The Stationary Bootstrap

Author(s): Dimitris N. Politis and Joseph P. Romano
Source: *Journal of the American Statistical Association*, Vol. 89, No. 428 (Dec., 1994), pp. 1303-1313
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2290993
Accessed: 10-04-2025 12:45 UTC

The Stationary Bootstrap

Dimitris N. POLITIS and Joseph P. ROMANO*

This article introduces a resampling procedure called the stationary bootstrap as a means of calculating standard errors of estimators and constructing confidence regions for parameters based on weakly dependent stationary observations. Previously, a technique based on resampling blocks of consecutive observations was introduced to construct confidence intervals for a parameter of the m-dimensional joint distribution of m consecutive observations, where m is fixed. This procedure has been generalized by constructing a "blocks of blocks" resampling scheme that yields asymptotically valid procedures even for a multivariate parameter of the whole (i.e., infinite-dimensional) joint distribution of the stationary sequence of observations. These methods share the construction of resampling blocks of observations to form a pseudo–time series, so that the statistic of interest may be recalculated based on the resampled data set. But in the context of applying this method to stationary data, it is natural to require the resampled pseudo–time series to be stationary (conditional on the original data) as well. Although the aforementioned procedures lack this property, the stationary procedure developed here is indeed stationary and possesses other desirable properties. The stationary procedure is based on resampling blocks of random length, where the length of each block has a geometric distribution. In this article, fundamental consistency and weak convergence properties of the stationary resampling scheme are developed.

KEY WORDS: Approximate confidence limit; Time Series.

1. INTRODUCTION

The bootstrap of Efron (1979) has proven to be a powerful nonparametric tool for approximating the sampling distribution and variance of complicated statistics based on iid observations. Recently, Künsch (1989) and Liu and Singh (1992) have independently introduced nonparametric versions of the bootstrap and jackknife that are applicable to weakly dependent stationary observations. Their resampling technique amounts to resampling or deleting one-by-one whole blocks of observations, to obtain consistent procedures for a parameter of the m-dimensional marginal distribution of the stationary series. Their resampling procedure has been generalized by Politis and Romano (1992a, 1992b) and by Politis, Romano, and Lai (1992) by resampling "blocks of blocks" of observations to obtain asymptotically valid procedures even for multivariate parameters of the whole (i.e., infinite-dimensional) joint distribution of the stationary time series.

In this article we introduce a new resampling method, called the stationary bootstrap, that is also generally applicable for stationary weakly dependent time series. Similar to the block resampling techniques, the stationary bootstrap involves resampling the original data to form a pseudo–time series from which the statistic or quantity of interest may be recalculated; this resampling procedure is repeated to build up an approximation to the sampling distribution of the statistic. In contrast to the aforementioned block resampling methods, the pseudo–time series generated by the stationary bootstrap method is actually a stationary time series. That is, conditional on the original data $X_1, \ldots, X_N$, a pseudo–time series $X_1^*, \ldots, X_N^*$ is generated by an appropriate resampling scheme that is actually stationary. Hence this procedure attempts to mimic the original model by retaining the stationarity property of the original series in the resampled pseudo–time series. As will be seen, the pseudo–time series is generated by resampling blocks of random size, where the length of each block has a geometric distribution.

In Section 2 the actual construction of the stationary bootstrap is presented and comparisons are made with the block resampling method of Künsch (1989) and Liu and Singh (1992). Some theoretical properties of the method are investigated in Section 3 in the case of the mean. In Section 4 it is shown how the theory may be extended beyond the case of the mean to construct asymptotically valid confidence regions for general parameters.

2. THE STATIONARY BOOTSTRAP RESAMPLING SCHEME

Suppose that $\{X_n, n \in \mathbf{Z}\}$ is a strictly stationary and weakly dependent time series, where the X_n are for now assumed real-valued. Suppose that μ is a parameter of the whole (i.e., infinite-dimensional) joint distribution of the sequence $\{X_n, n \in \mathbf{Z}\}$. For example, μ might be the mean of the process or the spectral distribution function. Given data $X_1, \ldots X_N$, the goal is to make inferences about μ based on some estimator $T_N = T_N(X_1, \ldots X_N)$. In particular, we are interested in constructing a confidence region for μ. Typically, an estimate of the sampling distribution of T_N is required, and the stationary bootstrap method proposed here is developed for this purpose. In general, we are led to considering a "root" or an approximate pivot $R_N = R_N(X_1, \ldots, X_N; \mu)$, which is just some functional depending on the data and possibly on μ as well. For example, R_N might be of the form $R_N = T_N - \mu$ or possibly a studentized version. The idea is that if the true sampling distribution of R_N were known, then probability statements about R_N could be inverted to yield confidence statements about μ. The stationary bootstrap is a method that can be applied to approximate the distribution of R_N. To describe the algorithm, let

$$B_{i,b} = \{X_i, X_{i+1}, \ldots, X_{i+b-1}\} \quad (1)$$

be the block consisting of b observations starting from X_i. In the case $j > N$, X_j is defined to be X_i, where $i = j(\text{mod } N)$ and $X_0 = X_N$. Let p be a fixed number in $[0, 1]$. Independent of $X_1, \ldots, X_N$ let $L_1, L_2, \ldots$ be a sequence of iid random

* Dimitris N. Politis is Assistant Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907. Joseph P. Romano is Associate Professor, Department of Statistics, Stanford University, Stanford, CA 94305.

Journal of the American Statistical Association
December 1994, Vol. 89, No. 428, Theory and Methods

variables having the geometric distribution, so that the probability of the event $\{L_i = m\}$ is $(1-p)^{m-1}p$ for $m = 1, 2, \ldots$. Independent of the X_i and the L_i, let $I_1, I_2, \ldots$ be a sequence of iid variables that have the discrete uniform distribution on $\{1, \ldots, N\}$. Now, a pseudo–time series $X_1^*, \ldots, X_N^*$ is generated in the following way. Sample a sequence of blocks of random length by the prescription $B_{I_1,L_1}, B_{I_2,L_2}, \ldots$. The first L_1 observations in the pseudo–time series $X_1^*, \ldots, X_N^*$ are determined by the first block B_{I_1,L_1} of observations $X_{I_1}, \ldots, X_{I_1+L_1-1}$, and the next L_2 observations in the pseudo–time series are the observations in the second sampled block B_{I_2,L_2}, namely $X_{I_2}, \ldots, X_{I_2+L_2-1}$. Of course, this process is stopped once N observations in the pseudo–time series have been generated (though it is clear that the resampling method allows for time series of arbitrary length to be generated). Once $X_1^*, \ldots, X_N^*$ has been generated, compute $T_N(X_1^*, \ldots, X_N^*)$ or $R_N(X_1^*, \ldots, X_N^*; T_N)$ for the pseudo–time series. The conditional distribution of $R_N(X_1^*, \ldots, X_N^*; T_N)$ given $X_1, \ldots, X_N$ is the stationary bootstrap approximation to the true sampling distribution of $R_N(X_1, \ldots, X_n, \mu)$. By simulating a large number B of pseudo–time series in the same manner, the true distribution of $R_N(X_1, \ldots, X_N; \mu)$ can be approximated by the empirical distribution of the B numbers $R_N(X_1^*, \ldots, X_N^*; T_N)$.

An alternative and perhaps simpler description of the resampling algorithm follows. Let X_1^* be picked at random from the original N observations, so that $X_1^* = X_{I_1}$. With probability p, let X_2^* be picked at random from the original N observations; with probability $1-p$, let $X_2^* = X_{I_1+1}$ so that X_2^* would be the "next" observation in the original time series following X_{I_1}. In general, given that X_i^* is determined by the Jth observation X_J in the original time series, let X_{i+1}^* be equal to X_{J+1} with probability $1-p$ and picked at random from the original N observations with probability p.

Proposition 1. Conditional on $X_1, \ldots, X_N$, $X_1^*, X_2^*, \ldots, X_N^*$ is stationary.

Much more is actually true. For example, if the original observations $X_1, \ldots, X_N$ are all distinct, then the new series $X_1^*, \ldots X_N^*$ is, conditional on $X_1, \ldots, X_N$, a stationary Markov chain. If, on the other hand, two of the original observations are identical and the remaining are distinct, then the new series $X_1^*, \ldots, X_N^*$ is a stationary second-order Markov chain. An obvious generalization, depending on the number of identical subsequences of observations, can be made. In fact, if m is the largest b such that, for some i distinct from j (and both i and j between 1 and N), $B_{i,b}$ and $B_{j,b}$ are identical (and $m = 0$ if all observations are distinct), then the series $X_1^*, \ldots, X_N^*$ is a $(m+1)$-order Markov chain.

The stationary bootstrap resampling scheme proposed here is distinct from that proposed by Künsch (1989) and Liu and Singh (1992). Their "moving blocks" method is described as follows. Suppose that $N = kb$. Resample with replacement from the blocks $B_{1,b}, \ldots, B_{N-b+1,b}$ to get k resampled blocks, say $B_1^*, \ldots, B_k^*$. The first b observations in the pseudo–time series are the sequence of b values in B_1^*, the next b observations in the pseudo–time series are the b values in B_2^*, and so on. In the case, N is not divisible by b, let k be the smallest integer satisfying $bk > N$. Resample k blocks as previously to generate $X_1^*, \ldots, X_{bk}^*$. Now simply delete the observations X_j^* for $j > N$.

Some of the similarities and differences between the stationary bootstrap and the moving blocks bootstrap algorithms should be apparent. To begin, the pseudo–time series generated by the moving blocks method is not stationary. Both methods involve resampling blocks of observations. In the moving blocks technique, the number of observations in each block is a fixed number b. In the stationary bootstrap method, the number of observations in each block is random and has a geometric distribution. The methods also differ in how they deal with end effects. For example, because there is no data after X_N, the moving blocks method does not define a block of length b beginning at X_N (if $b > 1$). To achieve stationarity for the resampled time series, the stationary bootstrap method "wraps" the data around in a "circle," so that X_1 "follows" X_N.

Variants on the stationary bootstrap based on resampling blocks of random length are possible. Instead of assuming that the L_i have a geometric distribution, one can consider other distributions. Alternative distributions for the I_i can be used as well. In this way the moving blocks may be viewed as a special case. The choice of L_i having a geometric distribution and I_i as the discrete uniform distribution was made so that the resampled series is stationary. Of course, other resampling schemes achieve stationary for the resamples series. For example, one could take the series $X_1^*, \ldots, X_N^*$ as previously constructed and add an independent series $Z_1^*, \ldots, Z_N^*$ to it, as a "smoothing" device. For the sake of concreteness, attention will focus on the particular scheme that we initially proposed.

Another way to think about the difference between the moving blocks method and the stationary bootstrap is as follows. For each fixed block size b, one can compute a bootstrap distribution or an estimate of standard error of an estimator. The stationary bootstrap method proposed here is essentially a weighted average of these moving blocks bootstrap distributions or estimates of standard error, where the weights are determined by a geometric distribution. It is important to keep in mind that a difficult aspect in applying these methods is how to choose b in the moving blocks scheme and how to choose p in the stationary scheme. Indeed, the issue becomes a "smoothing" problem.

3. THE MEAN

In this section, the special case of the sample mean is considered as a first step to justify the validity of the stationary bootstrap resampling scheme. Let $\mu = E(X_1)$ and set $T_N(X_1, \ldots, X_N) = \bar{X}_N = N^{-1}\sum_{i=1}^N X_i$. Note that under stationarity, if σ_N^2 is defined to be the variance of $N^{1/2}\bar{X}_N$, then

$$\sigma_N^2 = \operatorname{var}(X_1) + 2\sum_{i=1}^{N}\left(1 - \frac{i}{N}\right)\operatorname{cov}(X_1, X_{1+i}). \qquad (2)$$

Under the assumption that $\sum_{j=1}^{\infty} |\operatorname{cov}(X_1, X_j)| < \infty$, which is implied by typical assumptions of weak dependence, it follows that $\hat{\sigma}_N^2 \to \sigma_\infty^2$ as $N \to \infty$, where

$$\sigma_\infty^2 = \text{var}(X_1) + 2 \sum_{i=1}^{\infty} \text{cov}(X_1, X_{1+i}). \tag{3}$$

Moreover, we typically have that $R_N(X_1, \ldots, X_N; \mu) \equiv N^{1/2}(\bar{X}_N - \mu)$ tends in distribution to the normal distribution with mean 0 and variance σ_∞^2. A primary goal of this section is to establish the validity of the stationary bootstrap approximation defined by the conditional distribution of $R_N(X_1^*, \ldots, X_N^*; \bar{X}_N)$ given the data.

As a first step toward this end, and of interest in its own right, we first consider the mean and variance of $N^{1/2}\bar{X}_N^*$ (conditional on the data), where $\bar{X}_N^* = N^{-1} \sum_{i=1}^N X_i^*$. Because $E(X_1^* \mid X_1, \ldots, X_N) = \bar{X}_N$, a trivial consequence of stationarity is $E(\bar{X}_N^* \mid X_1, \ldots, X_N) = \bar{X}_N$. Because the true distribution of $N^{1/2}(\bar{X}_N - \mu)$ has mean 0, it follows that the bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ has this same mean.

Remark 1. For the moving blocks scheme, it is not the case that $E(\bar{X}_N^* \mid X_{13}, \ldots, X_N) = \bar{X}_N$. It is easy to see that

$$E(\bar{X}_N^* \mid X_1, \ldots, X_N) = \frac{\sum_{i=1}^{b-1} i(X_i + X_{N-i+1}) + b \sum_{j=b}^{N-b+1} X_j}{(N-b+1)b}. \tag{4}$$

Thus if $b/N \to 0$ as $N \to \infty$, $E(\bar{X}_N^* \mid X_1, \ldots, X_N) = \bar{X}_N + O_P(b/N)$. To see why, simply calculate the mean and variance of $E(\bar{X}_N^* \mid X_1, \ldots, X_N) - \bar{X}_N$ with the aid of (4) or see the proof of (iii) in theorem 6 of Liu and Singh (1992). In summary, the moving blocks bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ has a mean that is $O_P(b/N^{1/2})$ as $N \to \infty$ and $b/N \to 0$. As demonstrated by Liu and Singh (1992), to achieve consistency of the moving blocks bootstrap estimate of variance of $N^{1/2}\bar{X}_N$, it is necessary that $b \to \infty$ as $N \to \infty$. Moreover, Künsch (1989) proved that the choice $b \propto N^{1/3}$ is optimal to minimize the mean squared error of the moving blocks bootstrap estimate of variance. For such a choice, the moving blocks bootstrap distribution is centered at a location, $O_P(b/N^{1/2}) = O_P(N^{-1/6})$, which tends to zero quite slowly. Thus one cannot expect the moving blocks bootstrap to possess any second-order optimality properties, at least not without correcting for the bias by recentering the bootstrap distribution. One possibility is to approximate the distribution of $N^{1/2}(\bar{X}_N - \mu)$ by the (conditional) distribution of $N^{1/2}[\bar{X}_N^* - E(\bar{X}_N^* \mid X_1, \ldots, X_N)]$ (see Lahiri 1992). Such an approach may be satisfactory in the case of the mean, but it weakens the claim that the bootstrap is supposed to be a general purpose "automatic" technique. Moreover, this approach would not work as well outside the case of the mean. That is, in the general context of estimating a parameter μ by some estimator $T_N = T_N(X_1, \ldots, X_N)$, consider the approximation to the sampling distribution of $N^{1/2}(T_N - \mu)$ by the (conditional) distribution of $N^{1/2}[T_N^* - E(T_N^* \mid X_1, \ldots, X_N)]$, where $T_N^* = T_N(X_1^*, \ldots, X_N^*)$. In this case the approximating bootstrap distribution necessarily has mean 0 and hence does not account for the bias of T_N as an estimator of μ (unless T_N has zero bias).

Remark 2. In fact, if we consider the more general (possibly nonstationary) resampling scheme where the L_i's are iid with a common (possibly nongeometric) distribution, but the I_i's are iid uniform on $\{1, \ldots, N\}$, then the conditional mean of $\bar{X}_N^*$ is $\bar{X}_N$. In particular, a close cousin of the moving blocks bootstrap scheme that yields the correct (conditional) mean for the corresponding bootstrap distribution is obtained by letting L_i be the distribution assigning mass one to a fixed b (see Politis and Romano 1992c).

We now consider the stationary bootstrap estimate of variance of $N^{1/2}\bar{X}_N$ defined by $\hat{\sigma}_{N,p}^2 \equiv \text{var}(N^{1/2}\bar{X}_N^* \mid X_1, \ldots, X_N)$. In Lemma 1, a formula for $\hat{\sigma}_{N,p}^2$ is obtained, so that $\hat{\sigma}_{N,p}^2$ may be calculated without resampling. In the lemma, $\hat{\sigma}_{N,p}^2$ is given in terms of the circular autocovariances, defined by

$$\hat{C}_N(i) = \frac{1}{N} \sum_{j=1}^{N} [(X_j - \bar{X}_N)(X_{j+i} - \bar{X}_N)],$$

and the usual covariance estimates,

$$\hat{R}_N(i) = \frac{1}{N} \sum_{j=1}^{N} [(X_j - \bar{X}_N)(X_{j+i} - \bar{X}_N)].$$

Lemma 1.

$$\hat{\sigma}_{N,p}^2 = \hat{C}_N(0) + 2 \sum_{i=1}^{N-1} \left(1 - \frac{i}{N}\right)(1-p)^i \hat{C}_N(i). \tag{5}$$

Alternatively,

$$\hat{\sigma}_{N,p}^2 = \hat{R}_N(0) + 2 \sum_{i=1}^{N-1} b_N(i) \hat{R}_N(i), \tag{6}$$

where

$$b_N(i) = \left(1 - \frac{i}{N}\right)(1-p)^i + \frac{i}{N}(1-p)^{N-i}. \tag{7}$$

Evidently, Lemma 1 tells us that the bootstrap estimate of variance $\hat{\sigma}_{N,p}^2$, given by (6), is closely related to a lag window spectral density estimate of $f(0)$, where $f(\cdot)$ is the spectral density of the original process. Assuming that $f(\cdot)$ exists (which it does under summability of covariances), $f(0)$ is simply $\sigma_\infty^2/2\pi$, where σ_∞^2 is given by (3). Hence it is clear that, accounting for the factor $1/2\pi$, estimating σ_∞^2 or σ_N^2 [given in (2)] is equivalent to estimating $f(0)$ in a first-order asymptotic sense. We now prove a consistency property of $\hat{\sigma}_{N,p}^2$. Although many authors have developed theorems on the consistency properties of spectral estimates, including Priestley (1981), Zurbenko (1986), and Brillinger (1981), none fits easily in our framework. In Theorem 1, $\kappa_4(s, r, v)$ is the fourth joint cumulant of the distribution of $(xj, X_{j+r}, X_{j+s}, X_{j+s+r+v})$. The assumptions of the theorem are similar to those used by Brillinger (1981) and Rosenblatt (1984).

Theorem 1. Let $X_1, X_2, \ldots$ be a strictly stationary process with covariance function $R(\cdot)$ satisfying $R(0) + \sum_r |rR(r)| < \infty$. Assume that $p = p_N \to 0$, $Np_N \to \infty$, and

$$\sum_{u,v,w} |\kappa_4(u, v, w)| = K < \infty. \tag{8}$$

Then the bootstrap estimate of variance $\hat{\sigma}^2_{N,p_N}$ tends to σ^2_∞ in probability.

In fact, with only slightly more effort, it can be shown that, under the same conditions of Theorem 1, $\hat{\sigma}^2_{N,p_N}$ tends to σ^2_∞ in the sense $E(\hat{\sigma}^2_{N,p_N} - \sigma^2_\infty)^2 \to 0$. The proof actually shows much more. In particular [see (19)],

$$E(\hat{\sigma}^2_{N,p_N}) = \sigma^2_N - 2p_N \sum_{i=1}^{\infty} iR(i) + o(p_N) \tag{9}$$

and $\text{var}(\hat{\sigma}^2_{N,p_N}) = O(1/Np_N)$. Consequently, if the goal is to choose $p = p_N$ so that the mean squared error of $\hat{\sigma}^2_{N,p_N}$ as an estimator of σ^2_N is minimized, then the order of the squared bias, p^2_N. should be the same order as the variance, $(Np_N)^{-1}$. This occurs if $p_N \propto N^{-1/3}$. The calculation also points toward the difficulty in choosing p optimally. For if the goal remains minimizing the mean squared error of $\hat{\sigma}^2_{N,p}$, then p_N should satisfy $N^{1/3}p_N \to c$, where the constant c depends on intricate properties of the original process, such as $\sum_i iR(i)$. Estimation of this constant c appears difficult. Fortunately, fundamental consistency properties of the bootstrap are unaffected by not choosing p optimally. It is important to have p tending to 0 at the proper rate to achieve second-order properties, but getting the constant c right seems to enter in third-order properties.

Remark 3. We now compare the stationary bootstrap estimate of variance, $\hat{\sigma}^2_{N,p}$, with the moving blocks bootstrap estimate of variance. Suppose, for simplicity, that $N = kb$. Then the moving blocks bootstrap estimate of variance is $k/N \cdot \text{var}(X^*_1 + \cdots + X^*_b \mid X_1, \ldots X_N)$, where $(X^*_1, \ldots, X^*_b)$ is a block of fixed length b chosen at random from $B_{1,b}, \ldots B_{N-b,b}$. Except for end effects, the moving blocks bootstrap estimate of variance is equivalent to $m^2_{N,b} \equiv b^{-1}\text{var}(S_{I,b} \mid X_1, \ldots, X_N)$, where $S_{i,b}$ is the sum of the observations in $B_{i,b}$ defined in (1) and I is chosen at random from $\{1, \ldots, N\}$. By an argument similar to Lemma 1,

$$m^2_{N,b} = \hat{C}_N(0) + 2\sum_{i=1}^{b-1}\left(1 - \frac{i}{b}\right)\hat{C}_N(i). \tag{10}$$

Comparing $m^2_{N,b}$ with $\hat{\sigma}^2_{N,p}$ in (5), the two are quite close, in view of the approximation $(1 - iN^{-1})(1 - p)^i \approx 1 - ip$, provided that p^{-1} is approximately b. Intuitively, the stationary bootstrap scheme samples blocks of random length $1/p$, so the two approaches are roughly the same if the expected number of observations in each resampled block is the same for both methods. To further substantiate the claim that $m^2_{N,b} \approx \hat{\sigma}^2_{N,p}$ if $p = 1/b$, note that Künsch's expansion for the bias of the moving blocks estimate of variance exactly coincides with (9). In fact, (10) shows that the moving blocks—and hence also the stationary bootstrap—variance estimates are both approximately equivalent to a lag window spectral estimate using Bartlett's kernel (see Priestley 1981 for details). But a perhaps more interesting way to view the two variance estimates is as follows. One can compute $m^2_{N,b}$ defined by (10) for each b and then average over a distribution of b values. In particular, compute $E(m^2_{N,B})$, where B (independently) has a geometric distribution with mean p_N^{-1}, yielding

$$\begin{aligned} E(m^2_{N,B}) &= \hat{C}_N(0) + 2\sum_{b=1}^{\infty}\sum_{i=1}^{b-1}\left(1 - \frac{i}{b}\right)(1 - p_N)^{b-1}p_N\hat{C}_N(i) \\ &= \hat{C}_N(0) + 2\sum_{i=1}^{\infty}\sum_{b=i+1}^{\infty}\left(1 - \frac{i}{b}\right)(1 - p_N)^{b-1}p_N\hat{C}_N(i) \\ &= \hat{C}_N(0) + 2\sum_{i=1}^{\infty}\tilde{b}_N(i)\hat{C}_n(i), \end{aligned}$$

where

$$\tilde{b}_N(i) = (1 - p_N)^i + ip_N(1 - p_N)^{-1} \times \left[\log(p_N) + \sum_{j=1}^{i}\binom{i}{j}(-1)^j p^j/j!\right].$$

Because $p_N\log(p_N) \to 0$ as $p_N \to 0$, $\tilde{b}_N(i) \approx b_N(i)$, where $b_N(i)$ is given in (7). Hence the stationary bootstrap estimate of variance may be viewed approximately as a weighted average over b of estimates of variance based on resampling blocks of fixed length b; this suggests that the choice of p in the stationary scheme is less crucial than the choice of b in the moving blocks scheme. Moreover, by an argument similar to Theorem 1, $\text{var}(\hat{\sigma}^2_{N,p_N} - m^2_{N,b_N}) \to 0$ if $b = b_N = 1/p_N$, and the conditions of Theorem 1 are satisfied. The same claim can be made if $m^2_{N,b}$ is replaced by the exact moving blocks estimate of variance.

Two Simulated Examples. To empirically substantiate these claims, some numerical examples were considered, based on simulation. First, observations $X_1, \ldots, X_{200}$ were generated according to the model $X_t = Z_t + Z_{t-1} + Z_{t-2} + Z_{t-3} + Z_{t-4}$, where the Z_t are iid standard normal. Because $\sum_{i=1}^{200} X_i \approx 5\sum_{i=1}^{200} Z_i$, the variance of $N^{1/2}\bar{X}_N$, with $N = 200$, is very nearly 25. Note that the autocovariances $EX_0X_k \geq 0$, for any "lag" k. In Figure 1, the moving blocks and stationary bootstrap estimates of variance of $N^{1/2}\bar{X}_N$ are plotted as functions of block size b and $1/p$. Notice that the stationary bootstrap estimate of variance is much less variable; that is, it is less sensitive to the choice of p than the moving blocks bootstrap is to the choice of b.

Next, 200 observations from the model $X_t = Z_t - Z_{t-1} + Z_{t-2} - Z_{t-3} + Z_{t-4}$ were generated, where again the Z_t are

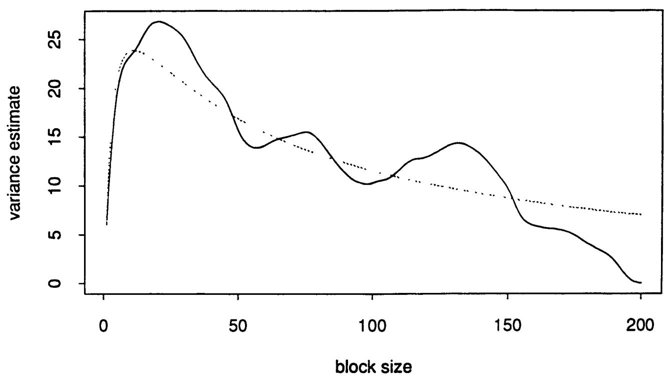

Figure 1. The Moving Blocks (Solid Line) and Stationary (Dotted Line) Bootstrap Estimates.

iid standard normal. In this case the autocovariances EX_0X_k, for $k = 1, 2, \ldots$, alternate in sign until they become 0 for lags k greater than 4. In Figure 2, the moving blocks and stationary bootstrap estimates of variance of $N^{1/2}\bar{X}_N$ are again plotted as functions of block size b and $1/p$. As before, it is observed that the stationary bootstrap estimate of variance is much less sensitive to the choice of p than the moving blocks is to the choice of b. In this second model, the true (standardized) variance of the sample mean is near 1, and the stationary bootstrap estimate is nearer to 1 for a wide range of p values; this behavior has been observed quite generally in other examples. Note that both Figure 1 and Figure 2 confirm our previous claim that the stationary bootstrap estimate of variance may be viewed approximately as a weighted average over b of moving blocks bootstrap estimates of variance.

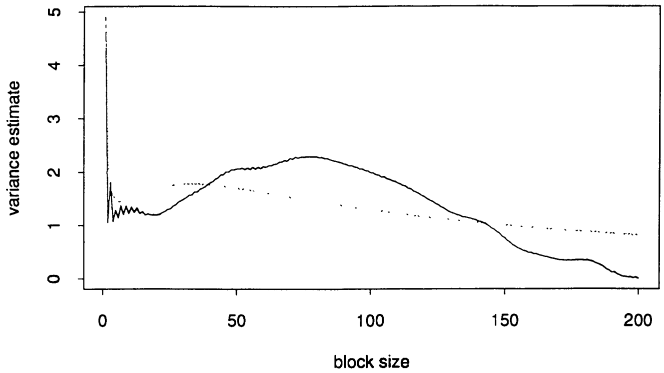

Figure 2. The Moving Blocks (Solid Line) and Stationary (Dotted Line) Bootstrap Estimates.

We now take up the problem of estimating the distribution of $N^{1/2}(\bar{X}_N - \mu)$, with the goal of constructing confidence intervals for μ. A strong mixing assumption on the original process will be in force. That is, it is assumed that data $X_1, \ldots, X_N$ are observed from an infinite sequence $\{X_n, n \in Z\}$. Let $\alpha_X(k) = \sup_{A,B}|P(AB) - P(A)P(B)|$, where A and B vary over events in the σ fields generated by $\{X_n, n \le 0\}$ and $\{X_n, n \ge k\}$.

The bootstrap approximation to the sampling distribution of $N^{1/2}(\bar{X}_N - \mu)$ is the distribution of $N^{1/2}(\bar{X}^*_N - \bar{X}_N)$, conditional on $X_1, \ldots, X_N$.

Theorem 2. Let $X_1, X_2, \ldots$ be a strictly stationary process with covariance function $R(\cdot)$ satisfying $R(0) + \sum_r |rR(r)| < \infty$. Assume (8) in Theorem 1. Assume, for some $d > 0$, that $E|X_i|^{d+2} < \infty$ and $\sum_k [\alpha_X(k)]^{d/(2+d)} < \infty$. Then, σ^2_∞ given in (3) is finite. Moreover, if $\sigma_\infty > 0$, then

$$\sup_x |P\{N^{1/2}(\bar{X}_N - \mu) \le x\} - \Phi(x/\sigma_\infty)| \to 0. \quad (11)$$

where $\Phi(\cdot)$ is the standard normal distribution function. Assume that $p_N \to 0$ and $Np_N \to \infty$. Then the bootstrap distribution is close to the true sampling distribution in the sense

$$\sup_x |P\{N^{1/2}(\bar{X}^*_N - \bar{X}_N) \le x | X_1, \ldots, X_N\} - P\{N^{1/2}(\bar{X}_N - \mu) \le x\}| \to 0 \quad (12)$$

in probability.

Remark 4. In Theorems 1 and 2, the condition (8) is implied by $E|X_i|^{6+\varepsilon} < \infty$ and $\sum_k k^2[\alpha(k)]^{[\varepsilon/(6+\varepsilon)]} < \infty$. To appreciate why, see (A.1) of Künsch (1989). Hence the conditions for Theorem 2 may be expressed solely in terms of a mixing condition and moment condition, without referring to cumulants. In summary, assume for some $\varepsilon > 0$ that $E|X_i|^{6+\varepsilon} < \infty$. Then the mixing conditions are implied by the single mixing condition $\alpha_X(k) = O(k^{-r})$ for some $r > 3(6 + \varepsilon)/\varepsilon$. This condition also implies $\sum_r |rR(r)| < \infty$.

The immediate application of Theorem 2 lies in the construction of confidence intervals for μ. For example, let $\hat{q}_N(1 - \alpha)$ be obtained from the bootstrap distribution by

$$P\{\bar{X}^*_N - \bar{X}_N \le \hat{q}_N(1 - \alpha)\} = 1 - \alpha.$$

Due to possible discreteness or uniqueness problems, $\hat{q}_N(1 - \alpha)$ should be defined to be the $1 - \alpha$ quantile of the (conditional) distribution of $\bar{X}^*_N - \bar{X}_N$; in general, let the $1 - \alpha$ quantile of an arbitrary distribution G be $\inf\{q : G(q) \ge 1 - \alpha\}$. Then it immediately follows that the bootstrap interval $[\bar{X}_N - \hat{q}_N(1 - \alpha/2), \bar{X}_N - \hat{q}_N(\alpha/2)]$ has asymptotic coverage $1 - \alpha$. Indeed, the theorem implies $\hat{q}_N(1 - \alpha) \to \sigma_\infty\Phi^{-1}(1 - \alpha)$ in probability.

Other bootstrap confidence intervals similarly may be shown to be asymptotically valid in the sense of having the correct asymptotic coverage; for example, a simple percentile method or the bootstrap t.

In practice, it is inevitable that a data-based choice for p would be made. For example, as previously mentioned, if p is chosen to minimize the mean squared error of $\hat{\sigma}^2_{N,p}$, then p should satisfy $N^3p_N \to C$. The constant C will depend on the spectral density and can be estimated consistently, say by some sequence $\hat{C}_N$. One could then choose $\hat{p}_N = N^{-1/3}\hat{C}_N$. In fact, with some additional effort, Theorem 2 can be generalized to consider a data-based choice for p. Subsequent work will focus on a proper choice of p. At this stage, it is clear that as long as p satisfies $p \to 0$ and $Np \to \infty$, the choice of p will not enter into first-order properties, such as coverage error, of the stationary bootstrap procedure. Getting the right rate for p to tend to 0 will undoubtedly enter into second-order properties, but getting "optimal" constants correct will be a third-order consideration. Such an investigation, though of vital importance, is beyond the scope of the present work. A step toward understanding second-order properties was presented by Lahiri (1992) in the case of moving blocks bootstrap.

4. EXTENSIONS

In this section we extend the results in Section 3 to more general parameters of interest. A basic theme is that results about the sample mean readily imply results for much more complicated statistics.

4.1 Multivariate Mean

Suppose that the X_i take values in $\mathbf{R}^d$, with jth component denoted by $X_{i,j}$. Interest focuses on the mean vector, $\mu = E(X_i)$, having jth component $\mu_j = E(X_{i,j})$. The definition of $\alpha_X(\cdot)$ readily applies to the multivariate case. As before, the stationary resampling algorithm is the same, yielding a pseudo–multivariate time series $X_1^*, \ldots, X_N^*$ with mean vector $\bar{X}_N^*$.

Theorem 3. Suppose, for some $\epsilon > 0$, that $E|X_{i,j}|^{6+\epsilon} < \infty$. Assume that $\alpha_X(k) = O(k^{-r})$ for some $r > 3(6+\varepsilon)/\varepsilon$. Then $N^{1/2}(\bar{X}_N - \mu)$ tends in distribution to the multivariate Gaussian distribution with mean 0 and covariance matrix $\mathbf{\Sigma} = (\sigma_{i,j})$, where

$$\sigma_{i,j} = \operatorname{cov}(X_{1,i}, X_{1,j}) + 2 \sum_{k=1}^{\infty} \operatorname{cov}(X_{1,i}, X_{1+k,j}).$$

Then if $p_N \to 0$ and $Np_N \to \infty$,

$$\sup_s | P^*\{ \|\bar{X}_N^* - \bar{X}_N\| \le s\} - P\{\|\bar{X}_N - \mu\| \le s\}| \to 0 \quad (13)$$

in probability, where $\|\cdot\|$ is any norm on $\mathbf{R}^d$ and P^* refers to a probability conditional on the original series.

The immediate application of the theorem is the construction of joint confidence regions for $\mu = (\mu_1, \ldots, \mu_d)$. Various choices for the norm yield different-shaped regions. Notice how easily the bootstrap handles the problem of constructing simultaneous confidence regions. An asymptotic approach would involve finding the distribution of the norm of a multivariate Gaussian random variable having a complicated (unknown) covariance structure. The resampling approach avoids such a calculation and handles all norms with equal facility.

4.2 Smooth Function of Means

Again, suppose that the X_i take values in $\mathbf{R}^d$. Suppose that $\theta = (\theta_1, \ldots, \theta_p)$, where $\theta_j = E[h_j(X_i)]$. Interest focuses on θ or some function f of θ. Let $\hat{\theta}_N = (\hat{\theta}_{N,1}, \ldots, \hat{\theta}_{N,j})$, where $\hat{\theta}_{N,j} = \sum_{i=1}^N h_j(X_i)/N$. Assume moment conditions of the h_j and mixing conditions on the X_i. Then, by the multivariate case, the bootstrap approximation to the distribution of $N^{1/2}(\hat{\theta}_N - \theta)$ is appropriately close in the sense

$$d(P\{N^{1/2}(\hat{\theta}_N - \theta) \le x\}, P^*\{N^{1/2}(\hat{\theta}_N^* - \hat{\theta}_N) \le x\}) \to 0 \quad (14)$$

in probability, where d is any metric metrizing weak convergence in $\mathbf{R}^p$. Moreover,

$$d(P\{N^{1/2}(\hat{\theta}_N - \theta) \le x\}, P\{Z \le x\}) \to 0, \quad (15)$$

where Z is multivariate Gaussian with mean 0 and covariance matrix $\mathbf{\Sigma}$ having (i, j) component

$$\operatorname{cov}(Z_i, Z_j) = \operatorname{cov}[h_i(X_1), h_j(X_1)] + 2 \sum_{k=1}^{\infty} \operatorname{cov}[h_i(X_1), h_j(X_{1+k})].$$

To see why, define $\mathbf{Y}_i$ to be the vector in $\mathbf{R}^p$ with jth component $h_j(X_i)$. Then the $\mathbf{Y}_i$ are weakly dependent if the original X_i are weakly dependent; in fact, $\alpha_Y(k) \le \alpha_X(k)$. Hence, with a moment assumption on the h_i, we are exactly back in the multivariate case. Now suppose that f is an appropriately smooth function from $\mathbf{R}^p$ to $\mathbf{R}^q$, and interest now focuses on the parameter $\mu = f(\theta)$. Assume that $f = (f_1, \ldots, f_q)$, where $f_i(y_1, \ldots, y_p)$ is a real-valued function from $\mathbf{R}^p$ having a nonzero differential at $(y_1, \ldots, y_p) = (\theta_1, \ldots, \theta_p)$. Let $\mathbf{D}$ be the $p \times q$ matrix with (i, j) entry $\partial f(y_1, \ldots, y_p)/\partial y_j$ evaluated at $(\theta_1, \ldots, \theta_p)$. Then the following is true.

Theorem 4. Suppose that f satisfies the aforementioned smoothness assumptions. Assume that for some $\varepsilon > 0$, $E[h_j(X_1)]^{6+\varepsilon} < \infty$, and that for some $r > 3(6+\varepsilon)/\varepsilon$, $\alpha_X(k) = O(k^{-r})$. Then, if $p_N \to 0$ and $Np_N \to \infty$, (14) and (15) hold. Moreover,

$$d(P\{N^{1/2}[f(\hat{\theta}_N) - f(\theta)] \le x\}, P^*\{N^{1/2}[f(\hat{\theta}_N^*) - f(\hat{\theta}_N)] \le x\}) \to 0$$

in probability and

$$\sup_s | P\{ \|f(\hat{\theta}_N) - f(\theta)\| \le s\} - P^*\{ \|f(\hat{\theta}_N^*) - f(\hat{\theta}_N)\| \le s\}| \to 0$$

in probability.

As an immediate application, consider the problem of constructing uniform confidence bands for $(R(1), \ldots, R(q))$, where $R(i) = \operatorname{cov}(X_1, X_{1+i})$. (To apply the previous theorem, let $W_i = (X_i, \ldots, X_{i+q})$, for $1 \le i \le N' = N - q$.) Although even asymptotic distribution theory for even Gaussian data seems formidable, the stationary bootstrap resampling approach handles the problem easily. The only caveat is to note that q is fixed as $N \to \infty$.

4.3 Differentiable Functionals

For simplicity, assume that the X_i are real-valued with common continuous distribution function F. Suppose that the parameter of interest μ is some functional T of F. A sensible estimate of F is $T(\hat{F}_N)$, where $\hat{F}_N$ is the empirical distribution of $X_1, \ldots, X_N$. Assume that T is Frechet differentiable; that is, suppose that

$$T(G) = T(F) + \int h_F \, d(G - F) + o(\|G - F\|),$$

for some (influence) function h_F, centered so that $\int h_F \, dF = 0$. For concreteness, suppose that $\|\cdot\|$ is the supremum norm, but this can be generalized. Then

$$N^{1/2}[T(\hat{F}_N) - T(F)] = N^{-1/2} \sum_{i=1}^{N} h_F(X_i) + o(N^{1/2}\|\hat{F}_N - F\|). \quad (16)$$

If for some $d \ge 0$, $E[h_F(X_1)]^{2+d} < \infty$ and $\sum_k [\alpha_X(k)]^{d/(2+d)}$, then $N^{-1/2} \sum_i h_F(X_i)$ is asymptotically normal with mean 0 and variance

$$E[h_F^2(X_i)] + 2 \sum_{k=1}^{\infty} \operatorname{cov}[h_F(X_1), h_F(X_{1+k})]. \quad (17)$$

To handle the remainder term in (16), Deo (1973) has shown that if $\sum_k k^2[\sigma_X(k)]^{1/2-\tau} < \infty$ for some $0 < \tau < \frac{1}{2}$, then $N^{1/2}[\hat{F}_N(\cdot) - F(\cdot)]$, regarded as a random element of the space of cadlag functions endowed with the supremum norm, converges weakly to $Z(\cdot)$, where $Z(\cdot)$ is a Gaussian process having continuous paths, mean 0, and

$$\operatorname{cov}[Z(t), Z(s)] = E[g_s(X_1)g_t(X_1)] + \sum_{k=1}^{\infty} E[g_s(X_1)g_t(X_{1+k})] + \sum_{k=1}^{\infty} E[g_s(X_{1+k})g_t(X_1)],$$

where $g_t(x) = I_{[0,t]}(x) - F(t)$. Hence Deo's result implies that $N^{1/2}[T(\hat{F}_N) - T(F)]$ is asymptotically normal with mean 0 and variance given by (17).

The bootstrap approximation to the distribution of $N^{1/2}[T(\hat{F}_N) - T(F)]$ is the distribution, conditional on $X_1, \ldots, X_N$, of $N^{1/2}[T(\hat{F}_N^*) - T(\hat{F}_N)]$, where $\hat{F}_N^*$ is the empirical distribution of $X_1^*, \ldots, X_N^*$ obtained by the stationary resampling procedure. If the error terms in the differential approximation of $T(\hat{F}_N^*)$ are negligible, then it is clear that the bootstrap will behave correctly, because Theorem 2 is essentially applicable. The key to justifying negligibility of error terms is to show $\rho(N^{1/2}[\hat{F}_N^*(\cdot) - \hat{F}_N(\cdot)], Z(\cdot)) \to 0$ in probability, where ρ is any metric metrizing weak convergence in the assumed function space. By Theorem 3, it is clear that the finite-dimensional distributions of $N^{1/2}[\hat{F}_N^*(\cdot) - F(\cdot)]$ will appropriately converge to those of $Z(\cdot)$. The only technical difficulty is showing tightness of the bootstrap empirical process. In fact, by an argument similar to Deo's, tightness can be shown if $Np_N^2 \to \infty$. The technical details will appear elsewhere.

In fact, the foregoing sketchy argument actually applies if T is only assumed compactly differentiable. For example, asymptotic validity for quantile functionals follows.

4.4 Linear Statistics Defined on Subseries

Assume that $X_i \in \mathbf{R}^d$. In this section we discuss how the stationary bootstrap may be applied to yield valid inferences for a parameter $\mu \in \mathbf{R}^D$ that may depend on the whole infinite-dimensional distribution of the process.

Consider the subseries $S_{i,M,L} = (X_{(i-1)L+1}, \ldots, X_{(i-1)L+M})$. These subseries can be obtained from the $\{X_i\}$ by a "window" of width M "moving" at lag L. Suppose that $T_{i,M,L}$ is an estimate of μ based on the subseries $S_{i,M,L}$, so $T_{i,M,L} = \phi_M(S_{i,M,L})$, for some function ϕ_M from $\mathbf{R}^{dM}$ to $\mathbf{R}^D$. Let $\bar{T}_N = \sum_{i=1}^{Q} T_{i,M,L}/Q$, where $Q = \{[(N-M)/L]\} + 1$; here $[\cdot]$ is the greatest integer function. To apply resampling to approximate the distribution of $\bar{T}_N$, just regard $(T_{1,M,L}, \ldots, T_{Q,M,L})$ as a time series in its own right. Note that M, L, and Q may depend on N. Weak dependence properties of the original series readily translate into weak dependence properties of this new series. Hence we are essentially back into the sample mean setting. A technical complication is that we are dealing with a triangular array of variables, so that Theorem 2 must be generalized. By taking this viewpoint, one can establish consistency and weak convergence properties of the stationary bootstrap. Indeed, this approach has been applied fruitfully in the moving blocks resampling scheme by Politis and Romano (1992a, 1992b). To appreciate the applicability of this approach, consider the problem of estimating the spectral density $f(\omega)$. Suppose that $T_{i,M,L}(\omega)$ is the periodogram evaluated at ω based on data $S_{i,M,L}$. Then, in fact, $\bar{T}_N(\omega)$ is approximately equal to Bartlett's kernel estimate of $f(\omega)$. Other kernel estimators can be (approximately) obtained by appropriate tapering of the individual periodogram estimates. A great advantage of the resampling approach is that it easily yields simultaneous confidence regions for the spectral density over some finite grid of ω values. Other examples falling in this framework are the spectral measure and cross-spectrum, where asymptotic approximations to sampling distributions are particularly intractable.

4.5 Future Work

Subsequent work will focus on three important problems. First, establish theoretical results to construct uniform confidence bands for the spectral measure. The discussion in Section 4.4 will readily allow one to construct confidence bands for the spectral measure over a finite grid of ω values, but this is theoretically unsatisfying. By constructing uniform confidence bands over the whole continuous range of ω, a basis for goodness-of-fit procedures can be established. Second, higher-order asymptotics are necessary, especially to compare procedures, just as in the iid case. Finally, the practical implementation, especially the choice of p, and the finite sample validity based on simulations will be addressed.

5. FURTHER NUMERICAL EXAMPLES

In Figure 3, the well-known Canadian lynx data are displayed, representing the number of lynx trappings in the Mackenzie River in the years 1821 to 1934; a histogram of the data reveals it is skewed and so not normal. Léger, Politis, and Romano (1992) analyzed the Canadian lynx data, together with the artificial series Y_t, $t = 1, \ldots, 200$, where $Y_t = X_t|X_t| + c$ and the X_t series follows the ARMA model

$$X_t - 1.352X_{t-1} + 1.338X_{t-2} - .662X_{t-3} + .240X_{t-4} = Z_t - .2Z_{t-1} + .04Z_{t-2},$$

with the Z_t's being independent normal $N(0, 1)$ random variables (and $c = 0$). A realization of the Y_t series is exhibited

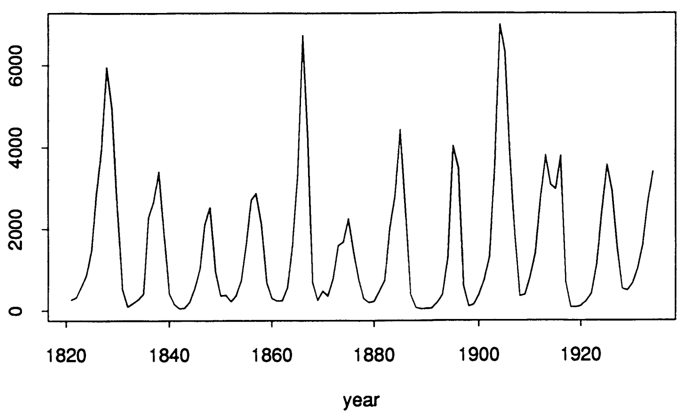

Figure 3. The Time Series of Annual Number of Lynx Trappings.

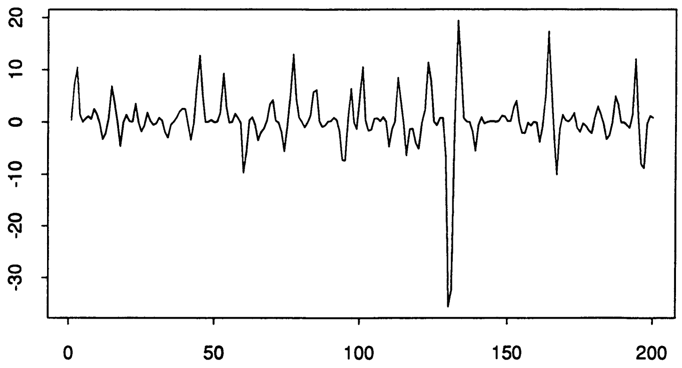

Figure 4. The Artificial Time Series Y.

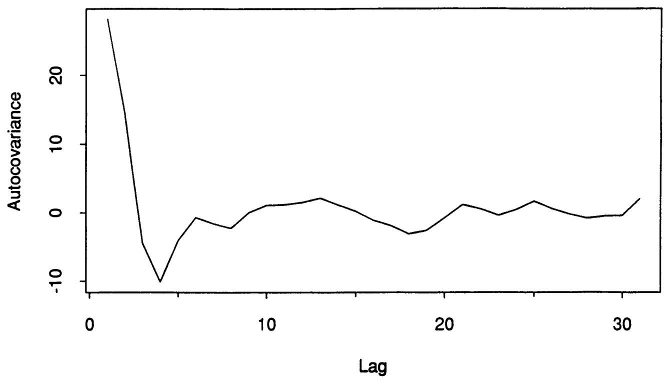

Figure 6. Sample Autocovariance of Y series.

in Figure 4; a histogram reveals this data is also not normal due to heavy tails.

In Léger et al. (1992) constructed confidence intervals for the mean of the Lynx series were constructed using the moving blocks technique. They also discussed the choice of b (and hence $p = 1/b$ for the stationary bootstrap). The stationary bootstrap "hybrid" (i.e., based on the approximation $P^*\{\sqrt{n}(T^* - \hat{T}) \leq x\} \simeq P\{\sqrt{n}(\hat{T} - \mu) \leq x\}$) 95% confidence interval for the mean μ of the Lynx data was [1,233.816, 1,832.719], based on 500 replications with $p = .05$. (The sample mean of the Lynx data is 1,538.018.) This is remarkably close to the Moving Blocks 95% confidence interval of [1,233.37, 1,826.07] presented by Léger et al. (1992), which was again based on 500 replications with $b = 25$. Note that in the stationary bootstrap simulation, p was chosen such that $1/p \simeq b$, where the choice of $b \simeq 25$ was explained by Léger et al. (1992).

But we might also consider the median m of the Lynx data as the parameter of interest. The obvious estimator is the sample median, which was equal to 771. But we need to attach a standard error or confidence interval to this estimate. The stationary bootstrap (i.e., "hybrid") 95% confidence interval for the median m of the Lynx data was [242.5, 957], based on 1,000 replications with $p = .05$.

Turning to the artificial Y_t series, it was mentioned that the distribution of Y_t, for some fixed t, is non-Gaussian. Indeed, it is a two-sided χ^2 distribution with 1 degree of freedom, centered and symmetric around the constant c. By analogy to the iid case, it is expected that the median and trimmed means would be more efficient than the sample mean for estimation of the location parameter c, because the two-sided χ^2 distribution with 1 degree of freedom can be thought of as being "close" to the double exponential distribution, (as close as the χ^2 distribution with 1 degree of freedom is to the χ^2 distribution with 2 degrees of freedom). For the simulation, the constant c was set to 0, and six different estimators of c were considered: the sample mean, the median, and the α-trimmed means (i.e., the mean of the remaining observations after throwing away the $[n\alpha]$ largest and the $[n\alpha]$ smallest ones), with $\alpha = .1, .2, .3$, and $.4$.

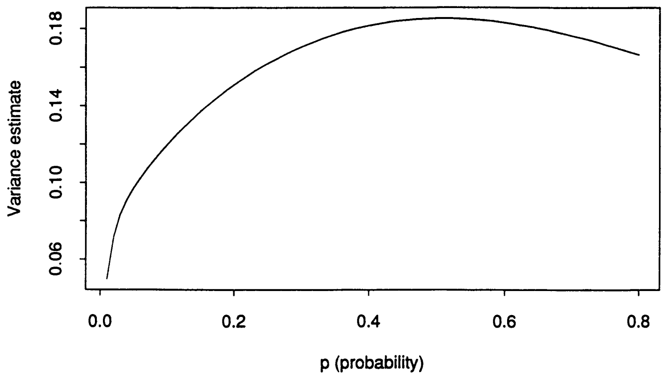

Figure 5. The Stationary Bootstrap Variance Estimate.

First, the problem of choosing p for the stationary bootstrap. To do this, look at the sample mean case, for which a simple expression of the variance exists:

$$\operatorname{var}\left(\frac{1}{200}\sum_{i=1}^{200} Y_i\right) = \frac{1}{200}\left(\operatorname{var}(Y_1) + 2\sum_{i=1}^{200}\left(1 - \frac{i}{200}\right)\operatorname{cov}(Y_1, Y_{1+i})\right).$$

The stationary bootstrap estimates of the variance of the sample mean for different choices of $p \in (0, .8)$ are pictured in Figure 5, and the sample autocovariance sequence of the Y_t series is pictured in Figure 6. It is seen that the autocovariances for lags greater than 6 are not significantly different from 0. This would lead to an empirically acceptable choice of b for the moving block method of the order of 10 (see Léger, Politis and Romano 1992). By the approximate correspondence of the moving blocks method and the stationary bootstrap with $p = 1/b$, the choice of $p = .1$ is suggested.

Having decided to use $p = .1$, let us proceed in comparing the six proposed estimators of c. Based on 500 stationary bootstrap replications, Table 1 reports the stationary bootstrap estimate of variance of the corresponding estimator,

Table 1. Trimmed Mean Confidence Intervals

α	$\hat{\sigma}^2_{\text{stat.bootstrap}}$	95% confidence interval
0	.1034	[−.282, .984]
.1	.0386	[−.089, .662]
.2	.0159	[−.092, .413]
.3	.0094	[−.030, .345]
.4	.0050	[−.080, .202]
.5	.0028	[−.082, .105]

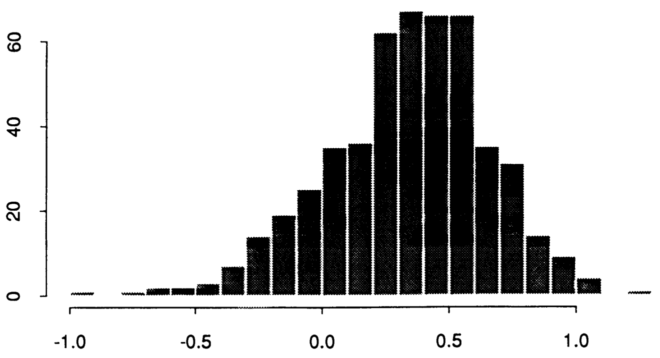

Figure 7. Bootstrap Distribution of the Y Series Sample Mean.

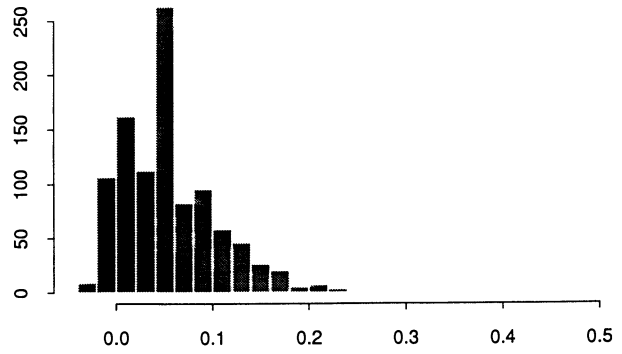

Figure 10. Bootstrap Distribution of the Y Series Sample Median.

as well as the 95% bootstrap confidence interval for the parameter c. For compact notation, the mean and the median were denoted as trimmed means, with α equal to 0 and .5. It is obvious from the table that the intuition suggesting the median as most efficient seems to be correct. Indeed, the median has the smallest (estimated by the stationary bootstrap) variance, and yields the shortest confidence interval for c; recall that c was taken to equal 0 in this simulation. According to this reasoning, the median should be preferred.

In Figures 7 through 10, the stationary bootstrap histograms of the distribution of the sample mean, the $\alpha = .1$ and .3 trimmed means, and the sample median of the Y_t series are pictured, based on 1000 bootstrap replications and $p = .1$. The bootstrap distribution of the sample median is clearly the least disperse. Based on the asymptotic theory justifying the bootstrap approximations to each of the trimmed means, the bootstrap can further be shown to be a viable method of choosing among competing estimators in an adaptive manner (see Léger and Romano 1990a,b).

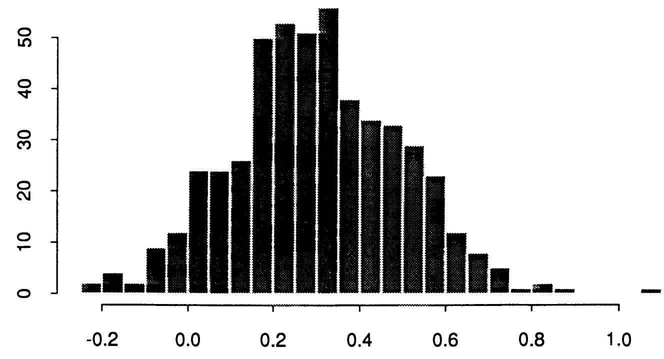

Figure 8. Bootstrap Distribution of the Y Series .1 Trimmed Mean.

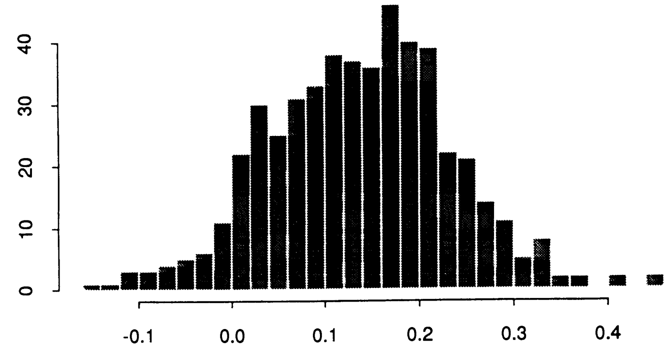

Figure 9. Bootstrap Distribution of the Y Series .3 Trimmed Mean.

APPENDIX: PROOFS

Proof of Lemma 1.

In the proof, all expectations and covariances are conditional on $X_1, \ldots, X_N$. Recall L_1 in the construction of the stationary resampling scheme. Then

$$E(X_1^* X_{1+i}^*) = E(X_1^* X_{1+i}^* \mid L_1 > i)P(L_1 > i) + E(X_1^* X_{1+i}^* \mid L_1 \le i)P(L_1 \le i)$$
$$= N^{-1} \sum_{j=1}^{N} X_j X_{j+i}(1-p)^i + \bar{X}_N^2[1-(1-p)^i].$$

Hence $\operatorname{cov}(X_1^*, X_{1+i}^*) = \hat{C}_N(i)(1-p)^i$. So, by (2) applied to the X_t^* series,

$$\hat{\sigma}_{N,p}^2 = \hat{C}_N(0) + 2\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)(1-p)^i\hat{C}_N(i),$$

yielding (5). To get (6), note that $\hat{R}_N(0) = \hat{C}_N(0)$, and if $1 \le i \le N-1$, then $\hat{C}_N(i) = \hat{R}_N(i) + \hat{R}_N(N-i)$. Therefore, by (5),

$$\hat{\sigma}_{N,p}^2 = \hat{R}_N(0) = 2\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)(1-p)^i\hat{R}_N(i) + 2\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)(1-p)^i\hat{R}_N(N-i).$$

Letting $j = N - i$ in the last sum yields the result.

Proof of Theorem 1.

For purposes of the proof, we may assume that $E(X_i) = 0$. Let

$$s_N^2 = s_{N,p_N}^2 = \hat{R}_{N,0}(0) + 2\sum_{i=1}^{N-1} b_N(i)\hat{R}_{N,0}(i), \qquad \text{(A.1)}$$

where $\hat{R}_{N,0}(i) = \sum_{j=1}^{N-i} X_j X_{j+i}/N$. By (5),

$$\hat{\sigma}_{N,P}^2 = s_{N,P}^2 - \bar{X}_N^2 - 2\bar{X}_N^2 \sum_{i=1}^{N-1} b_N(i).$$

Under the assumptions, $\bar{X}_N = O_P(N^{-1/2})$. Also, $\sum_{i=1}^{N-1} b_N(i) \le 2/p_N$, which implies $\bar{X}_N^2 \sum_{i+1}^{N-1} b_N(i) = o_P(1)$. Hence it suffices to show the estimator s_N^2 in (A.1) satisfies $s_N^2 \to \sigma_\infty^2$ in probability. To accomplish this, we show the bias and variance of s_N^2 tend to 0. By (7) and $E[\hat{R}_{N,0}(i)] = (N - i/N)R(i)$, it follows that

$$E(s_N^2) = R(0) + 2\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)^2(1-p_N)^i R(i) + 2\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)\frac{i}{N}(1-p_N)^{N-i}R(i).$$

The absolute value of the last term is bounded above by 2 $\sum_{i=1}^{\infty}|iR(i)|/N = O(N^{-1})$. To handle the first summation, use the approximation $(1-p_N)^i \approx 1 - ip_N$ to get this term is

$$2\sum_{i=1}^{N-1} R(i) - 2p_N\sum_{i=1}^{N-1} iR(i) + o(p_N). \tag{A.2}$$

Hence $E(s_N^2) = \sigma_\infty^2 + O(p_N)$. To calculate the variance of s_N^2, by the result (5.3.21) of Priestley (1981) originally due to Bartlett (1946), $\text{cov}[\hat{R}_{N,0}(i), \hat{R}_{N,0}(j)] \le S/N$, where $S = 2R(0)\sum_{m=-\infty}^{\infty}|R(m)| + K$. Now,

$$\text{var}(s_{N,p}^2) = \sum_{i=-(N-1)}^{N-1}\sum_{-(N-1)}^{N-1} b_N(i)b_N(j)\text{cov}[\hat{R}_{N,0}(i), \hat{R}_{N,0}(j)]$$

$$\le \frac{S}{N}\sum_{i=-(N-1)}^{N-1}\sum_{j=-(N-1)}^{N-1} b_N(i)b_N(j) \le \frac{S}{N}\cdot\frac{1-p_N+p_N^2}{p_N(1-p_N)} \to 0$$

if $Np_N \to \infty$ and $p_N \to 0$. Thus the result is proved.

Proof of Theorem 2

Without loss of generality, assume $\mu = 0$. The result (11) follows immediately from corollary 5.1 of Hall and Heyde (1980). To prove (12), for now assume the following three convergences hold for the sequence $X_1, X_2, \ldots$:

(C1) $N\bar{X}_N^2/(Np_N) \to 0$.
(C2) $\hat{C}_N(0) + 2\sum_{i=1}^{\infty}(1-p_N)^i\hat{C}_n(i) \to \sigma_\infty^2$.
(C3) $\frac{p_N}{N^{1+\delta/2}}\sum_{r=1}^{\infty}\sum_{i=1}^{N}\left|S_{i,r} - \frac{\bar{X}_N}{p_N}\right|^{2+\delta}(1-p_N)^{r-1}p_N \to 0$.

In (C3), $S_{i,b}$ is defined to be the sum of observations in $B_{i,b}$ defined in (1).

Claim. The distribution of $N^{1/2}(\bar{X}_N^* - \bar{X}_N)$, conditional on $X_1, \ldots, X_N$, tends weakly to the normal distribution with mean 0 and variance σ_∞^2, for every sequence $X_1, X_2, \ldots$ satisfying (C1), (C2), and (C3).

The proof of this claim will be given in five steps. In the proof, all calculations referring to this bootstrap distribution will be assumed conditional on $X_1, \ldots, X_N$. Set $E_{N,m} = (S_{I_1,L_1} + \cdots + S_{I_m,L_m})/N$, where the $I_1, I_2, \ldots$ are iid uniform on $\{1, \ldots, N\}$ and the $L_1, L_2, \ldots$ are iid geometric with mean $1/p_N$. Let M be the smallest integer m such that $L_1 + \cdots + L_m \ge N$. Also, let $J_1 = L_1 + \ldots L_{M-1}$ and $J = L_M + J_1$. Then $E_{N,M} - \bar{X}_N^*$ is just N^{-1} times the sum of the observations in B_{I_M,L_M}, after deleting the first $N - J_1$ of them. Let R_1 be the exact number of observations required from block B_{I_M,L_M} so that N observations from the M blocks have been sampled; that is, $R_1 = N - J_1$. Also, let $R = L_M - R_1$. Note that R, conditional on (R_1, J_1), has a geometric distribution with mean $1/p_N$. This follows from the "memoryless" property of the geometric distribution. Hence $E_{N,M} - \bar{X}_N^*$ is equal in distribution to $N^{-1}S_{I,R}$, where I is uniform on $\{1, \ldots N\}$.

Step 1. Show that $N^{1/2}(E_{N,M} - \bar{X}_N^*) \to 0$ in (conditional) probability. By the foregoing observation, it is enough to show that the mean and variance of $N^{-1/2}S_{I,R}$ tends to 0. But $E[S_{I,R}|R] = R\bar{X}_N$, so that $N^{-1/2}E(S_{I,R}) = N^{1/2}\bar{X}_N/(Np_N) \to 0$. Now,

$$N^{-1}\text{var}(S_{I,R}) = N^{-1}E[\text{var}(S_{I,R}|R] + N^{-1}\text{var}[E(S_{I,R}|R)].$$

But $\text{var}[E(S_{I,R}|R)] = \text{var}(R\bar{X}_N) = \bar{X}_N^2(1-p_N)/p_N^2$. Thus, by (C1) and $Np_N \to \infty$, $N^{-1}E[\text{var}(S_{I,R})|R] \to 0$, yielding $N^{-1}\text{var}(S_{I,R}) \to 0$ as well.

Step 2. Show that for any fixed sequence $m = m_N$ satisfying $Np_N/m_N \to 1$, the distribution of

$$N^{1/2}\left(E_{N,m_N} - \frac{m_N\bar{X}_N}{Np_N}\right) \tag{A.3}$$

tends to the normal distribution with mean 0 and variance σ_∞^2. First, note that $E(S_{I_1,L_1}) = \bar{X}_N p_N$. For $1 \le i \le m_N$, let $Y_{N,i} = m_N^{1/2}S_{I_i,L_i}/N^{1/2}$. Then (A.3) is $m_N^{1/2}[\bar{Y}_{m_N} - E(\bar{Y}_{m_N})]$, and $\bar{Y}_{m_N} = \sum_{i=1}^{m_N} Y_{N,i}/m_N$ is the average of iid variables. But, as in step 1, $\text{var}(Y_{N,i})$ is the same as the variance of m_N/N times the variance of $S_{I,R}$, where I is uniform on $\{1, \ldots, N\}$ and R is geometric with mean p_N. Again, apply the relationship

$$\text{var}(S_{I,R}) = E[\text{var}(S_{I,R}|R)] + \text{var}[E(S_{I,R}|R)]. \tag{A.4}$$

The second term on the right side of (A.4) is $\text{var}(R\bar{X}_N) = \bar{X}_N^2(1-p_N)/p_N^2 \to 0$ by (C1). Also, $r^{-1}\text{var}(S_{I,R}|R=r)$ is in fact given by $m_{N,r}^2$ defined in (10). Thus

$$\frac{m_N}{N}\text{var}(S_{I,R}) = \frac{m_N}{N}E(Rm_{N,R}^2) + o(1) = \frac{m_N}{Np_N}\hat{C}_N(0) + \frac{2m_N}{Np_N}\sum_{i=1}^{\infty}(1-p_N)^i\hat{C}_N(i) + o(1).$$

By the assumption $Np_N/m_N \to 1$ and (C2), it follows that $\text{var}(Y_{N,i}) \to \sigma_\infty^2$. To complete step 2, by Katz's (1963) Berry–Esseen bound, it suffices to show that

$$m_N^{-\delta/2}E|Y_{N,i} - E(Y_{N,i})|^{2+\delta} \to 0 \tag{A.5}$$

as $m_N \to \infty$. But the left side of (A.5) is (by conditioning on R) equal to

$$\frac{m_N}{N^{1+\delta/2}}E|S_{I,R}|^{2+\delta} = \frac{m_N}{N^{2+\delta/2}}\sum_r\sum_i\left|S_{i,r} - \frac{\bar{X}_N}{p_N}\right|^{2+\delta}(1-p_N)^{r-1}p_N,$$

which tends to 0 by (C3).

Step 3. The distribution of $N^{1/2}(E_{N,m_N} - \bar{X}_N)$ tends to normal with mean 0 and variance σ_∞^2. This follows by step 2 and (C1).

Step 4. The distribution of $N^{1/2}(E_{N,M} - \bar{X}_N)$ tends to normal with mean 0 and variance σ_∞^2. To see why, if $\tilde{M}$ is any random variable (sequence) satisfying $\tilde{M}/Np_N \to 1$ in probability, then $N^{1/2}(E_{N,\tilde{M}} - \bar{X}_N)$ tends to normal with mean 0 and variance σ_∞^2. This essentially follows by an extension of Theorem 7.3.2 (to a triangular array setting) of Chung (1974). In our case, $M = Np_N + O_P(N^{1/2}p_N^{1/2})$.

Step 5. Combine steps 1 and 4 to prove the claim.

Now to deduce (12), by a subsequence argument it suffices to show that the convergence (C1), (C2), and (C3) hold in probability for the original sequence $X_1, X_2, \ldots$. First, (C1) holds in probability because $N^{1/2}\bar{X}_N$ is order 1 in probability and $Np_N \to \infty$. Second, the convergence (C2) holds in probability by an argument very similar to Theorem 1. Finally, to show that (C3) holds in probability, write the term in question as

$$\frac{p_N}{N^{\delta/2}}E\left|S_{I,R} - \frac{\bar{X}_N}{p_N}\right|^{2+\delta}. \tag{A.6}$$

It suffices to show that (A.6) raised to the power $(2+\delta)^{-1}$ tends to 0 in probability, which by Minkowski's inequality is bounded above by

$$\left(\frac{p_N}{N^{\delta/2}}\right)^{1/(2+\delta)}[E|S_{I,R} - R\bar{X}_N|^{2+\delta}]^{1/(2+\delta)} + \left(\frac{p_N}{N^{\delta/2}}\right)^{1/(2+\delta)}\bar{X}_N E[|R - p_N^{-1}|^{2+\delta}]^{1/(2+\delta)}. \tag{A.7}$$

The second term in (A.7) is of order $\bar{X}_N N^{1/2}[Np_N]^{-(1+\delta)/(2+\delta)}$, which tends to 0 in probability. It now suffices to show

$$\frac{p_N}{N^{\delta/2}} E|S_{I,R} - R\bar{X}_N|^{2+\delta} \to 0$$

in probability, or that its expectation tends to 0; that is,

$$\frac{p_N}{N^{1+\delta/2}} \sum_r \sum_i E[|S_{i,r} - r\bar{X}_N|^{2+\delta}](1-p_N)^{r-1} p_N \to 0. \quad \text{(A.8)}$$

To bound $E|S_{i,r} - r\bar{X}_N|^{2+\delta}$, note that if $1 \le i \le i + r - 1 \le N$, then Yokoyama's (1980) moment inequality applies, yielding $E|S_{i,r}|^{2+\delta} \le Kr^{1+(\delta/2)}$, where the constant K depends only on the mixing sequence $\{\alpha(k)\}$. Thus, by Minkowski's inequality and then Yokoyama's inequality, we have

$$\begin{aligned} E|S_{i,r} - r\bar{X}_N|^{2+\delta} &\le [K^{1/(2+\delta)} r^{[1+(\delta/2)][1/(2+\delta)]} \\ &\quad + (E|r\bar{X}_N|^{2+\delta})^{1/(2+\delta)}]^{2+\delta} \\ &\le \left[K^{1/(2+\delta)} r^{[1+(\delta/2)][1/(2+\delta)]} \right. \\ &\quad \left. + \frac{r}{N} KN^{[1+(\delta/2)][1/(2+\delta)]}\right]^{2+\delta} \\ &\le (2K)^{2+\delta} r^{(1+\delta)/2}. \end{aligned}$$

In the case $i + r - 1 > N$ but $r < N$, write $S_{i,r} = (X_i + \cdots + X_N) + (X_1 + \cdots + X_{i+r-1-N})$. Apply Minkowski's inequality and Yokoyama's inequality to get $E|S_{i,r}| \le 2^{2+\delta} Kr^{1+(\delta/2)}$. Then, arguing as earlier, we find $E|S_{i,r} - r\bar{X}_N|^{2+\delta} \le (3K)^{2+\delta} r^{1+(\delta/2)}$. In the general case, suppose that $r + N(j-1) + \tilde{r}$, where $1 \le \tilde{r} \le N$. Then $S_{i,r} = (j-1)N\bar{X}_N + S_{i,\tilde{r}}$. So

$$E|S_{i,r} - r\bar{X}_N|^{2+\delta} = E|S_{i,\tilde{r}} - \tilde{r}\bar{X}_N|^{2+\delta},$$

and the general bound $(3K)^{2+\delta} r^{1+(\delta/2)}$ applies. Hence (A.8) is bounded above by

$$\frac{p_N}{N^{\delta/2}} \sum_{r=1}^{\infty} (3K)^{2+\delta} r^{1+\delta/2} (1-p_N)^{r-1} p_N = O\left(\frac{p_N}{N^{\delta/2}} \cdot \frac{1}{p_N^{1+\delta/2}}\right) = o(1).$$

Proof of Theorem 3

The proof follows immediately by considering linear combinations of the components and applying Theorem 2, which is applicable by Remark 4. Then (13) follows by the continuous mapping theorem (because a norm is almost everywhere continuous with respect to a Gaussian measure).

Proof of Theorem 4

The proof follows as (14) and (15) are immediate from Theorem 3, and the smoothness assumptions on f imply that $N^{1/2}[f(\hat{\theta}_N) - f(\theta)]$ has a limiting multivariate Gaussian distribution with mean 0 and covariance matrix $\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'$; see theorem A of Serfling (1980, p. 122).

[Received April 1992. Revised April 1993.]

REFERENCES

Bartlett, M. S. (1946), "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series," *Journal of the Royal Statistical Society Supplement,* 8, 27–41.

Brillinger, D. (1981), *Time Series: Data Analysis and Theory,* San Francisco: Holden-Day.

Chung, K. (1974), *A Course in Probability Theory* (2nd ed.), New York: Academic Press.

Deo, C. (1973), "A Note on Empirical Processes of Strong-Mixing Sequences," *Annals of Probability,* 1, 870–875.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics,* 7, 1–26.

Hall, P., and Heyde, C. (1980), *Martingale Limit Theory and its Application,* New York: Academic Press.

Katz, M. (1963), "Note on the Berry–Esseen Theorem," *Annals of Mathematical Statistics,* 34, 1107–1108.

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics,* 17, 1217–1241.

Lahiri, S. (1992), "Edgeworth Correction by Moving Block Bootstrap for Stationary and Nonstationary Data," in *Exploring the Limits of Bootstrap,* eds. R. LePage and L. Billard, New York: John Wiley.

Léger, C., Politis, D., and Romano, J. (1992), "Bootstrap Technology and Applications," *Technometrics,* 34, 378–398.

Léger, C., and Romano, J. (1990a), "Bootstrap Choice of Tuning Parameters," *Annal of the Institute of Statistical Mathematics,* 4, 709–735.

——— (1990b), "Bootstrap Adaptive Estimation: The Trimmed Mean Example," *Canadian Journal of Statistics,* 18, 297–314.

Liu, R. Y., and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap,* eds. R. LePage and L. Billard, New York: John Wiley.

Politis, D., and Romano, J. (1992a), "A General Resampling Scheme for Triangular Arrays of α-Mixing Random Variables With Application to the Problem of Spectral Density Estimation," *The Annals of Statistics,* 20, 1985–2007.

——— (1992b), "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, eds. C. Page and R. LePage, New York: Springer-Verlag, pp. 98–103.

——— (1992c), "A Circular Block Resampling Procedure for Stationary Data," in *Exploring the Limits of Bootstrap,* eds. R. LePage and L. Billard, New York: John Wiley, pp. 263–270.

Politis, D., Romano, J., and Lai, T. (1992), "Bootstrap Confidence Bands for Spectra and Cross-Spectra," *IEEE Transactions on Signal Processing,* 40, 1206–1215.

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

Rosenblatt, M. (1984), "Asymptotic Normality, Strong Mixing, and Spectral Density Estimates," *Annals of Probability,* 12, 1167–1180.

——— (1985), *Stationary Sequences and Random Fields,* Boston: Birkhauser.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics,* New York: John Wiley.

Yokoyama, R. (1980), "Moment Bounds for Stationary Mixing Sequences," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete,* 52, 45–57.

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