NONPARAMETRIC RESAMPLING FOR HOMOGENEOUS STRONG MIXING FIELDS

by

Dimitris N. Politis and Joseph P. Romano
Purdue University Stanford University

Technical Report #92-26

Department of Statistics
Purdue University

June 1992

Nonparametric Resampling for Homogeneous Strong Mixing Random Fields *

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

Künsch(1989) and Liu and Singh (1992) have recently introduced a block resampling method that is successful in deriving consistent bootstrap estimates of distribution and variance for the sample mean of a strong mixing sequence. Raïs and Moore (1990) and Raïs (1991) extended the results of Künsch and Liu and Singh in the case of the sample mean of a homogeneous strong mixing random field in two dimensions ($n = 2$).

In this report, the general case ($n \in \mathbf{Z}^+$) is considered, and a general resampling technique for strong mixing random fields is formulated, which is an extension of the 'blocks of blocks' resampling scheme for sequences in Politis and Romano (1992a, 1992c). The 'blocks of blocks' method can be used to construct asymptotically correct confidence intervals for parameters of the whole (infinite-dimensional) joint distribution of the random field, for example, the spectral density at a point. A variation of the 'blocks of blocks' resampling scheme that involves 'wrapping' the data around on a torus will also be studied, in view of its property to yield an unbiased bootstrap distribution.

AMS 1980 subject classifications. Primary 62M20; secondary 62G05.

Key words and phrases. Bootstrap, jackknife, nonparametric estimation, random fields, resampling methods, strong mixing, time series, weak dependence.

* Research partially supported by a Purdue Research Foundation Faculty Grant and NSF Grant DMS89-57217.

1. Introduction

Suppose $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{I}\}$ is a multivariate random field in n dimensions, with $n \in \mathbf{Z}^+$, i.e., a collection of random variables $X(\mathbf{t})$ taking values in $\mathbf{R}^d$, defined on a probability space $(\Omega, \mathcal{A}, P)$, and indexed by the variable $\mathbf{t} \in \mathbf{I} \subset \mathbf{R}^n$. The set $\mathbf{I}$ is assumed to be a regular discrete lattice in $\mathbf{R}^n$. If $n = 1$, $\mathbf{I}$ is just the set of integers $\mathbf{Z}$, and the random field reduces to being a random sequence. In two dimensions, the regular lattice can consist of triangular, rectangular, or hexagonal cells. For the case $n > 2$, the usual choice for $\mathbf{I}$ is the integer (rectangular) lattice $\mathbf{Z}^n$. The random field $\{X(\mathbf{t})\}$ will be assumed to be *homogeneous* (stationary, shift invariant), and weakly dependent (see Section 2 for the exact definitions).

Consider the statistical problem of estimating a certain parameter of the first marginal distribution of the random field $\{X(\mathbf{t})\}$, i.e., of the distribution of the random variable $X(\mathbf{0})$. To fix ideas, suppose that it is desired to obtain a confidence interval for $\mu = EX(\mathbf{0})$, on the basis of observing $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_N\}$, where $\mathbf{E}_N$ is the rectangle consisting of the points $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbf{Z}^n$ such that $1 \leq t_k \leq N_k$, where $k = 1, 2, \ldots, n$. Hence, the total number of observations is $|\mathbf{E}_N| = \prod_{i=1}^n N_i = N$. For this purpose, an approximation to the sampling distribution of the sample mean $\bar{X}_N = |\mathbf{E}_N|^{-1} \sum_{\mathbf{i} \in \mathbf{E}_N} X(\mathbf{i})$ is required. The Central Limit Theorem for random fields (cf. Bolthausen (1982), Rosenblatt (1985), Bradley (1992)) provides the convergence

$$(Var(\bar{X}_N))^{-1/2}(\bar{X}_N - \mu) \overset{\mathcal{L}}{\Longrightarrow} N(0,1) \tag{1}$$

as $N \to \infty$, under the assumptions $0 < EX(\mathbf{0})^2 < \infty$, and the strong mixing coefficient $\alpha_X(k)$ satisfies $\alpha_X(k) \to 0$ as $k \to \infty$, (the definition of $\alpha_X(k)$ is found in the next section). However, the variance $Var(\bar{X}_N)$ is unknown and must somehow be estimated for the Central Limit Theorem to be of any practical use. This is by no means a trivial problem, since it essentially requires an estimate of the spectral density of the random field at the origin.

Alternatively, a different, possibly non-normal, approximation of the distribution of the sample mean could be used for the purpose of constructing confidence intervals. For the case of a strong mixing random sequence ($n = 1$), Künsch (1989) and Liu and Singh (1992) have introduced a nonparametric version of the bootstrap and jackknife that yields confidence intervals for μ with asymptotically correct coverage. Their technique amounts to resampling or deleting one by one whole blocks of observations.

The block resampling method can be used to consistently estimate the distribution of the sample mean and related (differentiable) statistics, with the objective of setting confidence intervals for a parameter μ of the first marginal distribution of $\{X(\mathbf{t})\}$. As shown by Lahiri (1991), under appropriate conditions the block resampling method of Künsch and Liu and Singh yields a 'better-than-the-normal' approximation to the sampling distribution of the sample mean. This property is in exact analogy to the well-known (cf. Singh (1981)) optimality of the classical bootstrap for independent, identically distributed observations of Efron (1979, 1982).

The 'blocks of blocks' resampling technique (cf. Politis and Romano (1992a, 1992c)) was introduced as a generalization of the block resampling method of Künsch (1989) and Liu and Singh (1992), that permits construction of asymptotically valid confidence intervals for parameters of the whole (infinite-dimensional) distribution of $\{X(\mathbf{t})\}$. A prominent example of such a parameter is the spectral density function evaluated at a point, or at a grid of points.

Raïs and Moore (1990) and Raïs (1992) extended the results of Künsch (1989) and Liu and Singh (1992) to the case of the sample mean of a homogeneous random field in two dimensions ($n = 2$), observed on a finite part of a regular discrete lattice on the plane. All three regular lattices in two dimensions were examined by Raïs and Moore, and the block resampling bootstrap method was shown to be valid under reasonable moment and mixing conditions. Alternative resampling approaches for spatial data are found in Hall (1985, 1988) and Lele (1991).

It is the purpose of this report to simultaneously generalize the results of Raïs and Moore (1990) and Raïs (1992) in two directions: (a) to formulate a general resampling technique (the 'blocks of blocks' resampling scheme) for triangular arrays defined on a homogeneous random field, and (b) to allow for the possibility of having a random field in n dimensions, with n being *any* positive integer. As implied by its name, the 'blocks of blocks' resampling scheme for random fields is a generalization of the 'blocks of blocks' technique for sequences (cf. Politis and Romano (1992a, 1992c)).

The significance of extension (a) is that the 'blocks of blocks' method can be used to construct confidence intervals for parameters of the whole (infinite-dimensional) joint distribution of the random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$, such as the spectral density of the random field. The significance of extension (b) is apparent considering the following two broad classes of interesting examples.

- *Spatial random fields.* In this example the dimension n is usually 2 or 3, and $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_N \subset \mathbf{Z}^n\}$, are measurements of some physical quantity (pressure, temperature, etc.) at different points of a two-dimensional surface, or three-dimensional space.

- *Time series in* $(n-1)$ *dimensions.* Here t_1 (the first coordinate of $\mathbf{t} = (t_1, t_2, \ldots, t_n)$) is reserved to represent 'time'. As an example with $n = 3$, let t_1 denote the time parameter and (t_2, t_3) denote the spatial coordinates of a point on a surface. Then $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_N\}$ is a series of 'snap-shots' taken over time, where each snap-shot contains the measurements of a physical quantity at many different points of the surface and some point in time.

The importance of having such a resampling methodology for random fields lies in the fact that a whole host of different estimation problems can be approached and solved in a general framework. Results that are either unavailable or possibly just too cumbersome using classical asymptotic methods would now be immediate corollaries of a general theorem. In addition, from a computational point of view, a single general purpose resampling algorithm can be employed to produce point and interval estimates for practically *any* parameter of a random field model.

A variation of the 'blocks of blocks' resampling scheme that involves 'wrapping' the data around on a torus will also be studied, in view of its property to yield an unbiased bootstrap distribution. This 'circular blocks of blocks' methodology is new even for the one-dimensional ($n = 1$) case, since it has been previously defined and studied only for the sample mean of a stationary sequence (cf. Politis and Romano (1992b)). As will be discussed in Section 4, the 'circular' bootstrap is preferable to the 'non-circular' one, because, in addition to its unbiasedness property, it also yields a more accurate variance estimate.

2. Some definitions and assumptions

The random field $\{X(\mathbf{t})\}$ is assumed to be *homogeneous*, meaning that for any set $\mathbf{E} \subset \mathbf{I}$, and for any point $\mathbf{i} \in \mathbf{I}$, the joint distribution of the random variables $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}\}$ is identical to the joint distribution of $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}+\mathbf{i}\}$, where the set $\mathbf{E}+\mathbf{i} = \{\mathbf{t} \in \mathbf{I} : \mathbf{t} = \mathbf{t}' + \mathbf{i}, \mathbf{t}' \in \mathbf{E}\}$ is the set $\mathbf{E}$ 'translated' by $\mathbf{i}$ [1].

For simplicity and concreteness, attention will focus on the integer (rectangular) lattice $\mathbf{I} = \mathbf{Z}^n$ in the general n-dimensional case. For two points $\mathbf{t} = (t_1, \ldots, t_n)$ and $\mathbf{u} = (u_1, \ldots, u_n)$ in $\mathbf{Z}^n$, define the sup-distance in $\mathbf{Z}^n$ by

$$d(\mathbf{t}, \mathbf{u}) = \sup_j |t_j - u_j| \tag{2}$$

and for two sets $\mathbf{E}_1, \mathbf{E}_2$ in $\mathbf{Z}^n$, define

$$d(\mathbf{E}_1, \mathbf{E}_2) = \inf\{d(\mathbf{t}, \mathbf{u}) : \mathbf{t} \in \mathbf{E}_1, \mathbf{u} \in \mathbf{E}_2\} \tag{3}$$

In addition, the random field $\{X(\mathbf{t})\}$ is assumed to satisfy a weak dependence condition. Recall Rosenblatt's (1985) strong mixing coefficient, as applied to the setting of homogeneous random fields, which is defined by

$$\alpha_X(k) \equiv \sup |P(A \cap B) - P(A)P(B)| \tag{4}$$

where $A \in \mathcal{F}(\mathbf{E}_1), B \in \mathcal{F}(\mathbf{E}_2)$, and the supremum is over all sets $\mathbf{E}_1, \mathbf{E}_2$ in $\mathbf{I} = \mathbf{Z}^n$, such that the distance $d(\mathbf{E}_1, \mathbf{E}_2) = k$; note that $\mathcal{F}(\mathbf{E})$ is just the σ-algebra generated by $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}\}$.

The homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is said to be *α-mixing* if $\alpha_X(k) \to 0$ as $k \to \infty$. Examples of α-mixing homogeneous random fields include Gaussian fields with continuous and positive spectral density function (cf. Rosenblatt (1985)), and finitely dependent (moving average type) random fields for which $\alpha_X(k) = 0$, for $k >$ some r (cf. Moore (1988), Tjøstheim (1978)).

Suppose $\mu \in \mathbf{R}$ is a parameter of the whole (infinite-dimensional) joint distribution of the multivariate homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$, where $X(\mathbf{t})$ takes values in $\mathbf{R}^d$.

[1] Note that the fact that for any $\mathbf{E} \subset \mathbf{I}$, and $\mathbf{i} \in \mathbf{I}$, the set $\mathbf{E}+\mathbf{i}$ is a subset of $\mathbf{I}$, can be thought of as the defining property of the assumed regularity of the lattice $\mathbf{I}$.

The objective is to obtain confidence intervals for μ based on the observations $\{X(\mathbf{t})$, for $\mathbf{t} \in \mathbf{E}_N\}$, where $\mathbf{E}_N$ is the rectangle consisting of the points $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbf{Z}^n$ such that $1 \leq t_k \leq N_k$, where $k = 1, 2, \ldots, n$. Hence, the total number of observations is $N = \prod_{i=1}^n N_i$. For simplicity, in all that follows, only rectangular observation 'sites' $\mathbf{E}_N$ will be considered. Extensions to non-rectangular observation 'sites' (that possess some regularity) are possible, and will be apparent from the treatment of the rectangular case.

Furthermore it will be assumed that, as the sample size N increases, the corresponding set $\mathbf{E}_N$ 'expands', i.e. that if $N < N^*$, then $\mathbf{E}_N \subset \mathbf{E}_{N^*}$. Actually, it will be required that the set $\mathbf{E}_N$ 'expands' more or less uniformly in all directions by defining:

$N \Rightarrow \infty$ is equivalent to $N \to \infty$ in such a way that $\mathbf{E}_N \subset Ball(cN^{1/n})$,

where $Ball(r)$ is the ball of radius r in $\mathbf{R}^n$, and c is some positive constant.

The set-up to apply our estimation procedures is as follows. Let the block $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L}) = \{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$, where $\mathbf{i} = (i_1, i_2, \ldots, i_n)$ and $\mathbf{E}_{\mathbf{i},\mathbf{M},\mathbf{L}}$ is the smaller (and displaced) rectangle consisting of the points $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbf{Z}^n$ such that $(i_k - 1)L_k + 1 \leq t_k \leq (i_k - 1)L_k + M_k$, for $k = 1, 2, \ldots, n$, and where L_k and M_k are integers depending in general on the corresponding N_k. Denote $\mathbf{M} = (M_1, M_2, \ldots, M_n)$, $\mathbf{L} = (L_1, L_2, \ldots, L_n)$, and $M = \prod_{i=1}^n M_i$, $L = \prod_{i=1}^n L_i$. Note that, given the observations, $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L})$ is defined only for $\mathbf{i}$ such that $i_k = 1, \ldots, Q_k$, where $Q_k = [\frac{N_k - M_k}{L_k}] + 1$, and $[\cdot]$ is the integer part function. The total number of the $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L})$ blocks available from the data is therefore $Q = \prod_{i=1}^n Q_i$.

As in the time series case (cf. Politis and Romano (1993)), the general linear statistic can be formulated as

$$\bar{T}_N = \frac{1}{Q} \sum_{\mathbf{i} \in \mathbf{E}_Q} T_{\mathbf{i},\mathbf{M},\mathbf{L}} \tag{5}$$

where $T_{\mathbf{i},\mathbf{M},\mathbf{L}} = \phi_{\mathbf{M}}(\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L}))$, and $\phi_{\mathbf{M}} : \mathbf{R}^{dM} \to \mathbf{R}$ is some appropriately chosen function of the block $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L})$ that makes $\bar{T}_N$ an almost unbiased estimator of μ, (cf. assumption A_2 in what follows). This formulation permits the treatment of all standard estimation problems in the random field setting, e.g., estimating the mean, the autocovariances, the spectral density function, etc., as special cases of parameters estimable by a general linear statistic. The three abovementioned examples will be revisited in Section 4, together with specific comments on the implementation of the resampling methodology.

Note that, for each $\mathbf{E}_N$ considered, $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_Q\}$ is a homogeneous, weakly dependent random field in its own right, observed on the finite lattice $\mathbf{E}_Q$, where $\mathbf{E}_Q$ is the rectangle consisting of the points $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbf{Z}^n$ such that $1 \leq t_k \leq Q_k$, for $k = 1, 2, \ldots, n$. As a matter of fact, the $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$'s represent actually a triangular (more accurately: 'pyramidal') array of random fields, since for each N and $\mathbf{E}_N$ considered, the values of $\mathbf{M}, \mathbf{L}$ are generally different, resulting in a different random field. To make the notation easier, in what follows the dependence of the values of the $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$ field on N and $\mathbf{E}_N$ will not be explicitly denoted. While the homogeneity of $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$ is obvious, its weak dependence properties are the subject of the following lemma.

Lemma 1 *For each $\mathbf{E}_N$ fixed, if the $\{X(\mathbf{t})\}$ field is α-mixing with mixing coefficient $\alpha_X(k)$, then the following is true:*
(a) The $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$ field is α-mixing with mixing coefficient $\alpha_T(k) \leq \alpha_X(kL^\star - M^\star)$, for $k \geq [\frac{M^\star}{L^\star}] + 1$, where $M^\star = \max_i M_i$, $L^\star = \min_i L_i$.
(b) If $aM^\star \leq L^\star$, for some constant $a > 0$, then the $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$ field is also α-mixing with mixing coefficient $\alpha_T(k) \leq \alpha_X(kL^\star - M^\star)$, for $k \geq [\frac{1}{a}] + 1$.
(c) If, in addition to $aM^\star \leq L^\star$, $M^\star \to \infty$ as $N \Rightarrow \infty$, then for any fixed $k \geq [\frac{1}{a}] + 1$, $\lim_{N \Rightarrow \infty} \alpha_T(k) = 0$.

In the next sections we will make frequent use of the following assumptions, where all limits and order notations are taken as $N \Rightarrow \infty$, unless otherwise stated.

Assumptions:

(A_0) The random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is homogeneous and α-mixing.

(A_1) $E|T_{\mathbf{i},\mathbf{M},\mathbf{L}}|^{2p+\delta} < C$, for any $\mathbf{M}$, where p is an integer with $p > 2$, and $0 < \delta \leq 2, C > 0$ are some constants.

(A_2) $ET_{\mathbf{i},\mathbf{M},\mathbf{L}} = \mu + o(Q^{-1/2})$, where μ is a parameter of the infinite-dimensional joint distribution of the $\{X(\mathbf{t})\}$ random field, i.e., a parameter associated with the probability measure P.

(A_3) $\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \stackrel{\mathcal{L}}{\Longrightarrow} N(0, \sigma^2_\infty)$, and $\lim_{N \Rightarrow \infty} V_{Q/N} = \sigma^2_\infty > 0$, where $V_{Q/N} \equiv Var(\sqrt{Q}\bar{T}_N) = Var(\frac{1}{\sqrt{Q}} \sum_{\mathbf{i} \in \mathbf{E}_Q} T_{\mathbf{i},\mathbf{M},\mathbf{L}})$.

Regarding assumptions A_1 and A_2, note that $E|T_{\mathbf{i},\mathbf{M},\mathbf{L}}|^{2p+\delta}$ is the same for all $\mathbf{i}$, because of the homogeneity of the $\{X(\mathbf{t})\}$ and $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}\}$ fields; therefore, $E|T_{\mathbf{i},\mathbf{M},\mathbf{L}}|^{2p+\delta} = E|T_{\mathbf{1},\mathbf{M},\mathbf{L}}|^{2p+\delta}$, where the latter expectations do not even depend on $\mathbf{L}$. Similarly, $ET_{\mathbf{i},\mathbf{M},\mathbf{L}} = ET_{\mathbf{1},\mathbf{M},\mathbf{L}}$, where of course $\mathbf{1} = (1, 1, \ldots, 1)$.

The Central Limit Theorem of assumption A_3 will actually hold under common regularity conditions. For example (cf. Tikhomirov (1983)), sufficient conditions for A_3 to hold are the moment condition A_1 which is already assumed, an exponential mixing rate, i.e., $\alpha_X(k) \leq Ke^{-\beta k}$, for some positive K and β, and a variance condition of the type $\lim_{N \Rightarrow \infty} V_{Q/N}$ exists and equals $\sigma^2_\infty > 0$.

Assumptions A_2 and A_3 taken together can be used to obtain approximate confidence intervals for μ, given a finite sample of size N. Note however that to actually set the confidence intervals, the variance $V_{Q/N}$ or the asymptotic variance σ^2_∞ must be estimated. Estimating the variance and, in fact, the whole sampling distribution of $\bar{T}_N$, can be accomplished by one of the resampling methods that will be described in detail in the next section.

In parallel to the time series case, the 'blocks of blocks' resampling technique for homogeneous random fields amounts to resampling or deleting one by one whole blocks (rectangles) of the $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$'s. In Section 3 it will be shown that, under appropriate conditions, the resampling estimates of variance and sampling distribution are consistent. As a consequence, the corresponding confidence intervals for μ have asymptotically correct coverage probability.

It should be pointed out that assumption A_2 ensures that the asymptotic order of the bias of $\bar{T}_N$ is smaller than the asymptotic order of its standard deviation. This implies that the asymptotic confidence intervals for $E\bar{T}_N$ that are obtained from the Central Limit Theorem of assumption A_3 can be regarded as approximate confidence intervals for μ as well. However, for the purposes of variance estimation alone, assumption A_2', which is weaker than A_2, is sufficient, and allows for estimators $\bar{T}_N$ optimal from the point of view of Mean Squared Error.

(A_2') $ET_{\mathbf{i},\mathbf{M},\mathbf{L}} = \mu + O(Q^{-1/2})$.

The following moment inequality will be useful for our proofs. It was first proven by Yokoyama (1980) in the case of one dimension ($n = 1$), and by Raïs (1992) in the case of two dimensions ($n = 2$).

Lemma 2 *Let the univariate homogeneous random field $X(\mathbf{t})$ be observed at points $\mathbf{t} \in \mathbf{E}_N$, and assume $E|X(\mathbf{t})|^{2p+\delta} < \infty$, for some $p \in \mathbf{N}$, and $\delta > 0$, and that $\sum_{k=1}^{\infty} k^{np-1}\{\alpha_X(k)\}^{\delta/(2p+\delta)}$ is finite. If $N \Rightarrow \infty$, then a constant $0 \leq c < \infty$ exists that depends only on p and $\alpha_X(\cdot)$, and not on $\mathbf{E}_N$, such that*

$$E|\sum_{\mathbf{i}\in\mathbf{E}_N} X(\mathbf{i})|^{2p} \leq cN^p(E|X(\mathbf{t})|^{2p+\delta})^{2p/(2p+\delta)} \tag{6}$$

3. The 'blocks of blocks' resampling scheme for random fields

3.1 The jackknife. Focus attention on a particular sample size N, (and hence particular values of $\mathbf{M}, \mathbf{L}$ as well), and a corresponding set of observation 'sites' $\mathbf{E}_N$, and define the block $\mathcal{B}_{\mathbf{j}} = \{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}\}$, where $\mathbf{j} = (j_1, j_2, \ldots, j_n)$ and $\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}$ is the smaller rectangle consisting of the points $\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \mathbf{Z}^n$ such that $(j_k - 1)h_k + 1 \leq i_k \leq (j_k - 1)h_k + b_k$, for $k = 1, 2, \ldots, n$, and where $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{h} = (h_1, \ldots, h_n)$ are vectors in $\mathbf{Z}^n$ that depend in general on N and $\mathbf{E}_N$. As before, denote $b = \prod_{i=1}^n b_i$, and $h = \prod_{i=1}^n h_i$. Observe that, with $\mathbf{E}_N$ and N fixed, $\mathcal{B}_{\mathbf{j}}$ is defined only for $\mathbf{j}$ such that $1 \leq j_k \leq q_k$, where $q_k = [\frac{Q_k - b_k}{h_k}] + 1$, and thus the total number of the $\mathcal{B}_{\mathbf{j}}$ blocks available from the data is $q = \prod_{i=1}^n q_i$.

Analogously to the time series case (cf. Politis and Romano (1992a, 1992c)), let $\bar{T}_{N,-\mathbf{j}}$ be the average of the remaining $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$'s, after deleting the block $\mathcal{B}_{\mathbf{j}}$, i.e.,

$$\bar{T}_{N,-\mathbf{j}} = \frac{1}{Q - b} \sum T_{\mathbf{i},\mathbf{M},\mathbf{L}} \tag{7}$$

where the sum is over all $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ that are not in the block $\mathcal{B}_{\mathbf{j}}$. Then, define the pseudovalues $J_{\mathbf{j}} = \frac{1}{b}(Q\bar{T}_N - (Q - b)\bar{T}_{N,-\mathbf{j}})$, for $\mathbf{j} \in \mathbf{E}_q$, where $\mathbf{E}_q$ is the rectangle consisting of the points $\mathbf{j} = (j_1, j_2, \ldots, j_n)$ such that $1 \leq j_k \leq q_k$, for all $k = 1, 2, \ldots, n$.

The *'blocks of blocks' jackknife* estimate of the variance of $\sqrt{Q}\bar{T}_N$ is defined by

$$\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = \frac{b}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} (J_{\mathbf{j}} - \bar{T}_N)^2 \tag{8}$$

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

Theorem 1 *Under assumptions A_0, A_1, A_2', A_3, and if for $N \Rightarrow \infty$ we have*
(i) $Q \Rightarrow \infty$, $c_* < \frac{M_i}{M_j} < c^*$, and $\frac{L_i}{M_i} \to a_i$, for some constants $0 < c_* < c^*$, $a_i \in (0, 1]$, and for any $i, j = 1, \ldots, n$;
(ii) $b \Rightarrow \infty$ and $\frac{h_i}{b_i} \to d_i > 0$, $i = 1, \ldots, n$;
(iii) $b = o(Q)$;
(iv) $\sum_{k=1}^{\infty} k^{np-1}\{\alpha_X(k)\}^{\frac{\delta}{2p+\delta}} < \infty$;
then:

$$E\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = V_{b/N} + O(b/Q) \tag{9}$$

and

$$Var(\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)) = O(b/Q) \tag{10}$$

where $V_{b/N} \equiv Var(\frac{1}{\sqrt{b}}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{1},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}})$.

Note that in the assumptions of the theorem, $\mathbf{M}$ is allowed to either remain constant, or to satisfy $\mathbf{M} \Rightarrow \infty$ as $\mathbf{N} \Rightarrow \infty$, *as long as assumption* A_2' *is satisfied.* Choosing $\mathbf{M}$ (as well as $\mathbf{L}, \mathbf{b}, \mathbf{h}$, and l) appropriately will be discussed in Section 4. It will now be shown that it is not necessary that $\mathbf{h} \Rightarrow \infty$ for Theorem 1 to be true.

Theorem 2 *Under assumptions* A_0, A_1, A_2', A_3, *and if for* $N \Rightarrow \infty$ *we have*
(i) $Q \Rightarrow \infty$, $c_* < \frac{M_i}{M_j} < c^*$, *and* $\frac{L_i}{M_i} \to a_i$, *for some constants* $0 < c_* < c^*$, $a_i \in (0,1]$, *and for any* $i, j = 1, \ldots, n$;
(ii′) $b \Rightarrow \infty$ *and* $\mathbf{h} = (1, 1, \ldots, 1)$;
(iii) $b = o(Q)$;
(iv) $\sum_{k=1}^{\infty} k^{np-1}\{\alpha_X(k)\}^{\frac{\delta}{2p+\delta}} < \infty$;
then equations (9) and (10) remain true.

An alternative expression for $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$ is

$$\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = \frac{b}{q}\sum_{\mathbf{j}\in\mathbf{E}_q}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - \bar{T}_N)^2. \tag{11}$$

Thus, the jackknife variance estimate is identical to a 'sample variance' estimate; see Politis and Romano (1993) for a discussion of the notion of 'sample variance' in dependent samples.

As it turns out, instead of directly estimating $V_{Q/N} = Var(\sqrt{Q}\bar{T}_N)$ from the observations $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_Q\}$, the quantity $V_{b/N}$ (for $b << Q$) is estimated by looking at the variability of $\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ as $\mathbf{j}$ varies. Under our assumptions $Q \Rightarrow \infty$ and $b \Rightarrow \infty$ and assumption A_3, it is immediate that both $V_{Q/N} \to \sigma_\infty^2$, and $V_{b/N} \to \sigma_\infty^2$, and thus, $V_{Q/N} - V_{b/N} \to 0$. Hence the following corollary of Theorems 1 and 2 is true, which (in part) answers to the affirmative a conjecture of Cressie (1991, p. 492) regarding the suitability of a 'sample variance' estimator.

Corollary 1 *Under the assumptions of Theorem 1 or Theorem 2*

$$E\left(\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) - \sigma_\infty^2\right)^2 \to 0 \tag{12}$$

As a matter of fact, the closeness of the approximation of $V_{Q/N}$ by $V_{b/N}$ is crucial, since it determines the bias of $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$ as an estimate of $V_{Q/N}$ or σ^2_∞. The following theorem addresses this issue.

Theorem 3 *Under the assumptions of Theorem 1 or Theorem 2, and the additional condition* $\alpha_X(k) = O(k^{-\lambda})$, *with* $\lambda > \frac{np(2p+\delta)}{\delta}$,

$$V_{Q/N} - V_{b/N} = O(b^{-1/n}) \tag{13}$$

and

$$V_{Q/N} - \sigma^2_\infty = O(Q^{-1/n}) \tag{14}$$

Since $b = o(Q)$, it is apparent that $V_{Q/N}$ is 'closer' to σ^2_∞, than $V_{b/N}$ is to either $V_{Q/N}$ or σ^2_∞. Hence we can define the bias of $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$ either by $Bias(\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)) \equiv E\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) - V_{Q/N}$, or by $Bias(\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)) \equiv E\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) - \sigma^2_\infty$, and the following interesting corollary is immediate.

Corollary 2 *Under the assumptions of Theorem 3*

$$Bias(\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)) = O(b^{-1/n}) + O(b/Q) \tag{15}$$

which, combined with equation (10), implies that the choice $b \sim a_b Q^{\frac{n}{n+2}}$, *for some constant* $a_b > 0$, *minimizes the asymptotic order of the Mean Squared Error of* $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$ *as an estimator of* $V_{Q/N}$ *or* σ^2_∞.

The estimates offered by equations (10) and (15) cannot generally be improved. In particular, in the case where $\bar{T}_N$ is the sample mean of a stationary sequence (cf. Künsch (1989)), the asymptotic rates indicated by equations (10) and (15) are in fact attained. The same is true in the more general setting of the sample mean of a homogeneous random field, i.e., if $T_{\mathbf{i},\mathbf{M},\mathbf{L}} \equiv X(\mathbf{i})$, in which case $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$ is just a nonparametric estimate of (a constant multiple of) the spectral density of the random field evaluated at the origin.

3.2 The bootstrap. The 'blocks of blocks' bootstrap resampling is defined as follows. Sampling with replacement from the set $\{\mathcal{B}_{\mathbf{j}}, \mathbf{j} \in \mathbf{E}_q\}$ defines a (conditional on the original observations $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_N\}$) probability measure denoted by P^*. Let $Y_1, \ldots, Y_k$ be i.i.d. samples from P^*. Obviously each Y_i is a block (rectangle), with dimensions equal to the dimensions of $\mathcal{B}_1$. Let $\bar{Y}_i$ be the average of the $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$'s that are found in block Y_i, i.e.,

$$\bar{Y}_i = \frac{1}{b} \sum T_{\mathbf{i},\mathbf{M},\mathbf{L}} \tag{16}$$

where the sum is over all $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ that are found in block Y_i. Now define $\bar{T}_l^*$ to be the average of the $\bar{Y}_i$'s, i.e.,

$$\bar{T}_l^* = \frac{1}{k} \sum_{i=1}^{k} \bar{Y}_i \tag{17}$$

and note that $\bar{T}_l^*$ is actually the average of the $l = kb$ observations $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ that are found in the resampled blocks $Y_1, \ldots, Y_k$.

The following assumption on the order of magnitude of l will be needed for consistency of the 'blocks of blocks' bootstrap.

(A_4) $b = o(l)$.

Assumption A_4 together with the condition $b \Rightarrow \infty$ implies that $k \to \infty$ as $N \Rightarrow \infty$. It is easy to see that (because of condition (iii) or (iii') below) assumption A_4 is trivially satisfied if l is taken to be of the same asymptotic order as Q.

The *'blocks of blocks' bootstrap* approximation to the sampling distribution of $\bar{T}_N$ is provided by the following theorem.

Theorem 4 *Under the assumptions of Theorem 1 or 2, the additional assumption A_4, and the additional condition*

(iii') $b = o(Q^{2/3})$,

it is true that

$$E\left(Var^*(\sqrt{l}\bar{T}_l^*) - \sigma_\infty^2\right)^2 \to 0 \tag{18}$$

where $Var^*(\sqrt{l}\bar{T}_l^*)$ is the variance of $\sqrt{l}\bar{T}_l^*$ under the resampling probability P^*, and

$$\sup_x |P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \le x\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \le x\}| \xrightarrow{P} 0 \tag{19}$$

If assumption A_2 is adopted (instead of the weaker A_2'), we additionally have

$$\sup_x |P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \le x\} - P\{\sqrt{Q}(\bar{T}_N - \mu) \le x\}| \xrightarrow{P} 0 \tag{20}$$

as well as

$$\sup_x |P^*\{\sqrt{l}\frac{\bar{T}_l^* - E^*\bar{T}_l^*}{\sqrt{Var^*(\sqrt{l}\bar{T}_l^*)}} \le x\} - P\{\sqrt{Q}\frac{\bar{T}_N - \mu}{\sigma_\infty} \le x\}| \xrightarrow{P} 0 \tag{21}$$

where $\xrightarrow{P}$ denotes convergence in probability (as $N \Rightarrow \infty$).

An important observation is that $Var^*(\sqrt{l}\bar{T}_l^*) = \frac{b}{q}\sum_{\mathbf{j}\in\mathbf{E}_q}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - E^*\bar{T}_l^*)^2$, and $E^*\bar{T}_l^* = \frac{1}{q}\sum_{\mathbf{j}\in\mathbf{E}_q}\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}}$, both of which can be computed without resampling. As seen from the proof of the theorem, (cf. equation (34)), the bootstrap variance estimate $Var^*(\sqrt{l}\bar{T}_l^*)$ is asymptotically equivalent to the jackknife estimate $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$.

Equations (19), (20), and (21), would still be true with $\bar{T}_N$ substituted in place of $E^*\bar{T}_l^*$, provided a stricter bound is put on the block size b, e.g., $b = o(\sqrt{Q})$ and $l \sim Q$ (cf. Künsch(1989) and Liu and Singh (1992)), or $b = o(\sqrt{Q})$ and $l = o(Q^2/b^2)$ (cf. Politis and Romano (1992c)), in the one-dimensional case. However, it is now well known (cf. Lahiri (1991), Politis and Romano (1992b)) that this is not desirable since it introduces a bias in the bootstrap distribution, resulting in poorer approximations than the ones provided by (19), (20), and (21) as stated.

3.3 The 'circular' bootstrap. Equations (19), (20), and (21), all involve the re-centered bootstrap distributions, where subtraction of $E^*\bar{T}_l^*$ forces the bias of the bootstrap distribution to be exactly zero. A simple and 'automatic' way to have an unbiased 'blocks of blocks' bootstrap distribution is to 'wrap' the $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$'s around on a compact n-dimensional torus, that is, to define (for $\mathbf{i} \notin \mathbf{E}_Q$) $T_{\mathbf{i},\mathbf{M},\mathbf{L}} \equiv T_{\mathbf{i}^*,\mathbf{M},\mathbf{L}}$, where $\mathbf{i}^* = (i_1^*, \ldots, i_n^*)$, and $i_s^* = i_s(\text{mod}Q_s)$, for $s = 1, \ldots, n$.

The 'circular' block resampling bootstrap amounts to resampling whole rectangular 'patches' of the torus, and goes as follows. Define the blocks $\mathcal{B}_\mathbf{j}$ as previously, but note that now, *for any* $\mathbf{b}$, *there are Q such $\mathcal{B}_\mathbf{j}$, $\mathbf{j} \in \mathbf{E}_Q$,* (provided that $\mathbf{h} = (1, 1, \ldots, 1)$ as will be assumed in the following theorem).

Sampling with replacement from the set $\{\mathcal{B}_\mathbf{j}, \mathbf{j} \in \mathbf{E}_Q\}$ defines a (conditional on the original observations $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_N\}$) probability measure denoted by P^*. Let $Y_1, \ldots, Y_k$ be i.i.d.

samples from $P^\star$. Each Y_i is a block (rectangle), with dimensions equal to the dimensions of $\mathcal{B}_1$. As before, let

$$\bar{Y}_i = \frac{1}{b} \sum T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \tag{22}$$

where the sum is over all $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ that are found in block Y_i, and

$$\bar{T}_l^\star = \frac{1}{k} \sum_{i=1}^{k} \bar{Y}_i \tag{23}$$

i.e., $\bar{T}_l^\star$ is the average of the $l = kb$ observations $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ that are found in the resampled blocks $Y_1, \ldots, Y_k$.

The *circular 'blocks of blocks' bootstrap* approximation to the sampling distribution of $\bar{T}_N$ is provided by the following theorem.

Theorem 5 *Under the assumptions of Theorem 2 and the additional assumption A_4, it follows that*

$$E^\star \bar{T}_l^\star = \bar{T}_N \tag{24}$$

$$E\left(Var^\star(\sqrt{l}\bar{T}_l^\star) - \sigma_\infty^2\right)^2 \to 0 \tag{25}$$

where $E^\star \bar{T}_l^\star, Var^\star(\bar{T}_l^\star)$ are respectively the mean and variance of $\bar{T}_l^\star$ under the resampling probability $P^\star$, and

$$\sup_x |P^\star\{\sqrt{l}(\bar{T}_l^\star - \bar{T}_N) \le x\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \le x\}| \xrightarrow{P} 0 \tag{26}$$

If assumption A_2 is adopted (instead of the weaker A_2'), we additionally have

$$\sup_x |P^\star\{\sqrt{l}(\bar{T}_l^\star - \bar{T}_N) \le x\} - P\{\sqrt{Q}(\bar{T}_N - \mu) \le x\}| \xrightarrow{P} 0 \tag{27}$$

as well as

$$\sup_x |P^\star\{\sqrt{l}\frac{\bar{T}_l^\star - \bar{T}_N}{\sqrt{Var^\star(\sqrt{l}\bar{T}_l^\star)}} \le x\} - P\{\sqrt{Q}\frac{\bar{T}_N - \mu}{\sigma_\infty} \le x\}| \xrightarrow{P} 0 \tag{28}$$

Again observe that $Var^\star(\sqrt{l}\bar{T}_l^\star) = \frac{b}{Q}\sum_{\mathbf{j}\in\mathbf{E}_Q}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - \bar{T}_N)^2$, which can be computed without resampling. Also $E^\star \bar{T}_l^\star = \bar{T}_N$, implying that the circular 'blocks of blocks' bootstrap distribution is automatically centered around $\bar{T}_N$. Another most desirable feature of the circular bootstrap is that (like the jackknife) it is valid under the condition (*iii*), which is weaker (and more natural) than condition (*iii'*). This becomes apparent by comparing equation (36) to equation (34) in Section 5.

Indeed, the inherent improper centering of the noncircular bootstrap seems to bias the corresponding variance estimate as well, while the circular bootstrap variance estimate does not suffer from this defect. Although both the circular and non-circular 'blocks of blocks' bootstrap methods are valid asymptotically, and in particular they were both shown to be better than the normal approximation in the special case of the sample mean of a stationary sequence (cf. Lahiri (1991), Politis and Romano (1992b)), the existence of such a bias in the (non-circular) variance estimate suggests that the circular bootstrap might be more accurate in finite samples.

To intuitively see this, concider the usual sample variance of an i.i.d. (independent, identically distributed) sample, which is essentially the average squared deviation *from the sample mean* if the true mean is unknown. Since the true variance is the expected squared deviation from the true mean, centering the data around the sample mean, before squaring and averaging, results in a good estimate of the true variance, one of the reasons being that the sample mean is a good estimator of the true mean. It is obvious that centering the data around some other number than the sample mean would yield a variance estimator that is generally larger and not as accurate as the sample variance.

3.4 Multivariate extension. Using the δ-method, it is immediate that the jackknife, as well as the two bootstrap methods that were previously defined, remain asymptotically valid for smooth functions of the general linear statistic, i.e., statistics of the form $g(\bar{T}_N)$ as long as the function g has a non-zero derivative at μ.

Similarly, analogs of Theorems 1 – 5 hold true even if the function $\phi_{\mathbf{M}}$ is multivariate, i.e., $\phi_{\mathbf{M}} : \mathbf{R}^{dM} \to \mathbf{R}^{D}$, in which case both μ and the general linear statistic $\bar{T}_N$ are D-dimensional. In this case the bootstrap methodology is especially useful in that it can immediately yield approximate confidence regions for μ, that is, simultaneous confidence intervals for the coordinates $(\mu^{(1)}, \ldots, \mu^{(D)})$ of μ, (cf. Politis and Romano (1992a)).

For the multivariate limit theorems to hold, the assumptions A_1 and A_3 should be modified to accomodate the fact that $\bar{T}_N$ is D-dimensional, with coordinates $(\bar{T}_N^{(1)}, \ldots, \bar{T}_N^{(D)})$. The new assumptions should read:

($\mathbf{A}_1$) $E|T_{\mathbf{i},\mathbf{M},\mathbf{L}}^{(n_1)}|^{2p+\delta} < C$, for $n_1 = 1, \ldots, D$, and for any $\mathbf{M}$, where p is an integer with $p > 2$, and $0 < \delta \leq 2, C > 0$ are some constants.

($\mathbf{A}_3$) $\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \stackrel{\mathcal{L}}{\Longrightarrow} \mathbf{N}(0, \Sigma_\infty)$, the multivariate normal distribution with a positive definite covariance matrix $\Sigma_\infty = (\sigma_{n_1,n_2,\infty})$, where $\lim_{N \Rightarrow \infty} Cov(\sqrt{Q}\bar{T}_N^{(n_1)}, \sqrt{Q}\bar{T}_N^{(n_2)}) = \sigma_{n_1,n_2,\infty}$, and $\sigma_{n_1,n_1,\infty} > 0$, for all $1 \leq n_1 \leq n_2 \leq D$.

The 'blocks of blocks' jackknife, bootstrap, and circular bootstrap are defined the same way as in the univariate case. To elaborate, the 'blocks of blocks' jackknife estimate of $\sigma_{n_1,n_2,\infty}$ is

$$\hat{C}_{JACK}(\sqrt{Q}\bar{T}_N^{(n_1)}, \sqrt{Q}\bar{T}_N^{(n_2)}) = \frac{b}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} (J_{\mathbf{j}}^{(n_1)} - \bar{T}_N^{(n_1)})(J_{\mathbf{j}}^{(n_2)} - \bar{T}_N^{(n_2)}) \tag{29}$$

where $J_{\mathbf{j}}^{(n_1)}, \bar{T}_N^{(n_1)}$, etc., are the n_1th coordinates of $J_{\mathbf{j}}, \bar{T}_N$, and so forth. The 'blocks of blocks' bootstrap and circular bootstrap estimates of the (multivariate) sampling distribution of $\sqrt{Q}(\bar{T}_N - \mu)$ are $\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*)$ and $\sqrt{l}(\bar{T}_l^* - \bar{T}_N)$ respectively.

Theorem 6 *Under the assumptions of the respective univariate theorem, the 'blocks of blocks' jackknife, bootstrap, and circular bootstrap methods are asymptotically valid in the multivariate setting as well, i.e., they provide consistent estimates of the asymptotic covariance matrix and multivariate sampling distribution of the general linear statistic* $\bar{T}_N$.

4. Some remarks and examples

In this section it will be discussed in detail how the three archetypal problems of the nonparametric analysis of homogeneous random fields, namely estimating the mean, the autocovariance, and the spectral density function, can be approached by calculating and then resampling the general linear statistic $\bar{T}_N$. In addition, some comments on the practical choice of the design parameters $\mathbf{M}$, $\mathbf{L}$, $\mathbf{b}$, $\mathbf{h}$, and l will also be offered.

To start with, consider the choice of l. As indicated also in Section 3.2, l may be taken to be of the same asymptotic order as Q; since $l = kb$, this can easily be accomplished by letting $k = [Q/b]$. Now comparing Theorems 1 and 2, one might be led to think that the choice of $\mathbf{h}$ is immaterial, at least to a first approximation; however it does influence the constant factor in $Var(\hat{V}_{JACK}) = O(b/Q)$, and it is advisable to let $\mathbf{h} = (1, 1, \dots, 1)$. For example, in the case where $\bar{T}_N$ is the sample mean of a stationary sequence it has been shown (cf. Künsch(1989), Brillinger(1981)) that letting $\mathbf{h} = 1$ corresponds to a 33% reduction of $\lim \frac{Q}{b} Var(\hat{V}_{JACK})$ over letting $\mathbf{h} = \mathbf{b}$.

To turn to the choice of $\mathbf{b}$, note that by Corollary 2 an optimal choice would satisfy $b \sim a_b Q^{\frac{n}{n+2}}$, for some constant $a_b > 0$. This could be done by letting the coordinates of $\mathbf{b}$ satisfy $b_k \sim a_b^{\frac{1}{n}} Q^{\frac{1}{n+2}}$, for $k = 1, \dots, n$. However, properly choosing the constant a_b is most important (and difficult) in practice, and is quite analogous to choosing a bandwidth for a spectral estimator; see Politis and Romano (1993) for some practical guidelines in the one-dimensional ($n = 1$) case. Finally, regarding the choice of $\mathbf{M}$ and $\mathbf{L}$, two important and separate cases must be distinguished.

4.1 Parameters associated with a finite-dimensional marginal. For simplicity assume that the random field $\{X(\mathbf{t})\}$ is univariate. If one takes $M_k = L_k = 1$, for all $k = 1, 2, \dots, n$, then the $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L})$ block consists of just the observation $X(\mathbf{i})$. Leting $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ be equal to $X(\mathbf{i})$, it is seen that $\bar{T}_N$ is a regular sample mean, and the parameter μ would be $EX(\mathbf{0})$. Similarly, if $M_k = m$, for all $k = 1, 2, \dots, n$, then the $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L})$ blocks are cubes of side m, and by letting $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ be equal to $X(\mathbf{i})X(\mathbf{i} + \mathbf{s})$, it is seen that $\bar{T}_N$ is the (unbiased) sample autocovariance at lag $\mathbf{s} = (s_1, \dots, s_n)$, where $s_1, \dots, s_n$ are integers with absolute value less than m, provided of course that now $EX(\mathbf{i}) = 0$.

The two abovementioned examples correspond to parameters associated with a finite-dimensional distribution of the random field $\{X(\mathbf{t})\}$, i.e., the distribution of $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E_s}\}$, where $\mathbf{E_s}$ is a *finite* subset of $\mathbf{Z}^n$ consisting of the points $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ such that $1 \leq t_k \leq s_k$, for $k = 1, 2, \ldots, n$. Such parameters can in general be consistently estimated by the linear statistic $\bar{T}_N$, by choosing $M_k = s_k$ and $L_k = 1$, for $k = 1, 2, \ldots, n$, provided the function $\phi_{\mathbf{M}}$ is such that assumption A_2 is satisfied. In this case one would naturally chose $\phi_{\mathbf{M}}$ to make $\bar{T}_N$ an exactly unbiased estimator of μ.

4.2 Parameters associated with the whole infinite-dimensional distribution. If however μ happens to be a parameter of the whole (infinite-dimensional) distribution of the random field $\{X(\mathbf{t})\}$, consistent estimation of μ by $\bar{T}_N$ would generally require taking some (or all) of the M_k's to be increasing as the sample size N increases. The typical example of such a parameter is the spectral density function associated with the random field, evaluated at a point.

Let

$$T_{\mathbf{i},\mathbf{M},\mathbf{L}}(\mathbf{w}) = \frac{1}{(2\pi)^n M} \Big| \sum_{\mathbf{t} \in \mathbf{E}_{\mathbf{i},\mathbf{M},\mathbf{L}}} W_{\mathbf{t}} X(\mathbf{t}) e^{-j(\mathbf{w}\cdot\mathbf{t})} \Big|^2$$

that is, $T_{\mathbf{i},\mathbf{M},\mathbf{L}}(\mathbf{w})$ is the periodogram of block $\mathbf{B}(\mathbf{i}, \mathbf{M}, \mathbf{L})$ of data, 'tapered' by the function $W_{\mathbf{t}}$, and evaluated at some point $\mathbf{w} = (w_1, \ldots, w_n) \in (-\pi, \pi]^n$; note that $(\mathbf{w} \cdot \mathbf{t}) = \sum_{i=1}^n w_i t_i$ is just the inner product in $\mathbf{R}^n$, the symbol j denotes the imaginary unit $(\sqrt{-1})$, and for simplicity it was taken $\mathbf{M} = (M^{1/n}, \ldots, M^{1/n})$. Now define $\bar{T}_N = \frac{1}{Q} \sum_{i=1}^Q T_{\mathbf{i},\mathbf{M},\mathbf{L}}(\mathbf{w})$ as before. This is a so-called lag-window spectral estimator (cf. Zhurbenko(1986)) and it can be shown that, under suitable moment and mixing conditions, and if $M \to \infty$ with $M = o(N)$, $\bar{T}_N$ is a consistent estimator of the spectral density function $f(\mathbf{w})$ which is defined by

$$f(\mathbf{w}) = \frac{1}{(2\pi)^n} \sum_{\mathbf{t} \in \mathbf{Z}^n} R(\mathbf{t}) e^{-j(\mathbf{w}\cdot\mathbf{t})}$$

where $R(\mathbf{t}) = EX(\mathbf{0})X(\mathbf{t})$ is the autocovariance at 'lag' $\mathbf{t}$, and it was implicitly assumed that $EX(\mathbf{0}) = 0$.

For concreteness also assume that $W_{\mathbf{t}} \equiv 1$, i.e., that there is no tapering. As it turns out in this case, to have assumption A_2 satisfied, we must choose $M \sim a_M N^{\beta}$, for some constants

$a_M > 0$ and $1 > \beta > n/(n+2)$, which corresponds to *undersmoothing* the n-dimensional periodogram. Notably, we could chose an *optimal* (from the point of view of mean squared error) smoothing, i.e., $M \sim a_M N^{n/(n+2)}$, and still have assumption A_2' satisfied, which is sufficient for the purposes of just estimating the variance of $\bar{T}_N$.

In addition, to have condition (i) of the theorems satisfied, L_k for $k = 1, \ldots, n$ should be chosen to be approximately proportional to the corresponding M_k. This is not a mere technical assumption but is crucial to the validity of the resampling methodology, although the *point* estimator $\bar{T}_N$ would still be consistent even if $L_k = 1$ for $k = 1, \ldots, n$; see Politis and Romano (1992c) for an elaborate discussion of this phenomenon in the one-dimensional case.

The spectral density example helps outline the general problem associated with estimating parameters of the whole infinite-dimensional distribution of the random field. Loosely speaking, one should choose the function $\phi_{\mathbf{M}}$ such that the bias of $\bar{T}_N$ is as small as possible, as compared to the order of magnitude of its standard deviation (see assumptions A_2 and A_2'). This would invariably require the coordinates of $\mathbf{M}$ to be increasing functions of the sample size N; for consistency of the resampling estimates of variance and sampling distribution of $\bar{T}_N$, the coordinates of $\mathbf{L}$ should then be chosen to be approximately proportional to the corresponding $\mathbf{M}$ coordinate.

5. Technical Proofs

Proof of Lemma 1. (Proof of (a), the other parts being easy consequences of part (a).) Let $\mathbf{k} = (k_1, \ldots, k_n)$, and look at the mixing coefficient between the variables $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ and $T_{\mathbf{i}+\mathbf{k},\mathbf{M},\mathbf{L}}$, i.e., $\alpha\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, T_{\mathbf{i}+\mathbf{k},\mathbf{M},\mathbf{L}}\} = \sup |P(A \cap B) - P(A)P(B)|$, where the supremum is over all sets A, B that are in the σ-algebras generated by $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ and $T_{\mathbf{i}+\mathbf{k},\mathbf{M},\mathbf{L}}$ respectively.

It is easy to see that, if $\max_j |k_j| \geq [M^\star/L^\star] + 1$, then

$$\alpha\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, T_{\mathbf{i}+\mathbf{k},\mathbf{M},\mathbf{L}}\} \leq \alpha_X(\max_j(|k_j|L_j - M_j)) \leq \alpha_X(\max_j |k_j|L^\star - M^\star))$$

Since $\max_j |k_j| = d(\mathbf{i}, \mathbf{i}+\mathbf{k})$, it follows that the mixing coefficient $\alpha\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, T_{\mathbf{i}+\mathbf{k},\mathbf{M},\mathbf{L}}\}$ is bounded above by a function with argument the distance $d(\mathbf{i}, \mathbf{i}+\mathbf{k})$. The same idea can be used to show that the mixing coefficient corresponding to any two sets of $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$'s, e.g., $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_1\}$ and $\{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_2\}$, is bounded above by a function with argument $d(\mathbf{E}_1, \mathbf{E}_2)$, provided $d(\mathbf{E}_1, \mathbf{E}_2) \geq [M^\star/L^\star] + 1$, and that this function is in turn bounded above by $\alpha_X(d(\mathbf{E}_1, \mathbf{E}_2)L^\star - M^\star))$. □

Proof of Lemma 2. Extend the lattice homogeneous random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ to an inhomogeneous random field $\{\chi(\mathbf{t}), \mathbf{t} \in \mathbf{R}^n\}$ by defining $\chi(\mathbf{t}) = X([\mathbf{t}])$, where $[\mathbf{t}] = ([t_1], [t_2], \ldots, [t_n])$, and $[t]$ is the greatest integer $\leq t$. Then Lemma 2 follows from Lemma 1.8.1 of Ivanov and Leonenko (1986) as applied to the random field $\{\chi(\mathbf{t})\}$. □

Proof of Theorem 1. Note that conditions *(i)*, *(ii)* imply that $\frac{N_i}{M_i} \sim a_i Q_i \to \infty$, $i = 1, \ldots, n$.

Denote $\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}} = \frac{1}{\sqrt{b}} \sum_{\mathbf{i} \in \mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}}$. Then

$$\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = \frac{1}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} (\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}} - \bar{T}_N)^2$$

$$= \frac{1}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} \{\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}} - E\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}} - \sqrt{b}(\bar{T}_N - \frac{1}{\sqrt{b}} E\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}})\}^2 = A_N - 2C_N + D_N$$

where

$$A_N = \frac{1}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} (\tilde{B}_{\mathbf{j,b,h}} - E\tilde{B}_{\mathbf{j,b,h}})^2$$

$$C_N = \frac{1}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} \sqrt{b}(\bar{T}_N - \frac{1}{\sqrt{b}} E\tilde{B}_{\mathbf{j,b,h}})(\tilde{B}_{\mathbf{j,b,h}} - E\tilde{B}_{\mathbf{j,b,h}})$$

$$D_N = \frac{1}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} b(\bar{T}_N - \frac{1}{\sqrt{b}} E\tilde{B}_{\mathbf{j,b,h}})^2$$

The proof will consist of showing that A_N is the dominant part of $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$, both in terms of expected value and variance. In other words, C_N and D_N can be considered as negligible error factors.

Let $\xi(\mathbf{j}) = (\tilde{B}_{\mathbf{j,b,h}} - E\tilde{B}_{\mathbf{j,b,h}})^2$, in which case

$$A_N = \frac{1}{q} \sum_{\mathbf{j} \in \mathbf{E}_q} \xi(\mathbf{j}) = \frac{1}{q} \sum_{j_1=1}^{q_1} \sum_{j_2=1}^{q_2} \cdots \sum_{j_n=1}^{q_n} \xi(j_1, j_2, \ldots, j_n)$$

is just the sample mean of the homogeneous, strong mixing random field $\xi(\mathbf{j})$, which is observed on $\mathbf{E}_q$. Because the random field $\{X(\mathbf{t})\}$ is α-mixing with mixing coefficient $\alpha_X(\cdot)$, an argument similar to Lemma 1 shows that the random field $\xi(\mathbf{j})$ is also α-mixing with mixing coefficient $\alpha_\xi(\cdot)$, satisfying

$$\alpha_\xi(s) \leq \alpha_T(sh^\star - b^\star) \leq \alpha_X(sh^\star L^\star - b^\star L^\star - M^\star) \tag{30}$$

provided $s \geq s_0 = [\frac{M^\star}{h^\star L^\star} + \frac{b^\star - 1}{h^\star}] + 1$, where $b^\star = \max_i b_i$, and $h^\star = \min_i h_i$. Since by conditions $(i), (ii)$ we have that $M^\star = O(L^\star)$, and $b^\star = O(h^\star)$, it follows that there will be a smallest s^* that does *not* depend on N, and such that equation (30) will hold for all $s \geq s^*$.

Now $EA_N = E\xi(\mathbf{1}) = V_{b/N}$, and by the homogeneity of the $\xi(\mathbf{j})$ random field

$$VarA_N = \frac{1}{q} \sum_{j_1=-q_1}^{q_1} \sum_{j_2=-q_2}^{q_2} \cdots \sum_{j_n=-q_n}^{q_n} (1 - \frac{|j_1|}{q_1})(1 - \frac{|j_2|}{q_2}) \cdots (1 - \frac{|j_n|}{q_n}) \rho(j_1, j_2, \ldots, j_n)$$

where $\rho(j_1, j_2, \ldots, j_n) = \rho(\mathbf{j}) = Cov\{\xi(\mathbf{1}), \xi(\mathbf{1} + \mathbf{j})\}$. But (cf. Roussas and Ioannides (1987))

$$|Cov\{\xi(\mathbf{1}), \xi(\mathbf{1} + \mathbf{j})\}| \leq 10(E|\xi(\mathbf{1})|^{2p})^{2/p}\{\alpha_\xi(\max_k |j_k|)\}^{\frac{p-2}{p}}$$

By Lemma 2 now, $E|\xi(\mathbf{1})|^p \leq K_X(E|T_{\mathbf{1},\mathbf{M},\mathbf{L}}|^{2p+\delta})^{\frac{2p}{2p+\delta}}$, where K_X depends only on $\alpha_X(\cdot)$ and p. Observe that

$$\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\cdots\sum_{j_n=1}^{\infty}\{\alpha_\xi(\max_k |j_k|)\}^{\frac{p-2}{p}} = n\sum_{j_1=1}^{\infty} W(j_1)\{\alpha_\xi(j_1)\}^{\frac{p-2}{p}} \tag{31}$$

where $W(j_1)$ is the cardinality of the set $\{\mathbf{j} \in \mathbf{Z}^n : j_1 = \max_k j_k, \text{ and } j_k > 0, k = 1, 2, \ldots, n\}$. It can now be easily seen that $W(j_1) \leq j_1^{n-1}$, and from equation (30) it follows that the sum in equation (31) is finite. Invoking assumption A_1, it now follows that $VarA_N = O(1/q) = O(b/Q)$.

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

$$\frac{D_N}{b} = (\bar{T}_N - \mu)^2 + (\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}})^2 + 2(\bar{T}_N - \mu)(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}})$$

Using assumptions A_2' and A_3, we have that $\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}} = O(Q^{-1/2})$, and $E(\bar{T}_N - \mu)^2 = O(Q^{-1})$, and thus it follows that $ED_N = O(\frac{b}{Q})$.

Similarly, look at

$$\frac{D_N^2}{b^2} = (\bar{T}_N - \mu)^4 + (\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}})^4 + 6(\bar{T}_N - \mu)^2(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}})^2 +$$
$$+4(\bar{T}_N - \mu)(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}})^3 + 4(\bar{T}_N - \mu)^3(\mu - \frac{1}{\sqrt{b}}E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}})$$

Again by using Lemmas 1 and 2 and assumption A_2' we have $E(\bar{T}_N - \mu)^4 = O(Q^{-2})$ and $E(\bar{T}_N - \mu)^3 = O(Q^{-3/2})$, from which the result $ED_N^2 = O(\frac{b^2}{Q^2})$ is proved. □

Proof of Theorem 2. Look at the 'whole sample' $\mathcal{S} = \{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_Q\}$, and define the 'reduced samples' $\mathcal{S}_{\mathbf{j}} = \{T_{\mathbf{i},\mathbf{M},\mathbf{L}}, \mathbf{i} \in \mathbf{E}_Q^{(\mathbf{j})}\}$, for $\mathbf{j} = (j_1, j_2, \ldots, j_n)$ such that $1 \leq j_k \leq b_k$, (i.e., $\mathbf{j} \in \mathbf{E}_b$), and where $\mathbf{E}_Q^{(\mathbf{j})}$ is the rectangle consisting of the points $\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \mathbf{Z}^n$ such that $j_k \leq i_k \leq Q_k$, for $k = 1, 2, \ldots, n$. In this notation, $\mathcal{S} \equiv \mathcal{S}_{\mathbf{1}}$.

Let $\hat{V}_{JACK}^{(\mathbf{j})}(\sqrt{Q}\bar{T}_N)$ be the 'blocks of blocks' jackknife estimate of variance computed from the 'reduced sample' $\mathcal{S}_{\mathbf{j}}$, and using $\mathbf{h} = \mathbf{b}$. Since Theorem 1 applies here, it follows that

$$E\hat{V}_{JACK}^{(\mathbf{j})}(\sqrt{Q}\bar{T}_N) = V_{b/N} + O(b/Q) \tag{32}$$

and

$$Var(\hat{V}^{(\mathbf{j})}_{JACK}(\sqrt{Q}\bar{T}_N)) = O(b/Q) \tag{33}$$

for any $\mathbf{j}$ considered.

Now let $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)$ be the 'blocks of blocks' jackknife estimate of variance computed from the 'whole sample' $\mathcal{S}$, and using $\mathbf{h} = (1, 1, \ldots, 1)$. It is easy to see that

$$\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) \sim \frac{1}{b} \sum_{\mathbf{j} \in \mathbf{E}_b} \hat{V}^{(\mathbf{j})}_{JACK}(\sqrt{Q}\bar{T}_N)$$

from which it is immediate that $E\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = V_{b/N} + O(b/Q)$, and, since by the Cauchy-Schwarz inequality and equation (33) we have $Cov(\hat{V}^{(\mathbf{j})}_{JACK}(\sqrt{Q}\bar{T}_N), \hat{V}^{(\mathbf{j}')}_{JACK}(\sqrt{Q}\bar{T}_N)) = O(b/Q)$, it follows that

$$Var(\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N)) \sim \frac{1}{b^2} \sum_{\mathbf{j} \in \mathbf{E}_b} \sum_{\mathbf{j}' \in \mathbf{E}_b} Cov(\hat{V}^{(\mathbf{j})}_{JACK}(\sqrt{Q}\bar{T}_N), \hat{V}^{(\mathbf{j}')}_{JACK}(\sqrt{Q}\bar{T}_N)) = O(b/Q)$$

and the Theorem is proved. □

Proof of Theorem 3. To make the calculations easier assume $b_1 = b_2 = \cdots = b_n = \sqrt[n]{b}$. Now

$$V_{Q/N} = \sum_{j_1=-Q_1}^{Q_1} \sum_{j_2=-Q_2}^{Q_2} \cdots \sum_{j_n=-Q_n}^{Q_n} (1 - \frac{|j_1|}{Q_1})(1 - \frac{|j_2|}{Q_2}) \cdots (1 - \frac{|j_n|}{Q_n}) Cov_T(\mathbf{j}) = \sum_{\mathbf{j} \in \mathbf{E}_Q^{\pm}} K_Q(\mathbf{j}) Cov_T(\mathbf{j})$$

where $\mathbf{E}_Q^{\pm} = \{\mathbf{j} \in \mathbf{Z}^n : |j_k| \leq Q_k, k = 1, \ldots, n\}$, $K_Q(\mathbf{j}) = (1 - \frac{|j_1|}{Q_1})(1 - \frac{|j_2|}{Q_2}) \cdots (1 - \frac{|j_n|}{Q_n})$, and $Cov_T(\mathbf{j}) = Cov(T_{\mathbf{i},\mathbf{M},\mathbf{L}}, T_{\mathbf{i}+\mathbf{j},\mathbf{M},\mathbf{L}})$; note that $Cov_T(\mathbf{j})$ depends on N in general, but it is not explicitly denoted.

Similarly

$$V_{b/N} = \sum_{j_1=-b_1}^{b_1} \sum_{j_2=-b_2}^{b_2} \cdots \sum_{j_n=-b_n}^{b_n} (1 - \frac{|j_1|}{b_1})(1 - \frac{|j_2|}{b_2}) \cdots (1 - \frac{|j_n|}{b_n}) Cov_T(\mathbf{j}) = \sum_{\mathbf{j} \in \mathbf{E}_b^{\pm}} K_b(\mathbf{j}) Cov_T(\mathbf{j})$$

where $\mathbf{E}_b^{\pm} = \{\mathbf{j} \in \mathbf{Z}^n : |j_k| \leq b_k, k = 1, \ldots, n\}$, and $K_b(\mathbf{j}) = (1 - \frac{|j_1|}{b_1})(1 - \frac{|j_2|}{b_2}) \cdots (1 - \frac{|j_n|}{b_n})$.

Look at

$$V_{Q/N} - V_{b/N} = \sum_{\mathbf{j} \in \mathbf{E}_b^{\pm}} (K_Q(\mathbf{j}) - K_b(\mathbf{j})) Cov_T(\mathbf{j}) + \sum_{\mathbf{j} \in \mathbf{E}_Q^{\pm} - \mathbf{E}_b^{\pm}} K_Q(\mathbf{j}) Cov_T(\mathbf{j}) = C_1 + C_2 + C_3$$

where

$$C_1 = \sum_{\mathbf{j} \in \mathbf{E}_m^{\pm}} (K_Q(\mathbf{j}) - K_b(\mathbf{j}))\, Cov_T(\mathbf{j})$$

$$C_2 = \sum_{\mathbf{j} \in \mathbf{E}_b^{\pm} - \mathbf{E}_m^{\pm}} (K_Q(\mathbf{j}) - K_b(\mathbf{j}))\, Cov_T(\mathbf{j})$$

$$C_3 = \sum_{\mathbf{j} \in \mathbf{E}_Q^{\pm} - \mathbf{E}_b^{\pm}} K_Q(\mathbf{j}) Cov_T(\mathbf{j})$$

where $\mathbf{E}_m^{\pm} = \{\mathbf{j} \in \mathbf{Z}^n : |j_k| \leq m, k = 1, \ldots, n\}$, and the constant $m = [\frac{1}{a}]$ is defined in Lemma 1(b); since $b \Rightarrow \infty$, $Q \Rightarrow \infty$, and $b = o(Q)$, it was implicitly assumed that N is large enough so that $m < b_k < Q_k$, $k = 1, \ldots, n$.

Now note that for $\mathbf{j} \in \mathbf{E}_b^{\pm}$, $0 \leq K_Q(\mathbf{j}) - K_b(\mathbf{j}) \leq 1 - (1 - \frac{\max_k |j_k|}{b_1})^n \leq c_1 \frac{\max_k |j_k|}{b_1}$, for some constant $c_1 > 0$, from which it follows immediately that $C_1 = O(\frac{1}{b_1})$. (As a matter of fact, $|K_Q(\mathbf{j}) - K_b(\mathbf{j})|$ is of exact rate $\frac{\max_k |j_k|}{b_1}$, as can be shown by the lower bound $K_Q(\mathbf{j}) - K_b(\mathbf{j}) \geq \frac{\max_k |j_k|}{b_1} - c_2 \frac{\max_k |j_k|}{Q_1} \geq (1 - \epsilon) \frac{\max_k |j_k|}{b_1}$, for large enough N, and for some constants $c_2 > 0$ and $0 < \epsilon < 1$.)

By Ibragimov's strong-mixing inequality (cf. Roussas and Ioannides (1987)) and assumption A_1, it follows that

$$Cov_T(\mathbf{j}) = O(\{\alpha_T(\max_k |j_k|)\}^{\frac{2(p-1)+\delta}{2p+\delta}})$$

But from Lemma 1(b) we have $\alpha_T(k) \leq \alpha_X(kL^\star - M^\star)$, for $k > m = [\frac{1}{a}]$. Combining this with assumption $\alpha_X(k) = O(k^{-\lambda})$, it follows that $\alpha_T(k) = O(kL^\star - M^\star)^{-\lambda} = O((M^\star)^{-\lambda}(ak - 1)^{-\lambda}) = O((ak-1)^{-\lambda})$, for $k \geq m+1$, and thus

$$Cov_T(\mathbf{j}) = O((a \max_k |j_k| - 1)^{-\nu}) = O((\max_k |j_k|)^{-9n})$$

where $\nu > \frac{np}{\delta}(2p - 2 + \delta) \geq 9n$, for $\mathbf{j}$ such that $\max_k |j_k| \geq m + 1$, which actually holds if $\mathbf{j} \notin \mathbf{E}_m^{\pm}$.

As in the proof of Theorem 1, let $W(j_1)$ be the cardinality of the set $\{\mathbf{j} \in \mathbf{Z}^n : j_1 = \max_k j_k$, and $j_k > 0, k = 1, 2, \ldots, n\}$, and note that

$$C_2 = O\left(\sum_{\mathbf{j} \in \mathbf{E}_b^{\pm} - \mathbf{E}_m^{\pm}} \frac{\max_k |j_k|}{b_1} Cov_T(\mathbf{j}) \right) = O\left(\sum_{j_1 = m+1}^{b_1} W(j_1) \frac{j_1}{b_1} j_1^{-9n} \right) = O\left(\frac{1}{b_1} \sum_{j_1 = m+1}^{b_1} j_1^{-8n} \right) = O(\frac{1}{b_1})$$

Similarly

$$C_3 = O\left(\sum_{\mathbf{j}\in\mathbf{E}_Q^{\pm}-\mathbf{E}_b^{\pm}} Cov_T(\mathbf{j})\right) = O\left(\sum_{j_1=b_1+1}^{Q^*} W(j_1)j_1^{-9n}\right) = O\left(\sum_{j_1=b_1+1}^{Q^*} W(j_1)\frac{j_1}{b_1}j_1^{-9n}\right) = o(\frac{1}{b_1})$$

where $Q^* = \max_k Q_k$, and equation (13) is proven.

By an almost identical calculation equation (14) is also proven. □

PROOF OF THEOREM 4. First look at

$$Var^*(\sqrt{l}\bar{T}_l^*) = lVar^*\{\frac{1}{k}\sum_{i=1}^{k}\bar{Y}_i\} = Var^*\tilde{Y}_1$$

where $\tilde{Y}_i = \sqrt{b}\bar{Y}_i$, $i = 1, \ldots, k$, and the facts that the $\bar{Y}_i$'s are i.i.d. under P^*, and $l = kb$ were used. But

$$Var^*\tilde{Y}_1 = \frac{1}{q}\sum_{\mathbf{j}\in\mathbf{E}_q}(\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}} - E^*\tilde{Y}_1)^2$$

where

$$E^*\tilde{Y}_1 = \frac{1}{q}\sum_{\mathbf{j}\in\mathbf{E}_q}\frac{1}{\sqrt{b}}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}}$$

In this double sum, each $T_{\mathbf{i},\mathbf{M},\mathbf{L}}$ with index $\mathbf{i}$ such that $b_k \le i_k \le Q_k - b_k$, $k = 1, \ldots, n$, is represented exactly $\prod_{k=1}^{n}[\frac{b_k}{h_k} + 1] \sim b/h$ times. This implies that

$$E^*\tilde{Y}_1 = \frac{\prod_{k=1}^{n}[\frac{b_k}{h_k} + 1]}{q\sqrt{b}}\sum_{\mathbf{i}\in\mathbf{E}_Q} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - \frac{1}{q\sqrt{b}}O_{MS}(b^2/h)$$

where the notation $Z_N = O_{MS}(m_N)$, with Z_N representing a sequence of random variables, and m_N a sequence of numbers, means that $EZ_N^2 = O(m_N^2)$, and it was used that, by assumption A_1 and Lyapunov's inequality, each $T_{\mathbf{i},\mathbf{M},\mathbf{L}} = O_{MS}(1)$; note that the 'order in mean square' O_{MS} is quite weaker than the definition of 'expected order' O_E of Woodroofe (1970), but stronger than the order 'in probability' O_P.

Recalling that $q \sim Q/h$ and that $b = o(Q)$ yields:

$$|E^*\tilde{Y}_1 - \sqrt{b}\bar{T}_N| = O_{MS}(\frac{b^{3/2}}{hq}) = O_{MS}(\frac{b^{3/2}}{Q})$$

Therefore

$$Var^*\tilde{Y}_1 = \frac{1}{q}\sum_{i=1}^{q}(\tilde{B}_i - \sqrt{b}\bar{T}_N + O_{MS}(\frac{b^{3/2}}{Q}))^2 = \hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) + O_{MS}(\frac{b^{3/2}}{Q}) \qquad (34)$$

Now by Corolllary 1, and condition (iii'), equation (18) follows.

Since by assumptions A_2, A_3 we have that

$$\sup_x |P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x\} - \Phi(x/\sigma_\infty)| \to 0$$

to prove equation (21), it suffices to show that

$$\sup_x |P^*\{\frac{\bar{T}_l^* - E^*\bar{T}_l^*}{Var^*\bar{T}_l^*} \leq x\} - \Phi(x)| \xrightarrow{P} 0 \qquad (35)$$

Note that $\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) = \sqrt{k}\sum_{i=1}^{k}(\tilde{Y}_i - E^*\tilde{Y}_i)$, where the $\tilde{Y}_i$'s are i.i.d under P^*. Recall that the theorem's assumptions imply that $k \to \infty$, and therefore by the Berry-Esseen theorem it follows that (35) holds provided $E^*|\tilde{Y}_1 - E^*\tilde{Y}_1|^3$ is bounded in probability. However,

$$E^*|\tilde{Y}_1 - E^*\tilde{Y}_1|^3 = \frac{1}{q}\sum_{\mathbf{i}\in\mathbf{E}_q}|\tilde{B}_{\mathbf{i},\mathbf{b},\mathbf{h}} - \frac{1}{q}\sum_{\mathbf{j}\in\mathbf{E}_q}\tilde{B}_{\mathbf{j},\mathbf{b},\mathbf{h}}|^3 \leq \frac{1}{q}\{(\sum_{\mathbf{i}\in\mathbf{E}_q}|\tilde{B}_{\mathbf{i},\mathbf{b},\mathbf{h}}|^3)^{\frac{1}{3}} + q^{-2/3}|\sum_{\mathbf{i}\in\mathbf{E}_q}\tilde{B}_{\mathbf{i},\mathbf{b},\mathbf{h}}|\}$$

where the quantity on the right-hand-side converges in probability to $\{(E|\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}}|^3)^{1/3} + |E\tilde{B}_{\mathbf{1},\mathbf{b},\mathbf{h}}|\}^3$, because a weak law of large numbers holds for the α-mixing random fields $\tilde{B}_{\mathbf{i},\mathbf{b},\mathbf{h}}$ and $|\tilde{B}_{\mathbf{i},\mathbf{b},\mathbf{h}}|^3$, $\mathbf{i} \in \mathbf{Z}^n$, similarly as in the proof of Theorem 1. Hence equation (19) is proved.

To complete the proof, note that equations (20), (21) follow immediately from equations (18), (19), and assumption A_2 by Slutsky's theorem. □

PROOF OF THEOREM 5. Since $E^\star\bar{T}_l^\star = \frac{1}{q^\star}\sum_{\mathbf{j}\in\mathbf{E}_{q^\star}}\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} = \bar{T}_N$, we will focus on $Var^\star(\sqrt{l}\bar{T}_l^\star)$.

We have

$$Var^\star(\sqrt{l}\bar{T}_l^\star) = \frac{b}{Q}\sum_{\mathbf{j}\in\mathbf{E}_Q}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - \bar{T}_N)^2$$

$$= \frac{b}{Q}\{\sum_{\mathbf{j}\in\mathbf{E}_q}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - \bar{T}_N)^2 + \sum_{\mathbf{j}\in\mathbf{E}_Q-\mathbf{E}_q}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}} T_{\mathbf{i},\mathbf{M},\mathbf{L}} - \bar{T}_N)^2\}$$

Observe that in the second summation above we have a sum of $Q - q$ terms, each of which is of order $O_{MS}(1/b)$. This can be shown by the same argument used in the proof of Theorem 1 to show that the dominant term in the sum

$$\frac{b}{Q}\sum_{\mathbf{j}}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E_{j,b,h}}} T_{\mathbf{i,M,L}} - \bar{T}_N)^2$$

is actually

$$\frac{b}{Q}\sum_{\mathbf{j}}(\frac{1}{b}\sum_{\mathbf{i}\in\mathbf{E_{j,b,h}}} T_{\mathbf{i,M,L}} - ET_{\mathbf{1,M,L}})^2$$

and an application of Lemma 1 (b) and Lemma 2.

Note that $Q - q = O(\sum_{i=1}^{n} b_i \prod_{j\neq i} Q_j) = O(b^{\frac{1}{n}} Q^{\frac{n-1}{n}})$, since by assumptions $\mathbf{b} \Rightarrow \infty$ and $\mathbf{Q} \Rightarrow \infty$, all the b_i's are of order $b^{1/n}$, and all the Q_i's are of order $Q^{1/n}$. Thus

$$Var^{\star}(\sqrt{l}\bar{T}_l^{\star}) = \frac{q}{Q}\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) + \frac{b}{Q}O_{MS}(\sqrt[n]{(\frac{Q}{b})^{n-1}}) = \frac{q}{Q}\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) + O_{MS}(\sqrt[n]{\frac{b}{Q}})$$

Recall that $\mathbf{q} = (q_1, \ldots, q_n)$, with $q_i = Q_i - b_i + 1$, $i = 1, \ldots, n$. It follows that $1 - \frac{q}{Q} = O(\sqrt[n]{\frac{b}{Q}})$, and so finally

$$Var^{\star}(\sqrt{l}\bar{T}_l^{\star}) = \hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) + O_{MS}(\sqrt[n]{\frac{b}{Q}}) \tag{36}$$

Now by Corollary 1, and condition (iii), equation (25) follows.

By a similar argument to the proof of Theorem 4 it is shown that $\sqrt{l}(\bar{T}_l^{\star} - \bar{T}_N)$ is asymptotically normal (with high probability) and the theorem is proved. □

PROOF OF THEOREM 6. The proof follows easily by the same arguments as in Politis and Romano (1992a), using as a reference the univariate theorems concerning random fields of the present paper.□

References

[1] Bolthausen, E. (1982), On the central limit theorem for stationary random fields, *Ann. Prob.*, 10, 1047-1050.

[2] Bradley, R.C. (1992), On the spectral density and asymptotic normality of weakly dependent random fields, *J. Theor. Prob.*, vol. 5, no. 2, 355-373.

[3] Brillinger, D. (1981), *Time Series: Data Analusis and Theory, 2nd ed.*, Holden-Day, San Fransisco.

[4] Cressie, N. (1991), *Statistics for Spatial Data*, John Wiley, New York.

[5] Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *Ann. Statist.*, 7, 1-26.

[6] Efron, B. (1982),*The Jackknife, the Bootstrap, and other Resampling Plans*, SIAM NSF-CBMS, Monograph 38.

[7] Hall, P. (1985), Resampling a coverage pattern, *Stoch. Process. Appl.*, 20, 231-246.

[8] Hall, P. (1988), On confidence intervals for spatial parameters estimated from nonreplicated data, *Biometrics*, 44, 271-277.

[9] Ivanov, A.V. and Leonenko, N.N. (1986), *Statistical analysis of random fields*, Kluwer Academic Publishers, The Netherlands.

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

[11] Lahiri, S.N.(1991), Second order optimality of stationary bootstrap, *Statist. Prob. Letters*, 11, 335-341.

[12] Lele, S. (1991), Jackknifing linear estimation equations: Asymptotic theory and applications in stochastic processes, *J. Royal Statist. Soc., Ser. B*, 53, pp. 253-267.

[13] Liu, R.Y. and Singh, K. (1992), Moving Blocks Jackknife and Bootstrap Capture Weak Dependence, in *Exploring the Limits of Bootstrap*, (edited by Raoul LePage and Lynne Billard), John Wiley, pp. 225-248.

[14] Moore, M. (1988), Spatial linear processes, *Commun. Statist. – Stochastic Models*, 4 (1), 45-75.

[15] Politis, D.N. and Romano, J.P. (1992a), 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*, (edited by Connie Page and Raoul LePage), Springer-Verlag, pp. 98-103.

[16] Politis, D.N. and Romano, J.P. (1992b), A circular block-resampling procedure for stationary data, in *Exploring the Limits of Bootstrap*, (edited by Raoul LePage and Lynne Billard), John Wiley, pp. 263-270.

[17] Politis, D.N. and Romano, J.P. (1992c), A General Resampling Scheme for Triangular Arrays of α-mixing Random Variables with application to the problem of Spectral Density Estimation, *Ann. Statist.*, vol. 20, No. 4.

[18] Politis, D.N. and Romano, J.P. (1993), On the Sample Variance of Linear Statistics Derived from Mixing Sequences, to appear in *Stoch. Proc. Appl.*.

[19] Raïs, N. (1992), *Méthodes de reéchantillonage et de sous échantillonage dans le contexte spatial et pour des données dépendantes* , Ph.D. thesis, Department of Mathematics and Statistics, University of Montreal, Montreal, Canada.

[20] Raïs, N. and Moore, M. (1990), Bootstrap for some stationary α-mixing processes, Abstract, *INTERFACE '90*, 22nd Symposium on the Interface of Computing Science and Statistics.

[21] Rosenblatt, M. (1985), *Stationary sequences and random fields*, Birkhäuser, Boston.

[22] Roussas, G.G. and Ioannides, D. (1987), Moment Inequalities for Mixing Sequences of Random Variables, *Stoch. Analysis and Applications* , 5(1), p.61-120, Marcel Dekker.

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

[24] Tikhomirov, A.N., (1983), On normal approximation of sums of vector-valued random fields with mixing, *Soviet Math. Dokl.*, 28, No.2, 396-397.

[25] Tjøstheim, D. (1978), Statistical spatial series modelling, *Adv. Applied Prob.*, 10, 130-154.

[26] Woodroofe, M. (1970), On choosing a delta-sequence, *Ann. Math. Statist.*, vol. 41, no. 5, 1665-1671.

[27] Yokoyama, R. (1980), Moment Bounds for Stationary Mixing Sequences, *Z. Wahrsch. verw. Gebiete* , 52, p.45-57.

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