like Theorem 3 is based on our corollaries.

THEOREM 4. *Suppose* $M \sim AN^{\beta}$, $\beta > 1/3$. *Under assumptions* (A_0) *and* (A_4), *conditions* (i), (ii′), (iii′) *and* (vii) *and under the additional condition*:

(ix) $E|X_{t,1}|^{2r+\varepsilon} < \infty$, $E|X_{t,2}|^{2r+\varepsilon} < \infty$, *for some* $\varepsilon > 0$, *and for* $r = 2p + \delta$, *where* $p > 2$ *is some integer and* $\delta \in (0, 2]$ *is some real constant*,

the estimators $|\overline{T}_N|$, $\overline{T}_N^R$ *and* $\overline{T}_N^I$ *are asymptotically normal, and their respective asymptotic variance and sampling distribution can be estimated consistently via the blocks of blocks jackknife and bootstrap procedures, analogously to Theorem* 3.

As in the discussion after Theorem 3, analytical corrections for the bias can also be employed in order to work with mean squared error optimal estimators, in which $M \sim AN^{1/3}$.

In addition, recalling the fact that "the bootstrap commutes with smooth functions" (or from the discussion at the end of the previous section), it is seen that the theorem also implies

$$\begin{aligned}\sup_x \Big| & P^*\left\{\sqrt{l}\left(\log|\overline{T}_l^*| - \log|\overline{T}_N|\right) \le x\right\} \\ & - P\left\{\sqrt{Q}\left(\log|\overline{T}_N| - \log|f(w)|\right) \le x\right\}\Big| \to_p 0,\end{aligned} \tag{21}$$

provided $f(w) \ne 0$. We would again prefer to work with the above version because, in view of the asymptotic formula for $\mathrm{Var}(|\overline{T}_N|)$, the quantity $\sqrt{Q}(|\overline{T}_N| - |f(w)|)/|f(w)|$ would still be approximately pivotal in case the squared coherency $|f(w)|^2/f_1(w)f_2(w)$ is approximately constant along w. Notably if $X_{t,2}$ is the output of a linear time-invariant filter with input $X_{t,1}$, the squared coherency is identically equal to 1.

Estimating the coherency function $\psi(w) = |f(w)|/\sqrt{f_1(w)f_2(w)}$ is itself of some interest. A consistent estimator can be formed if lag-window estimates of the modulus of the cross-spectral density and of the f_1 and f_2 spectral densities are substituted for $|f(w)|$, $f_1(w)$ and $f_2(w)$ in the definition of $\psi(w)$. As a smooth function of the lag-window estimates, this estimator, [call it $\overline{\psi}(w)$], would be asymptotically normal, and its asymptotic variance and sampling distribution could be estimated consistently via the blocks of blocks resampling scheme. Note also that Fisher's z-transformation is applicable here [cf. Priestley (1981)] and could be used to derive confidence intervals for $\psi(w)$ through the asymptotic normal distribution of the variance stabilized statistic $\tanh^{-1}(\overline{\psi}(w))$. If the blocks of blocks bootstrap method were used, variance stabilization would not affect the asymptotic first-order accuracy, but it is likely to improve the rate of the asymptotic approximation, analogously to the increased accuracy of the bootstrap approximation of the sampling distribution of a *studentized* sample mean [see also Hall (1992) and Hall, Martin and Schucany (1989)].

7. Some practical considerations. In this section we will discuss the practical implementation of the blocks of blocks resampling procedure in the spectral density problem. Two comments are in order.

The problem of bias. Although by undersmoothing the bias will be asymptotically negligible, in a finite-sample setting it usually is not, especially if the true spectrum has sharp and narrow-band peaks or troughs. This can be understood intuitively through the analogy with the kernel smoothed estimators in which it is obvious that a sharp peak of f at w_0 can significantly influence the value of $\hat{f}(w)$ for $w \in (w_0 - \Delta w, w_0 + \Delta w)$, thus introducing bias. Here Δw is of the order of π/M, since the Fourier transform of the Bartlett kernel is $(4M/w^2)\sin^2(w/2M)$.

Now observe that the bias in $\overline{T}_N(w)$ is due entirely to the bias of the periodogram of a block of M observations (cf. Section 5). By recalling that the bias of the periodogram of white noise is zero, the technique of *whitening* is suggested [cf. Brillinger (1975)] in order to alleviate the problem of bias in practice. This technique amounts to first fitting a parametric (usually autoregressive) model to the data and then filtering the data with the estimated whitening filter. The filtered data now would have an approximately constant (true) spectrum, and applying the blocks of blocks resampling procedure to them will give more accurate results because of reduced bias.

Choosing the design parameters. The asymptotic restrictions on M, L, b, h and $l = kb$ imposed by the assumptions and conditions of the theorems do not give practical guidelines for choosing these parameters in a finite-sample situation. However, we will describe some intuitive considerations that would lead to reasonable choices.

In practice, the choice of M and L is made in an exploratory fashion, by examining plots of the lag-window estimator for different choices. One can choose a combination of M and L that leads to a visibly undersmoothed estimator. At this stage one can also decide on whether it is necessary to perform a whitening procedure, by looking at the peaks and troughs of the estimator as a function of w. In addition, since the computation of periodograms is efficiently done by the fast Fourier transform (FFT) algorithm, then the choice of M also determines for *which* frequencies w the estimator will be evaluated. (Recall that the FFT algorithm computes a periodogram for frequencies on the discrete grid: $w_j = 2\pi j/M$, for $j = 1, \dots, M$.)

By the discussion after Corollary 1, we are justified in choosing $h = 1$. This is most desirable since the effective sample to be bootstrapped or jackknifed is of size $q \sim (Q - b)/h$. Also, to satisfy assumption (A_4), it is reasonable to let $k = [Q/b]$ or $k = [Q/b] + 1$, and then $l = kb$.

Regarding the problem of choosing b in a finite-sample situation, an interplay between the choice of M and L and that of b should be noted. It can be shown that to have a reasonable variance estimate provided by the bootstrap we should *at least* require $b > [1/a] \sim M/L$. Since each $\mathscr{B}_j$ consists of $(b - 1)L + M$ observations X_t, this requirement can be interpreted to read as follows: Each resampled block $\mathscr{B}_j$ should at least contain about $2M$ of the X_t observations. Of course in practice we cannot take b too large as well. An upper limit on b must also be set, which is reflected in the condition that $b = o(\sqrt{Q})$ of Theorem 2 or the $b \sim A_b Q^{1/3}$ dictated for MSE optimality.

Let us elaborate on this last point by using a numerical example. Suppose the sequence $\{Y_t, t \in \mathbb{Z}\}$ is actually m-dependent and the statistic in question is the sample mean. This case is discussed in detail in Liu and Singh (1988), with particular mention to the situation that, for all t, n, $\mathrm{Cov}(Y_t, Y_n) \geq 0$, in which case it is claimed that $\hat{V}_{\mathrm{JACK}}$ increases monotonically with the block size b and approaches the true value. However, this requires an infinite sample size. The following numerical example shows that in a finite-sample case, taking too large a block size might spoil the correction that the moving blocks jackknife offers versus the classical jackknife.

A sample $Y_1, \dots, Y_{100}$ was generated from the moving average model: $Y_t = Z_t + Z_{t-1} + Z_{t-2}$, where the Z_t's are i.i.d. $N(0, 1)$. By observing that $\Sigma_{i=1}^{100} Y_i \simeq 3\Sigma_{i=1}^{100} Z_i$, we see that $\mathrm{Var}(\Sigma_{i=1}^{100} Y_i/10) \simeq 9$. A plot of $\hat{V}_{\mathrm{JACK}}(\Sigma_{i=1}^{100} Y_i/10)$ as a function of the block size b is shown in Figure 1. It is seen that although $\hat{V}_{\mathrm{JACK}}$ increases and "captures" the true variance at about $b = 10$, taking a greater b *worsens* the approximation and, for $b \geq 20$, the moving blocks correction is totally lost. The part for $80 \leq b \leq 100$ is understandably bad, due to a very small equivalent sample to be jackknifed. Obviously, for $b = 100$ the equivalent sample is of size 1, which of course leads to a variance estimate of zero.

The variability of $\hat{V}_{\mathrm{JACK}}(\Sigma_{i=1}^{100} Y_i/10)$ as a function of b is an indication of the difficulty of estimation in problems with dependent data. The moving average

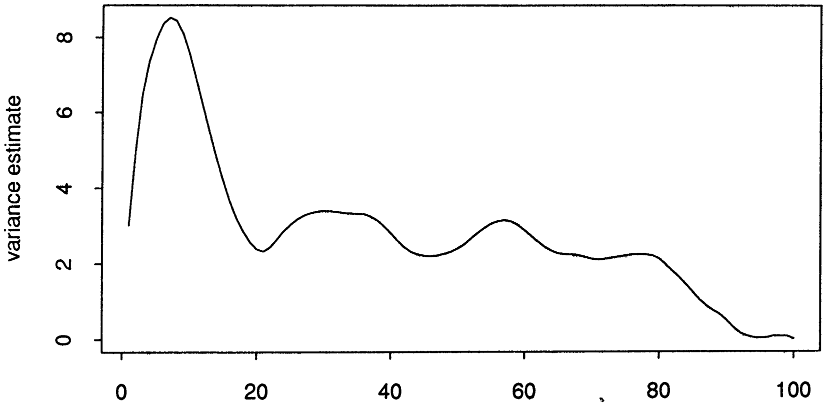

FIG. 1. $\hat{V}_{\text{JACK}}$ *for the sample mean case, as a function of the block size b.*

model that was used was very simple, and the situation itself (estimating the variance of the sample mean) is the simplest of the ones discussed in our paper. Even in slightly more complicated examples, for example, a moving average model of higher order or an autoregressive model, it is not uncommon to have variance estimates that are off by an order of magnitude from the true parameter, for a sample size of 100 (which was used in our example). It follows that a result like Lemma 2 on higher-order accuracy is of theoretical value only, unless a huge sample size is available. One cannot expect higher-order accuracy of the bootstrap distribution approximation if the scale parameter is not estimated accurately.

To return to the spectral density example, it is noteworthy that in the special case of nonoverlapping $B_{i,M,L}$ blocks, that is, $L = M$, Brillinger (1975) proved that the $T_{i,M,L}$'s are in fact asymptotically *independent* (i.e., 0-dependent), under the conditions that all cumulants of X_t exist and are summable. This is a refinement of our Lemma 1 in that particular case. So for $L = M$ we could theoretically use $b = 1$, that is, the classical jackknife and bootstrap, and have an asymptotically valid procedure. However, even here, it would be advisable in a practical application to take b larger than 1, since in a finite sample setup the $T_{i,M,L}$'s are *not* independent.

Acknowledgment. Thanks are due to the anonymous referees for many helpful suggestions that significantly ameliorated the presentation of our results.

REFERENCES

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BARTLETT, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *J. Roy. Statist. Soc. Suppl.* **8** 27–41.

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

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BLACKMAN, R. B. and TUKEY, J. W. (1959). *The Measurement of Power Spectra from the Point of View of Communication Engineering*. Dover, New York.

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

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

DANIELL, P. J. (1946). Discussion on Symposium on autocorrelation in time series. *J. Roy. Statist. Soc. Suppl.* **8** 88–90.

DICICCIO, T. J. and ROMANO, J. P. (1988). A review of bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **50** 338–370.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. SIAM, Philadelphia.

EFRON, B. and TIBSHIRANI, R. J. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statist. Sci.* **1** 54–77.

FRANKE, J. and HÄRDLE, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **20** 121–145.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

GÖTZE, F. and HIPP, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–240.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Dover, New York.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–953.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

HALL, P., MARTIN, M. A. and SCHUCANY, W. R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Statist. Comput. Simulation* **33** 161–172.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

HÄRDLE, W. and BOWMAN, A. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 100–110.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes*. Springer, New York.

JENKINS, G. M. (1963). Cross-spectral analysis and the estimation of linear open loop transfer functions. In *Time Series Analysis* (M. Rosenblatt, ed.) 267–276. Wiley, New York.

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

LAHIRI, S. N. (1990). Second order optimality of stationary bootstrap. Technical Report 90-1, Dept. Statistics, Iowa State Univ.

LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708.

LIU, R. Y. and SINGH, K. (1988). Moving blocks jackknife and bootstrap capture weak dependence. Unpublished manuscript, Dept. Statistics, Rutgers Univ.

PARZEN E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics* **3** 167–190.

POLITIS, D. N. (1990). A general resampling scheme for time series with weak dependence. Ph.D. dissertation, Dept. Statistics, Stanford Univ.

POLITIS, D. N. and ROMANO, J. P. (1992). A nonparametric resampling procedure for multivariate confidence regions in time series analysis. In *Computing Science and Statistics: Proceedings of the 22nd Symposium on the Interface* (C. Page and R. LePage, eds.) 98–103. Springer, New York.

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

ROSENBLATT, M. (1984). Asymptotic normality, strong mixing and spectral density estimates. *Ann. Probab.* **12** 1167–1180.

ROUSSAS, G. G. (1988). A moment inequality of S_{n, k_n} for triangular arrays of random variables under mixing conditions, with applications. In *Statistical Theory and Data Analysis II* (K. Matusita, ed.) 273. North-Holland, Amsterdam.

ROUSSAS, G. G. and IOANNIDES, D. (1987). Moment inequalities for mixing sequences of random variables. *Stochastic Anal. Appl.* **5** 61–120.

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

SWANEPOEL, J. W. H. and VAN WYK, J. W. J. (1986). The bootstrap applied to spectral density function estimation. *Biometrika* **73** 135–142.

THOMSON, D. and CHAVE, A. (1988). Jackknifed error estimates for spectra, coherences, and transfer functions. AT&T Bell Laboratories, Murray Hill, N.J.

TIKHOMIROV, A. N. (1980). On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.* **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. Electron.* **AU-15** 70.

YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. Gebiete* **52** 45–57.

ZHURBENKO, I. G. (1979). On spectral density statistics obtained by means of time shift. *Soviet Math. Dokl.* **20** 547–552.

ZHURBENKO, I. G. (1980). A limit theorem for spectral density statistics with time shift. *Ukrainian Math. J.* **32** 316–326.

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

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305