APPLICATIONS OF RESAMPLING SCHEMES FOR STATIONARY TIME SERIES

by

Dimitris N. Politis
Purdue University

Technical Report # 91-13

Department of Statistics
Purdue University

March, 1991

△

APPLICATIONS OF RESAMPLING SCHEMES FOR STATIONARY TIME SERIES

by

Dimitris N. Politis
Purdue University

ABSTRACT

The bootstrap has been proven to be a powerful tool for the nonparametric estimation of standard errors and confidence limits in situations where the sample consists of independent observations. Recently, block-resampling schemes have been proposed in order to apply the bootstrap to time series problems. In this report, the different variants of resampling schemes for stationary time series are compared, and examples of their implementation are presented using real and simulated data sets.

1. Introduction: The bootstrap in time series problems

The problem of nonparametrically estimating the variance and/or the sampling distribution of statistics based on data $X_1, X_2, \ldots, X_N$ is considerably more difficult in the case the data are not independent. Let us first consider the simplest such case where $X_1, \ldots, X_N$ are observations from the (univariate) m-dependent stationary sequence $\{X_n, n \in \mathbf{Z}\}$, and the statistic of interest is the sample mean $\bar{X}_N = N^{-1}\sum_{i=1}^{N} X_i$. Recall that a sequence of random variables $\{X_n, n \in \mathbf{Z}\}$ is called (strictly) stationary if, for all $n \in \mathbf{N}$, the joint distribution of $(X_k, X_{k+1}, \ldots, X_{k+n})$ does not depend on k; a stationary sequence is called m-dependent if the set of random variables $\{X_n, n = -1, -2, \ldots\}$ is independent of $\{X_n, n = m, m+1, \ldots\}$. In this setting, independence can be thought of as 0-dependence.

It is easy to see that $E\bar{X}_N = \mu$, where $\mu = EX_1$, and that (due to stationarity)

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

In fact, taking into account the m-dependence, it is implied that

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

since $Cov(X_1, X_{1+i}) = 0$, for $i > m$. Hence, it is immediate that $\bar{X}_N \xrightarrow{P} \mu$, i.e. the sample mean is consistent. It is also easy to show that $\bar{X}_N$ is asymptotically normal, so to obtain confidence intervals for μ we would just need a consistent estimate of $\sigma_\infty^2 \equiv \lim_{N\to\infty}\sigma_N^2 = Var(X_1) + 2\sum_{i=1}^{m} Cov(X_1, X_{1+i})$.

Looking at the classical (i.i.d.) bootstrap estimate of variance of $\sqrt{N}\bar{X}_N$, it is apparent that it would asymptotically converge to $Var(X_1)$, and therefore it would be inconsistent for σ_∞^2. The reason of course is that the classical bootstrap resampling scheme is valid for *independent* data, and is insensitive to their 'time-order'. In other words, the classical bootstrap procedure utilizes the data only through the empirical distribution that they define. This empirical distribution is an approximation to the first-marginal distribution of the sequence $\{X_n\}$, and hence, based on it, we can not expect to estimate a parameter like σ_∞^2 which depends on the $(m+1)$th-marginal distribution of the m-dependent observations (and on the whole infinite-dimensional joint distribution for observations with more general dependence structure).

Recently Künsch(1989) and Liu and Singh(1988) have independently proposed a block-resampling scheme that takes care of this problem by working with empirical estimates of the b-dimensional marginal distribution of the sequence $\{X_n\}$, where b is an integer that is allowed to increase with the sample size N. This method (hereafter termed the 'moving blocks' method), can be described as follows:

- Define $\mathcal{B}_i$ to be the block of b consecutive observations starting from X_i, that is $\mathcal{B}_i = (X_i, \ldots, X_{i+b-1})$, where $i = 1, \ldots, q$ and $q = N - b + 1$. Sampling with replacement from the set $\{\mathcal{B}_1, \ldots, \mathcal{B}_q\}$, defines a (conditional on the original data) probability measure P^* which is used in the 'moving blocks' bootstrap procedure. If k is an integer such that $kb \sim N$, then letting $\xi_1, \ldots, \xi_k$ be drawn i.i.d. from P^*, it is seen that each ξ_i is a block of b observations $(\xi_{i,1}, \ldots, \xi_{i,b})$. If all $l = kb$ of the $\xi_{i,j}$'s are concatenated in one long vector denoted by $Y_1, \ldots, Y_l$, then the 'moving blocks' bootstrap estimate of the variance of $\sqrt{N}\bar{X}_N$ is the variance of $\sqrt{l}\bar{Y}_l$ under P^*, and the 'moving blocks' bootstrap estimate of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ is $P^*\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \leq x\}$, where $\bar{Y}_l = \frac{1}{l}\sum_{i=1}^{l} Y_i$.

  As a final step, confidence intervals for μ can be obtained either by means of the Central Limit Theorem using the 'moving blocks' bootstrap estimate of variance, or by approximating the quantiles of the distribution $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ by the corresponding quantiles of $P^*\{\sqrt{l}(\bar{Y}_l - \bar{X}_N) \leq x\}$. If P^* probabilities turn out to be cumbersome to analytically calculate, one can always resort to Monte Carlo, i.e. drawing a large number of samples $\xi_1^{(j)}, \ldots, \xi_k^{(j)}$ i.i.d. from P^*, where $j = 1, \ldots, J$, and evaluating the required probabilities or quantiles empirically from the Monte Carlo set of the J *re-samples*. It is obvious that taking $b = 1$ makes the 'moving blocks' bootstrap coincide with the classical (i.i.d.) bootstrap of Efron(1979).

It can be shown (cf. Lahiri(1990)) that a slightly modified 'moving blocks' bootstrap estimate of sampling distribution turns out to be more accurate than the normal approximation, under some regularity conditions, resulting to more accurate confidence intervals for μ. The modification amounts to approximating the quantiles of $P\{\sqrt{N}(\bar{X}_N - \mu) \leq x\}$ by the corresponding quantiles of $P^*\{\sqrt{l}(\bar{Y}_l - E^*\bar{Y}_l) \leq x\}$, where $E^*\bar{Y}_l$ denotes the expected value of $\bar{Y}_l$ under the P^* probability (conditional on the original data).

To illustrate the 'moving blocks' method, consider the following numerical example taken from Politis and Romano(1989). A sample $Y_1, \ldots, Y_{100}$ was generated from the moving average model: $Y_t = Z_t + Z_{t-1} + Z_{t-2}$, where the Z_t's are i.i.d. N(0,1). By observing that $\sum_{i=1}^{100} Y_i \simeq 3\sum_{i=1}^{100} Z_i$, it is immediate that $Var(\frac{1}{10}\sum_{i=1}^{100} Y_i) \simeq 9$. A plot of the 'moving blocks' bootstrap estimate of the variance of $\frac{1}{10}\sum_{i=1}^{100} Y_i$ as a function of the block size b is shown in Figure 1. As expected, the classical (i.i.d.) bootstrap (that corresponds to the choice $b = 1$) underestimates the variance, yielding an estimate about 3. The 'moving blocks' bootstrap with b near the value 10 seems to give the most accurate estimate, while taking a greater b *worsens* the approximation, and for $b \geq 20$ the 'moving blocks' correction is totally lost.

In Politis and Romano(1989,1990), the 'blocks of blocks' resampling scheme was introduced, in order to address the problem of setting confidence intervals for parameters associated with the whole (infinite-dimensional) distribution of the $X_1, X_2, \ldots$ observations. A prime example of such a parameter is the spectral density function of the $\{X_n\}$ sequence, evaluated at a point. As a by-product, the 'blocks of blocks' method also provides more accurate confidence intervals for parameters associated with a finite-dimensional distribution of the observations, as compared to confidence intervals obtained by the normal approximation. Examples of such parameters include the autocovariance $Cov(X_0, X_s)$ and the autocorrelation $Cov(X_0, X_s)/Var(X_0)$ at lag s. The 'blocks of blocks' scheme is a generalization of the 'moving blocks' method, and the two coincide if the parameter under consideration is the mean EX_1.

The 'blocks of blocks' resampling scheme will be discussed in the next section, together with a general description of the method and its properties. Also discussed is a related resampling scheme, the 'stationary bootstrap' (cf. Politis and Romano(1991)), that has the additional property that the pseudo-sequence obtained by resampling is stationary, as is the original sequence. In addition, some interesting applications and examples of the aforementioned resampling schemes will be presented, where again the objective is to construct accurate estimates of sampling distribution for statistics that are smooth functions of sample-mean type estimators.

Although our main emphasis is in nonparametric estimation, let us mention here that, analogously to the independent case, a parametric bootstrap can be formulated for time series models as well. For example, if it is assumed that the time series is Gaussian, with mean zero and autocovariance sequence R (unknown), bootstrap replications can be constructed by simulating Gaussian time series with mean zero and estimated autocovariance $\hat{R}$, (cf. Ramos(1988)).

For completeness, let us also describe how the classical (i.i.d.) bootstrap can be successfully applied to a certain class of time series models, in the same way it is applied to regression problems. These models are characterized by the fact that they are generated by a 'white' noise sequence (also called the 'residuals'), that is just a sequence of i.i.d. random variables. Although this noise sequence is not directly observable (as is in the classical bootstrap setting), it is in general estimable. The empirical distribution of the estimated residuals is then used in the same fashion the empirical distribution of the sample is used in the classical bootstrap.

To fix ideas, suppose $\{Y_n, n \in \mathbf{Z}\}$ is a stationary autoregressive process of order p, i.e. $Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + e_n$, where $\mathbf{a} = (a_1, \ldots, a_p)$ is a p-dimensional unknown parameter, and the residuals e_n are i.i.d. with (unknown) distribution F, such that $Ee_n = 0$ and $Ee_n^2 = 1$. After observing a finite sample, the problem at hand is to establish confidence intervals for the parameters $a_1, \ldots, a_p$, or for some other related quantity, e.g. $h(a_1, \ldots, a_p)$, where $h(\cdot)$ is a smooth function. The bootstrap algorithm in this situation would be the following (cf. Freedman(1984), Freedman and Peters(1986), Efron and Tibshirani(1986), Swanepoel and van Wyk(1986), Bose(1988)).

- Suppose that a sample $Y_{1-p}, \ldots, Y_N$ is observed and the least-squares estimates $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_p)$ are calculated based on the $Y_1, \ldots, Y_N$ observations. Then, estimates of the residuals are formed by letting $\hat{e}_n = Y_n - \sum_{i=1}^{p} \hat{a}_i Y_{n-i}$, for $n = 1, \ldots, N$. Define the centered estimated residuals as $\tilde{e}_n = \hat{e}_n - \frac{1}{N}\sum_{n=1}^{N} \hat{e}_n$, for $n = 1, \ldots, N$, and let $\tilde{F}_N(\cdot)$ denote the distribution function that puts mass $1/N$ at each $\tilde{e}_n$. Now let $e_1^*, \ldots, e_N^*$ be an i.i.d. sample with distribution $\tilde{F}_N$. Finally generate the resampled sequence Y_n^* as follows: Let $Y_n^* = Y_n$, for $n = 1-p, \ldots, 0$, and let $Y_n^* = \sum_{i=1}^{p} \hat{a}_i Y_{n-i}^* + e_n^*$, for $n = 1, \ldots, N$. Based on the $Y_1^*, \ldots, Y_N^*$ re-sample, new least-squares estimates $\mathbf{a}^* = (a_1^*, \ldots, a_p^*)$ can be computed, and the bootstrap estimate of the (multivariate) sampling distribution of $\sqrt{N}(\hat{\mathbf{a}} - \mathbf{a})$ is the (conditional on the data $Y_{1-p}, \ldots, Y_N$) distribution of $\sqrt{N}(\mathbf{a}^* - \hat{\mathbf{a}})$.

It has been shown (cf. Bose(1988)) that the bootstrap approximation of the sampling distribution of the $\hat{\mathbf{a}}$ estimate is quite accurate; indeed, under some conditions, it is more accurate than the asymptotic normal distribution for $\hat{\mathbf{a}}$. In addition, the bootstrap estimate of the covariance matrix of $\hat{\mathbf{a}}$ is the (conditional on the data) covariance matrix of $\mathbf{a}^*$, and the bootstrap estimate of the sampling distribution of $\sqrt{N}(h(\hat{a}_1, \ldots, \hat{a}_p) - h(a_1, \ldots, a_p))$ is the (conditional on the data) distribution of $\sqrt{N}(h(a_1^*, \ldots, a_p^*) - h(\hat{a}_1, \ldots, \hat{a}_p))$.

The above boostrap algorithm can be easily generalized to mixed autoregressive moving average models, in which the residuals can be estimated in a similar way (cf. Box and Jenkins (1970)). It can also be generalized to nonlinear autoregressions, e.g. of the type $Y_n = f(Y_{n-1}, \ldots, Y_{n-p}) + e_n$, provided the function f can be estimated consistently. Examples (with $p = 1$) of assumed models for the function f include a smooth parametric model, e.g. $f(x) = a_1 cos(x - a_2)$, where a_1, a_2 are the unknown parameters, or a threshold model (cf. Tong(1983)), e.g. $f(x) = a_1 x$, if $x > a$, and $f(x) = a_2 x$, if $x \leq a$, where a, a_1, a_2 are the unknown parameters. Alternatively, f could be estimated nonparametrically, using methods of nonparametric regression such as kernel smoothing and multivariate splines (cf. Friedman(1991), Lewis and Stevens(1990)). Finally, it should be mentioned that in Franke and Härdle (1987) the ideas related to bootstrapping a nonparametric regression (cf. Härdle and Bowman (1988)) are applied in the setting of kernel smoothed spectral estimation.

2. The 'blocks of blocks' general resampling scheme

A general resampling procedure for stationary time series, termed the 'blocks of blocks' resampling scheme (cf. Politis and Romano (1989, 1990)), will now be outlined, and some examples of its implementation will be presented. The 'blocks of blocks' scheme is a non-parametric procedure that yields confidence intervals of asymptotically correct coverage for a parameter associated with the infinite-dimensional joint distribution of the terms in the time series.

Let $X_1, \ldots, X_N$ be observations from the (strictly) stationary multivariate time series $\{X_n, n \in \mathbf{Z}\}$, where X_1 takes values in $\mathbf{R}^d$. The time series $\{X_n, n \in \mathbf{Z}\}$ is assumed to have a weak dependence structure. Specifically, the α-mixing (also called strong mixing) condition will be assumed, i.e. that $\alpha_X(k) \to 0$, as $k \to \infty$, where $\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$, and $A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty$ are events in the σ-algebras generated by $\{X_n, n \leq 0\}$ and $\{X_n, n \geq k\}$ respectively.

With the α-mixing assumption there is associated a certain notion of asymptotic independence of the 'future' of X_k with the 'past' of X_0. Examples where $\alpha_X(k) = O(k^{-\lambda})$, for some $\lambda > 0$, include (but are not limited to) a Gaussian time series with a sufficiently smooth spectral density, or a time series generated by an ARMA (autoregressive moving average) model with innovations that have a general absolutely continuous distribution. In particular, Gaussian ARMA processes with bounded spectral density have the α-mixing coefficients decreasing geometrically, and a moving average (MA) process of order p will have $\alpha_X(k) = 0, \forall k > p$.

As a first step, let us set up the estimation problem in the following manner. Suppose $\mu \in \mathbf{R}^D$ is a parameter of the m-dimensional joint distribution of sequence $\{X_n, n \in \mathbf{Z}\}$, where m could be infinite. For each $N = 1, 2, \ldots$ let $B_{i,M,L}$ be the block of M consecutive observations starting from $(i-1)L+1$, i.e., the subseries $X_{(i-1)L+1}, \ldots, X_{(i-1)L+M}$, where M, L are integer functions of N. Define $T_{i,M,L} = \phi_M(B_{i,M,L})$, where $\phi_M : \mathbf{R}^{dM} \to \mathbf{R}^D$ is some function. So for fixed N, the $T_{i,M,L}$ for $i \in \mathbf{Z}$ constitute a strictly stationary sequence. In practice we would observe a segment $X_1, \ldots, X_N$ from the time series $\{X_n\}$, which would permit us to compute $T_{i,M,L}$ for $i = 1, \ldots, Q$ only, where $Q = [\frac{N-M}{L}] + 1$ and $[\cdot]$ is the integer part function. Also, define the general linear statistic:

$$\bar{T}_N = \frac{1}{Q}\sum_{i=1}^{Q} T_{i,M,L} \tag{3}$$

Under broad regularity conditions $\bar{T}_N$ is a consistent estimator of μ. Loosely stated, these regularity conditions consist of a weak dependence structure (allowing the variance of $\bar{T}_N$ to tend to zero as $N \to \infty$), and a condition of unbiasedness or asymptotic unbiasedness of $T_{1,M,L}$, i.e., $ET_{1,M,L} = \mu$, or $ET_{1,M,L} \to \mu$ as $M \to \infty$.

Some examples of time series statistics that can fit in this framework are the following. For the examples assume X_n is univariate, that is $d = 1$.

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

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

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

$$\phi_M(B_{i,M,L}) = \frac{1}{2\pi M}\Big|\sum_{t=L(i-1)+1}^{L(i-1)+M} W_t^{(M)} X_t e^{-jtw}\Big|^2 \tag{4}$$

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

Note that in example (I), μ is just EX_1, i.e. it is a parameter of the m-dimensional marginal distribution of sequence $\{X_n, n \in \mathbf{Z}\}$, with m=1. Similarly, in example (II), $\mu = EX_0X_s$ is a parameter of the m-dimensional marginal, with m=$s + 1$, and in example (III), μ is the spectral density evaluated at the point w, i.e. a parameter of the whole (infinite-dimensional) joint distribution of $\{X_n, n \in \mathbf{Z}\}$.

With the objective of setting confidence intervals for μ, the 'blocks of blocks' bootstrap procedure goes as follows:

- Define $\mathcal{B}_{j,b}$ to be the block of b consecutive $T_{i,M,L}$'s starting from $T_{j,M,L}$; that is, let $\mathcal{B}_{j,b} = (T_{j,M,L}, \ldots, T_{j-1+b,M,L})$. Note that there are $q = Q - b + 1$ such $\mathcal{B}_{j,b}, j = 1, \ldots, q$. Sampling with replacement from the set $\{\mathcal{B}_{1,b}, \ldots, \mathcal{B}_{q,b}\}$ defines (conditionally on the original observations $X_1, \ldots, X_N$) a probability measure denoted by P^*, which is used in the 'blocks of blocks' bootstrap procedure. Let $Y_1, \ldots, Y_k$ be i.i.d. samples from P^*, where k is of the same asymptotic order as Q/b, (for instance, let $k = [Q/b]+1$). Obviously, each Y_i is a block of size b which we denote as $Y_i = (y_{i1}, \ldots, y_{ib})$. Let us concatenate the y_{ij} in one long vector of size $l = kb$ denoted by $T_1^*, \ldots, T_l^*$, where $T_i^* = y_{rv}$, for $r = [i/b], v = i - br$. Now both $P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \leq x\}$ and $P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x\}$ constitute 'blocks of blocks' bootstrap estimates of $P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x\}$, where $\bar{T}_l^* = \frac{1}{l}\sum_{i=1}^{l} T_i^*$, and the variance of $\sqrt{l}\bar{T}_l^*$ under the P^* probability constitutes the 'blocks of blocks' bootstrap estimate of the variance of $\sqrt{Q}\bar{T}_N$.

As can be easily checked, the 'moving blocks' technique is a special case of the 'blocks of blocks' resampling scheme, as applied to the sample mean example (I). Under mixing and moment conditions, consistency of the 'blocks of blocks' bootstrap estimate of sampling distribution was proved in Politis and Romano(1989,1990) in the general case (where m might be infinite). This is summarized in the following theorem, in which it is assumed that $D = 1$, that is, $\mu \in \mathbf{R}$.

Theorem 1 *Suppose the stationary sequence $\{X_n, n \in \mathbf{Z}\}$ is α-mixing, with $\alpha_X(k) = O(k^{-\lambda})$, for $\lambda > \frac{n_0(2n_0+\delta)}{\delta}$, where n_0 is an integer with $n_0 > 2$, and $0 < \delta \leq 2, C > 0$ are some constants. Also suppose that $E|T_{1,M,L}|^{2n_0+\delta} < C$, for all M, and that $ET_{1,M,L} = \mu + O(Q^{-1/2})$ and $\lim_{N\to\infty} Var(\frac{1}{\sqrt{Q}}\sum_{i=1}^{Q} T_{i,M,L})$ exists and equals $\sigma_\infty^2 > 0$.*
If $M = o(N)$ and $L \sim aM$, for some $a \in (0,1]$, and $b \to \infty$ with $b = o(Q^{1/2})$, then as $N \to \infty$, $Var^*(\sqrt{l}\bar{T}_l^*) \xrightarrow{P} \sigma_\infty^2$, where Var^* denotes the variance under the P^* probability, and the following are true:

$$\sup_x |P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x(Var^*(\sqrt{l}\bar{T}_l^*))^{1/2}\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \leq x\sigma_\infty\}| \xrightarrow{P} 0 \quad (5)$$

$$\sup_x |P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \leq x\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \leq x\}| \xrightarrow{P} 0 \quad (6)$$

where E^* denotes expectation under the P^* probability.
If in addition $ET_{1,M,L} = \mu + o(Q^{-1/2})$, then μ can be substituted instead of $E\bar{T}_N$ in equations (5) and (6).

To actually compute confidence intervals using the (5) or (6) approximations, one has to compute the corresponding quantiles of $P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x(Var^*(\sqrt{l}\bar{T}_l^*))^{1/2}\}$ or $P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \leq x\}$. As mentioned previously, this is usually done by Monte-Carlo, that is, drawing a large number of independent samples from P^* (that are refered to as *re-samples*), and looking at the quantiles of the resulting empirical distribution. Finding a constant C^* that satisfies $P^*\{E^*\bar{T}_l \leq \bar{T}_l^* + C^*\} = 1 - \epsilon$, would then immediately imply that $P\{E\bar{T}_N \leq \bar{T}_N + C^*\} \simeq 1 - \epsilon$, and, also that $P\{\mu \leq \bar{T}_N + C^*\} \simeq 1 - \epsilon$, the latter provided $ET_{1,M,L} = \mu + o(Q^{-1/2})$, that is, if the bias of the estimator $T_{1,M,L}$ is of smaller order than its standard deviation.

To fix ideas, suppose we are looking for a 95% equal tailed confidence interval for μ, under the assumption $ET_{1,M,L} = \mu + o(Q^{-1/2})$. Other types of confidence intervals (e.g. symmetric around μ or smallest length) can be treated in the same way. Starting with the approximation $.975 = P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \leq x^*\} \simeq P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x^*\}$, note that $l^{-1/2}x^* + E^*\bar{T}_l^*$ is just the .975 quantile of the bootstrap distribution $P^*\{\bar{T}_l^* \leq x\}$. Similarly, define y^* such that $l^{-1/2}y^* + E^*\bar{T}_l^*$ is the .025 quantile of $P^*\{\bar{T}_l^* \leq x\}$. Then, the 95% equal tailed confidence interval for μ is $[\bar{T}_N - x^*/\sqrt{Q}, \bar{T}_N - y^*/\sqrt{Q}]$. This, in the terminology of Hall(1988), is a 'hybrid' bootstrap confidence interval based on the approximation (6). Using the approximation (5) would lead to a 'bootstrap-t' confidence interval.

An important observation is that $Var^*(\sqrt{l}\bar{T}_l^*) = \frac{b}{q}\sum_{i=1}^{q}(\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L} - E^*\bar{T}_l^*)^2$, and $E^*\bar{T}_l^* = \frac{1}{q}\sum_{i=1}^{q}\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L}$, both of which can be computed without resampling. The variance estimate $Var^*(\sqrt{l}\bar{T}_l^*)$ is asymptotically equivalent to the estimate $\hat{V}_{JACK}(\sqrt{Q}\bar{T}_N) = \frac{b}{q}\sum_{i=1}^{q}(\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L} - \bar{T}_N)^2$, which is refered to as the 'blocks of blocks' jackknife estimate of variance.

Let us break for a moment to illustrate the practical implementation of the block-resampling scheme just presented by means of an example. Consider the Canadian Lynx data (annual number of lynx trappings in the Mackenzie River for the period 1821 to 1934), that are available to all users of the S statistical language (cf. Becker, Chambers, and Wilks(1988)). It is important to note that the Lynx data have been shown (cf. Subba Rao and Gabr (1980)) to be non-linear (and non-Gaussian), and hence would not succumb to usual linear parametric modeling, such as fitting ARMA (autoregressive moving average) models.

Suppose that we are interested in obtaining a 95% confidence interval for the mean of the annual number of lynx trappings, i.e. EX_1, where the data $X_1, \ldots, X_{114}$ are pictured in Figure 2. Note that $\bar{X}_{114} = \frac{1}{114}\sum_{i=1}^{114} X_i = 1538.018$.

The first step is to compute a variance estimate for $\sqrt{114}\bar{X}_{114}$. Using the estimator $\hat{V}_{JACK}$, (the bootstrap estimate $Var^*(\sqrt{l}\bar{T}_l^*)$ is practically indistinguishable from $\hat{V}_{JACK}$ for all $b = 1, 2, \ldots, 40$), involves chosing the design parameter b appropriately. Different choices of b lead to quite different estimates as shown in Figure 3.

Now, it is known in this case that to have a variance estimator with asymptotically smallest Mean Squared Error (M.S.E.) it is necessary that $b^3 \sim \frac{3N}{4\sigma_\infty^4}(\sum_{s=-\infty}^{\infty} |s|R(s))^2$, where $\sigma_\infty^2 = \lim_{N\to\infty} Var(\sqrt{N}\bar{X}_N)$, and $R(s) = Cov(X_1, X_{1+s})$ (cf. Künsch(1989)). This is hardly a new result, since, for the particular case of the sample mean, $\hat{V}_{JACK}$ is essentially a nonparametric estimate of the spectral density at zero, smoothed by Bartlett's kernel. A rough estimate of $\sum_{s=-\infty}^{\infty} |s|R(s)$ based on the sample autocovariances can be obtained. However, σ_∞^2 is unknown. In fact, this is exactly what $\hat{V}_{JACK}$ asymptotically estimates. Looking at the variability of $\hat{V}_{JACK}$ for different choices of b, indicates that this route for choosing b is really a vicious circle.

Therefore, the choice of b requires a heuristic guess based on studying the data more deeply. In the case of the lynx data, looking at the estimated sample autocovariance sequence $\hat{R}(s) = \frac{1}{114}\sum_{i=1}^{114-|s|} X_i X_{i+|s|}$, or at its unbiased version $\frac{114}{114-|s|}\hat{R}(s)$, is particularly enlightning. In Figure 4, a plot of $\hat{R}(s)/\hat{R}(0)$ is presented, i.e. the sample autocorrelation sequence. There is a clear indication of a certain cyclic behaviour with a period of about ten years (cf. Cambell and Walker (1977) and the references therein). Bearing in mind that $Var(\sqrt{N}\bar{X}_N) = R(0) + 2\sum_{i=1}^{N-1}(1 - \frac{i}{N})R(i)$, it can be seen that chosing $b = b_0$ effectively retains only the first b_0 autocovariances in the summation. But by the almost-periodicity of $R(s)$, one expects that there will be a lot of cancellations among the terms in the above sum. Hence, it would be advisable to let b be of the order of two or three cycles, in order to allow for the cancellations to take place. We would opt to take $b = 25$ in this case, leading to a variance estimate of 2,873,828.

Returning to the problem of determining b via minimum M.S.E. considerations, it is quite interesting that simultaneously trying to solve for b and $\hat{V}_{JACK}$ the equation presented in Figure 3, and the equation $b^3\hat{V}^2_{JACK} = (3N/4)\tilde{R}^2$, with $\tilde{R} = 2\sum_{s=1}^{40} s(1-s/40)\hat{R}(s) = -30,533,255$ being an estimate of $\sum_{s=-\infty}^{\infty} |s|R(s)$ using Bartlett's kernel, actually leads to $b = 22$ and an estimated variance of 2,853,373 (taken from Figure 3 with $b = 22$). However, it should be noted that due to the asymptotic nature of this way of determining b (and due to the high variance of $\tilde{R}$) we would not a *priori* trust this result if it were not supported by some other reasoning regarding the process.

Now that a variance estimate for $\bar{X}_N$ is available (the variance estimate is 2,873,828/114 = 25209, and the standard deviation estimate is 158.8), a 95% confidence interval for EX_1 using the Central Limit Theorem would be [1226.8, 1849.3]. Note that this confidence interval could be inaccurate because the distribution of the lynx data seems to be excessively skewed (see Figure 5). This observation amounts to worrying about the rate of convergence in the Central Limit Theorem, which is made worse by the dependence among the data.

A traditional way out of this difficulty is the use of transformations. In the literature (cf. Subba Rao and Gabr(1980), Cambell and Walker(1977)) it is suggested to use the logarithmic transformation on the data $X_1, \ldots, X_{114}$. However, this would lead to a confidence interval for $E \log X_1$, that does not immediately relate to a confidence interval for EX_1. Alternatively, one can use a transformation on $\bar{X}_N$ and the δ-method (cf. Miller(1986)). Using the logarithmic transformation on $\bar{X}_N$ (and the asymptotic normal distribution of $\log \bar{X}_N$) leads to the [1256.3, 1882.9] 95% confidence interval for EX_1, which is markedly different from the interval based on a normal approximation for $\bar{X}_N$.

The question then is: which of the two confidence intervals is better? The bootstrap solves this dilemma, because it automatically captures the skewness without the need for transformations. Heuristically, it is like if the bootstrap implicitly employs an 'optimal' in some sense normalizing transformation before constructing its distribution estimate (cf. Efron (1979, 1982, 1987)).

In Figures 6 and 7, histograms of the 'moving blocks' bootstrap distribution of $\bar{X}_N$ are presented for two bootstrap simulations. The choice of design parameters was for Figure 6, $k = 20$, $b = 25$, and the number of bootstrap replications was $J = 500$; for Figure 7, $k = 5$,

$b = 25$, and $J = 100$. The 95% bootstrap (hybrid) confidence intervals for EX_1 were [1233.37, 1826.07] (Figure 6), and [1221.03, 1862.62] (Figure 7). Note that because in general $l = kb$ is not equal to Q, the bootstrap confidence limits can not be 'read' immediately from the histograms. A re-normalization is in order, as discussed after the statement of Theorem 1.

It is apparent that the bootstrap confidence intervals are quite close to the confidence interval derived by the Central Limit Theorem. This could be guessed by looking at the bootstrap histograms which do not appear to be too skewed (the histogram of Figure 6 has coefficient of skewness 0.333, compared to 1.33 of the empirical distribution of the lynx data). It is also apparent that a reasonably accurate bootstrap confidence interval can be obtained by as few as 100 Monte Carlo replications (although it is *always* advisable to take a larger number).

It is interesting to note that if μ is a parameter of the m-dimensional marginal distribution of sequence $\{X_n\}$, with m finite, then M could be taken to be a *fixed* constant equal to m, and L can be taken equal to one in the 'blocks of blocks' procedure. In this case, and under some additional regularity conditions (including that $ET_{1,M,L} = \mu$, and that $\alpha(k)$ has an exponential decay), it has been proved (Lahiri(1990), Politis and Romano(1990)) that the approximation provided by equation (5) is more than first-order accurate. This fact establishes that the bootstrap approximation (5) is preferable to the normal approximation provided by a Central Limit Theorem for $\bar{T}_N$, especially if there is significant skewness in the distribution of the $T_{i,M,L}$'s.

The reason that (5) provides a more accurate approximation than (6) is that $E^*\bar{T}_l^* = \bar{T}_N + O_p(b/Q)$, i.e. the distribution of $\bar{T}_l^*$ under P^* possesses a random bias of significant order. This bias is associated with the 'blocks of blocks' resampling scheme that assigns reduced weight to $T_{i,M,L}$'s with $i < b$ or $i > Q - b + 1$. In other words, if we let P_i^* be the limit (almost sure in P^*) of the proportion l^{-1}(number of the T_j^*'s that equal $T_{i,M,L}$) as $l \to \infty$, (and assuming no ties among the $T_{i,M,L}$'s), although $P_i^* = b/R$, with $R = b(Q - b + 1)$, for any i such that $b \leq i \leq Q - b + 1$, this proportion drops to $P_i^* = i/R$, for any $i < b$, and $P_i^* = (Q - i + 1)/R$, for any $i > Q - b + 1$.

A way to dispense with this difficulty is to 'wrap' the $T_{i,M,L}$'s around in a 'circle', that is, to define (for $i > Q$) $T_{i,M,L} \equiv T_{i_Q,M,L}$, where $i_Q = i(\text{mod}Q)$, and $T_{0,M,L} \equiv T_{Q,M,L}$. This idea is incorporated in the following general formulation which, for reasons to become apparent later, will be called the *stationary* (blocks of blocks) resampling scheme.

- With the $T_{i,M,L}$'s defined for all i as above, define the $\mathcal{B}_{j,b}$ as previously, but note that now for any integer b there are Q such $\mathcal{B}_{j,b}, j = 1, \ldots, Q$. Let p be a number in [0,1]. Independent of $X_1, \ldots, X_N$, let $L_1, L_2, \ldots$ be a sequence of independent and identically distributed random variables having the geometric distribution, so that the probability of the event $\{L_i = n\}$ is $(1-p)^{n-1}p$ for $n \in \mathbf{N}$. Independent of the X_i and the L_i, let $I_1, I_2, \ldots$ be a sequence of independent and identically distributed variables which have the discrete uniform distribution on $\{1, \ldots, N\}$. Now, a new pseudo-sequence $T_1^\star, \ldots, T_Q^\star$ is generated in the following way. Sample a sequence of blocks of random length by the prescription $\mathcal{B}_{I_1,L_1}, \mathcal{B}_{I_2,L_2}, \ldots$. The first L_1 observations in the pseudo-sequence $T_1^\star, \ldots, T_Q^\star$ are determined by the first resampled block $\mathcal{B}_{I_1,L_1}$, the next L_2 observations in the pseudo-sequence are the observations in the second resampled block $\mathcal{B}_{I_2,L_2}$, and so on. The process is stopped once Q observations in the pseudo-sequence have been generated. It is not hard to see that the pseudo-sequence $T_1^\star, \ldots, T_Q^\star$ is *stationary*, conditionally on the original data.

  This method of resampling and generating $T_1^\star, \ldots, T_Q^\star$ defines a (conditional on the original data $X_1, \ldots, X_N$) probability measure $P^\star$. If we define $\bar{T}_Q^\star = \frac{1}{Q}\sum_{i=1}^{Q} T_i^\star$, then it is easy to see that $E^\star \bar{T}_Q^\star = \bar{T}_N$, where $E^\star$ denotes expectation under the $P^\star$ probability. Hence, the stationary blocks of blocks bootstrap estimate of $P\{\sqrt{Q}(\bar{T}_N - \mu) \leq x\}$ is $P^\star\{\sqrt{Q}(\bar{T}_Q^\star - \bar{T}_N) \leq x\}$.

Under mixing and moment conditions, it can be shown (cf. Politis and Romano (1991)) that the stationary bootstrap estimates of variance and of distribution are consistent. Considering the Lynx data again, a problem equivalent to determining b in the 'moving blocks' or 'blocks of blocks' boostrap presents itself, namely choosing the design parameter p. It can be shown (cf. Politis and Romano(1991)) that the stationary bootstrap estimate of the variance of $\sqrt{N}\bar{X}_N$

can be analytically calculated by

$$\hat{V}_{St.B.} = \hat{R}(0) + 2\sum_{i=1}^{N-1} h_N(i)\hat{R}(i) \tag{7}$$

where $h_N(i) = (1 - \frac{i}{N})(1-p)^i + \frac{i}{N}(1-p)^{N-i}$. Note that in the stationary bootstrap there are blocks of random length. Since the average length of these blocks is $1/p$, it is expected that the quantity $1/p$ should play a similar role as the parameter b in the moving blocks method. In Figure 8, the $\hat{V}_{St.B.}$ variance estimates are pictured for different values of p between 0 and 1/2. Choosing $p = 0.05$ corresponds to a variance estimate of 2,335,502.

It should be stressed that all the abovementioned block-resampling methods give valid results for statistics that are smooth functions of sample means, since 'the bootstrap commutes with smooth functions' (cf. Bickel and Freedman (1981)). In addition, the moving blocks and the stationary bootstrap methods have been shown to work with statistics that are representable by smooth functionals, i.e. that can only be approximated by sample means (cf. Künsch(1989), Liu and Singh(1988), Politis and Romano(1991)), as is, for example, a trimmed-mean.

It should also be pointed out that the 'blocks of blocks' method is asymptotically correct in the general case where $\mu \in \mathbf{R}^D$, in which case $\bar{T}_N$ and the $T_{i,M,L}$'s are multivariate. Denote $\mu^{(n)}, \bar{T}_N^{(n)}$, etc. to be the nth coordinate of $\mu, \bar{T}_N$, and so forth. The following multivariate theorem was proved in Politis and Romano(1990), while a similar result is true for the multivariate stationary bootstrap (Politis and Romano(1991)).

Theorem 2 *Suppose the stationary sequence $\{X_n, n \in \mathbf{Z}\}$ is α-mixing, with $\alpha_X(k) = O(k^{-\lambda})$, for $\lambda > \frac{n_0(2n_0+\delta)}{\delta}$, where n_0 is an integer with $n_0 > 2$, and $0 < \delta \le 2, C > 0$ are some constants. Suppose that $E|T_{1,M,L}^{(n)}|^{2n_0+\delta} < C$, for all M, and for all $n = 1, 2, \ldots, D$. Also suppose that $ET_{1,M,L} = \mu + O(Q^{-1/2})$ and $\lim_{N\to\infty} Cov(\frac{1}{\sqrt{Q}}\sum_{i=1}^{Q} T_{i,M,L}^{(n_1)}, \frac{1}{\sqrt{Q}}\sum_{i=1}^{Q} T_{i,M,L}^{(n_2)})$ exists and equals $\Sigma_\infty^{(n_1),(n_2)}$, with $\Sigma_\infty^{(n),(n)} > 0$ for all $n = 1, 2, \ldots, D$.*
If $M = o(N)$ and $L \sim aM$, for some $a \in (0,1]$, and $b \to \infty$ with $b = o(Q^{1/2})$, then as $N \to \infty$, $Cov^*(\sqrt{l}\bar{T}_l^*) \xrightarrow{P} \Sigma_\infty$, where Cov^* denotes the covariance matrix of $\sqrt{l}\bar{T}_l^*$ under the P^* probability, and the following are true:

$$\sup_{x\in\mathbf{R}^D} |P^*\{\sqrt{l}(\bar{T}_l^* - E^*\bar{T}_l^*) \le (Cov^*(\sqrt{l}\bar{T}_l^*))^{1/2}x\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \le \Sigma_\infty^{1/2}x\}| \xrightarrow{P} 0 \tag{8}$$

$$\sup_{x \in \mathbf{R}^D} |P^*\{\sqrt{l}(\bar{T}_l^* - \bar{T}_N) \le x\} - P\{\sqrt{Q}(\bar{T}_N - E\bar{T}_N) \le x\}| \xrightarrow{P} 0 \tag{9}$$

where $(Cov^*(\sqrt{l}\bar{T}_l^*))^{1/2}$ and $\Sigma_\infty^{1/2}$ represent the 'square root' of these positive definite matrices given by some decomposition procedure, and the (n_1, n_2) element of $Cov^*(\sqrt{l}\bar{T}_l^*)$ can be computed analytically as $\frac{b}{q}\sum_{i=1}^{q}(\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L}^{(n_1)} - E^*\bar{T}_l^{*(n_1)})(\frac{1}{b}\sum_{j=i}^{i-1+b} T_{j,M,L}^{(n_2)} - E^*\bar{T}_l^{*(n_2)})$.
If in addition $ET_{1,M,L} = \mu + o(Q^{-1/2})$, then μ can be substituted instead of $E\bar{T}_N$ in equations (8) and (9).

An important implication of theorem 2 is that by the multivariate blocks of blocks resampling scheme we can get asymptotically correct approximations to the sampling distributions of continuous functions of $\bar{T}_N$. For example, with no extra effort we can get approximations to the distribution of $\max_{n=1,2,\ldots,D}\{\sqrt{Q}|\bar{T}_N^{(n)} - \mu^{(n)}|\}$, which allows for the possibility of constructing simultaneous confidence intervals for all coordinates of μ (cf. Politis and Romano (1990)). In particular, in Politis, Romano, and Lai (1990), the case where μ is the spectral or cross-spectral density function sampled at a grid of points was studied, with the objective of setting uniform confidence bands.

As our final example, consider the important case where the parameter of interest is the autocorrelation coefficient at lag s, i.e. the parameter $\rho(s) = R(s)/R(0)$, where $R(s) = EX_0X_s$ and for simplicity it is assumed that $EX_0 = 0$. In that case, the linear statistic $\bar{T}_N$ is $(s+1)$-dimensional, with $\bar{T}_N^{(n)} = \frac{1}{N-s}\sum_{i=1}^{N-s} X_iX_{i+n-1}$, and $L = 1$, $M = s+1$ and $\phi_M^{(n)}(x_1, \ldots, x_M) = x_1x_n$, for $n = 1, \ldots, s+1$. It is easy to see that $\bar{T}_N^{(n)}$, for $n = s+1$, is just the sample autocovariance $\hat{R}(s)$ at lag s. Via the 'blocks of blocks' resampling scheme applied to the linear statistic $\bar{T}_N$, accurate confidence intervals for the autocovariances can be obtained, as well as variance estimates of the sample autocovariance estimates. Considering the complicated form of the asymptotic variance of the sample autocovariances (that involves estimates of the fourth order cumulants, cf. Anderson(1971)), the advantage of using an automatic procedure like the bootstrap is apparent.

Now the estimator $\hat{\rho}(s) = \bar{T}_N^{(s+1)}/\bar{T}_N^{(1)}$ is a smooth function of the linear statistic $\bar{T}_N$, and its statistical properties can be analyzed via the 'blocks of blocks' bootstrap. Of course, if we are only interested in $\rho(s)$, a 2-dimensional linear statistic, consisting of just $\bar{T}_N^{(1)} = \hat{R}(0)$ and $\bar{T}_N^{(s+1)} = \hat{R}(s)$, would suffice. The advantage of considering the $(s+1)$-dimensional statistic $\bar{T}_N$

is that we can instantly obtain *simultaneous* confidence intervals (confidence band) for $\rho(k), k = 1, \ldots, s$ (and for $R(0)$), that are *not* available by classical methods (cf. Priestley(1981)). An obvious use of such confidence bands is in testing hypotheses regarding the covariance structure.

The way this can be done is as follows. For concreteness, assume that we are looking for a 95% confidence band for $\rho(k), k = 1, \ldots, s$. That is, we are looking for two sequences $c_1(k), c_2(k)$ such that $P\{\forall k \in \{1, \ldots, s\} : \hat{\rho}(k) - c_1(k) \leq \rho(k) \leq \hat{\rho}(k) + c_2(k)\} = 0.95$. To start with, apply Fisher's z transformation to approximately stabilize the variance of the estimates at different lags, i.e. let $\zeta(k) = \frac{1}{2}\log\frac{1+\rho(k)}{1-\rho(k)}$, and $\hat{\zeta}(k) = \frac{1}{2}\log\frac{1+\hat{\rho}(k)}{1-\hat{\rho}(k)}$, for $k = 1, \ldots, s$. Then, by the 'blocks of blocks' bootstrap, obtain an approximation to the distribution of the 'maximum modulus' $\sqrt{N}\max_{k=1,\ldots,s}|\hat{\zeta}(k) - \zeta(k)|$. This immediately leads to a uniform width (i.e. $c_1(k) = c_1$ and $c_2(k) = c_2, k = 1, \ldots, s$) and symmetric (i.e. $c_1(k) = c_2(k)$) confidence band for $\zeta(k), k = 1, \ldots, s$, and can be translatedto a confidence band (of non-uniform width) for $\rho(k), k = 1, \ldots, s$. Alternatively, we can get a (non-symmetric in general) equal-tailed uniform width confidence band for $\zeta(k), k = 1, \ldots, s$, by finding bootstrap approximations to x and y such that $P\{\sqrt{N}\max_{k=1,\ldots,s}(\hat{\zeta}(k) - \zeta(k)) \leq x\} = 0.975$, and $P\{\sqrt{N}\min_{k=1,\ldots,s}(\hat{\zeta}(k) - \zeta(k)) \leq y\} = 0.025$.

As an illustration, a time series $Y_1, \ldots, Y_{200}$ was generated according to the model

$$X_t - 1.352X_{t-1} + 1.338X_{t-2} - 0.662X_{t-3} + 0.240X_{t-4} = Z_t - 0.2Z_{t-1} + 0.04Z_{t-2}$$

$$Y_t = X_t|X_t|$$

where the Z_t's are independent normal N(0,1) random variables. A plot of the $Y_1, \ldots, Y_{200}$ data set is presented in Figure 9, and Figure 10 contains a histogram of the empirical distribution of the data, that reflects the non-normal distribution of the Y_t's.

In Figures 11 and 12, 95% equal-tailed confidence bands are set for $\zeta(k)$ and $\rho(k), k = 1, \ldots, 10$, via the 'blocks of blocks' bootstrap, with design parameters, $b = 15, k = 20$, and number of bootstrap replications $J = 500$. Observe that although the estimates $\hat{\zeta}(k)$ and $\hat{\rho}(k)$ are indistinguishable, (a consequence of the fact that Fisher's z transformation is almost an identity for $|\rho(k)| < 0.4$), the corresponding to them confidence bands are quite different. In fact, the confidence band for $\zeta(k)$ is constructed so that it has the property of uniform width along k, which can not be carried over to the band for $\rho(k)$.

In both figures, the middle curve can be considered to be the 'true' values of $\zeta(k)$ and $\rho(k)$. These are really estimates of $\zeta(k)$ and $\rho(k)$ obtained by generating a Y_t time series of length 20,000. From this extra long stretch of the Y_t time series, one hundred approximately independent series of length 200 were extracted in order to get an empirical estimate of the distribution of the 'maximum modulus', to be compared to its boostrap approximation.

In Figures 13 and 14, histograms of the boostrap and the empirical estimates of the distribution of $\sqrt{N}\max_{k=1,\ldots,s}|\hat{\zeta}(k) - \zeta(k)|$ are pictured. Maybe due to the small number (one hundred) of available series, the empirical distribution has a shorter right tail than its bootstrap counterpart. This implies that the bootstrap confidence band would be more conservative (i.e. wider) than a confidence band based on this empirical distribution, provided the latter was somehow available.

References

[1] Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley.

[2] Becker, R.A., Chambers, J.M., Wilks, A.R. (1988) *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole.

[3] Bickel, P.J. and Freedman, D.A. (1981), Some Asymptotic Theory for the Bootstrap, *Ann.Statist.*, 9, 1196-1217.

[4] Bose, A. (1988), Edgeworth correction by bootstrap in autoregressions, *Ann.Statist.*, 16, 1709-1722.

[5] Box, G.E.P. and Jenkins, G.M. (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

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

[7] Cambell, M.J. and Walker, A.M. (1977), A Survey of Statistical Work on the MacKenzie River Series of Annual Canadian Lynx Trappings for the Years 1821 - 1934 and a New Analysis, *J. Roy. Statist. Soc. (A)*, 140, 411-431.

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

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

[10] Efron, B. (1987), Better Bootstrap Confidence Intervals, *J. Amer. Statist. Assoc.*, 82, 171-185.

[11] Efron, B. and Tibshirani, R.J. (1986), Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statist. Sci.* 1, 54-77.

[12] Franke,J. and Härdle,W. (1987), On bootstrapping kernel spectral estimates, Preprint (to appear in *Ann. Statist.*).

[13] Freedman, D.A.(1981), Bootstrapping regression models, *Ann.Statist.*, 9, 1218-1228.

[14] Freedman, D.A. and Peters, S.C. (1984), Bootstrapping an econometric model: Some empirical results, *J. Bus. Econom. Statist.*, 2, 150-158.

[15] Friedman, J. (1991), Multivariate Adaptive Regression Splines, to appear in *Ann.Statist.*.

[16] Hall, P. (1988), Theoretical Comparison of Bootstrap Confidence Intervals, *Ann. Statist.*, 16, 927-953.

[17] Härdle,W. and Bowman, A.(1988), Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands, *J. Amer. Statist. Assoc.*, 83, 100-110.

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

[19] Lahiri, S.N.(1990), Second order optimality of stationary bootstrap, *Proc. of 214th IMS Meeting*, East Lancing, Michigan. (Also: Technical Report No.90-1, Department of Statistics, Iowa StateUniversity.)

[20] Lewis, P. and Stevens, J. C. (1990), Using MARS to investigate non-linear time series, *Proc. Interface'90 Symposium*, East Lancing, Michigan.

[21] Liu, R.Y. and Singh, K. (1988), Moving Blocks Jackknife and Bootstrap Capture Weak Dependence, Unpublished manuscript, Department of Statistics, Rutgers University.

[22] Miller, R.G. (1986), *Beyond ANOVA, Basics of Applied Statistics*, John Wiley.

[23] Politis D.N. and Romano, J.P. (1989), A General Resampling Scheme for Triangular Arrays of α-mixing Random Variables with application to the problem of Spectral Density Estimation, Technical Report No.338, Department of Statistics, Stanford University.

[24] Politis D.N. and Romano, J.P. (1990), A Nonparametric Resampling Procedure for Multivariate Confidence Regions in Time Series Analysis, Proceedings of INTERFACE '90, 22nd Symposium on the Interface of Computing Science and Statistics.

[25] Politis D.N. and Romano, J.P. (1991), The Stationary Bootstrap, Technical Report No. 365, Department of Statistics, Stanford University.

[26] Politis D.N., Romano, J.P., and Lai, T.-L. (1990), Boostrap Confidence Bands for Spectra and Cross-Spectra, Technical Report No. 342, Department of Statistics, Stanford University.

[27] Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press.

[28] Ramos, E. (1988), Resampling methods for time series, Technical Report No. ONR-C-2, Department of Statistics, Harvard University.

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

[30] Subba Rao, T. and Gabr, M.M. (1980), A test for linearity of stationary time series, *J. Time Ser. Anal.*, 1, 145-158.

[31] Swanepoel, J.W.H. and van Wyk, J.W.J. (1986), The bootstrap applied to spectral density function estimation, *Biometrika* **73**, 135-142.

[32] Tong, H.(1983), *Threshold Models in Non-linear Time Series Analysis*, Springer Lecture Notes in Statistics, 21.

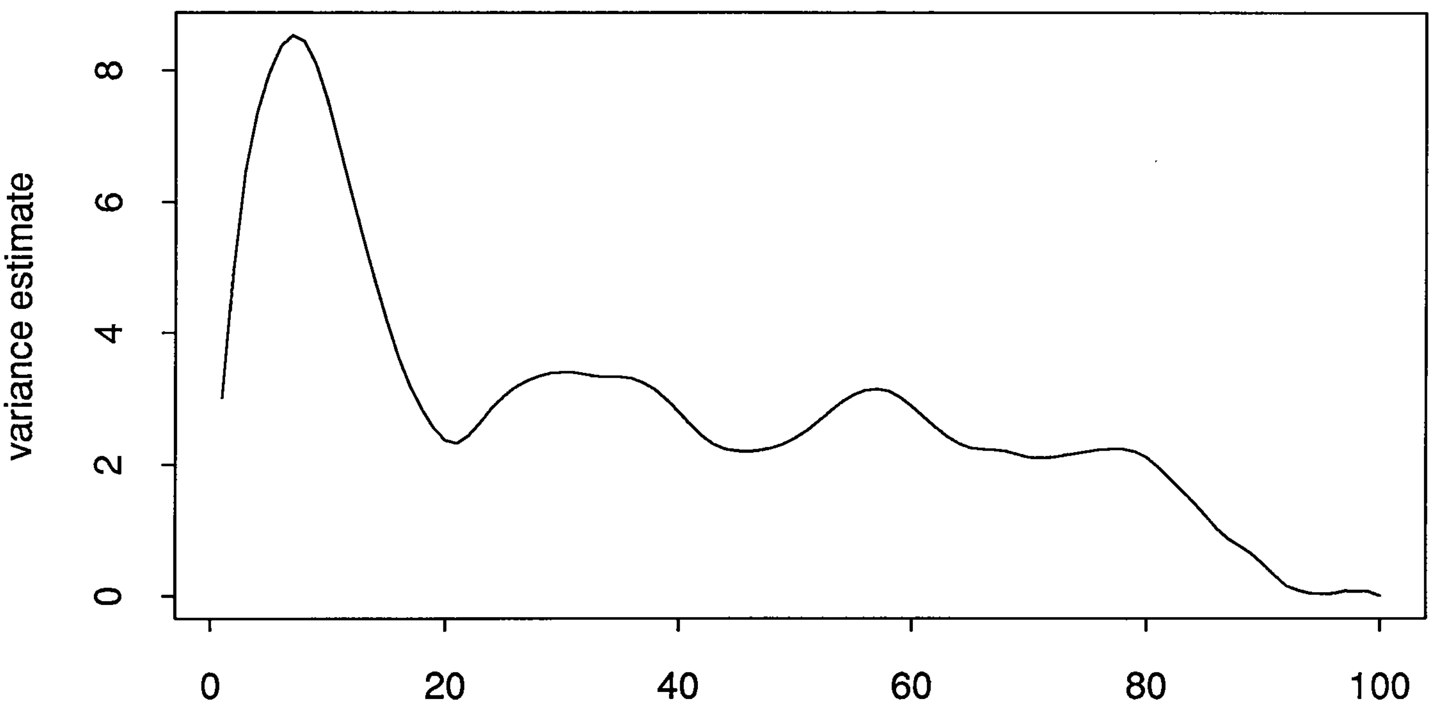

FIGURE 1. The moving blocks estimate as a function of the block size b

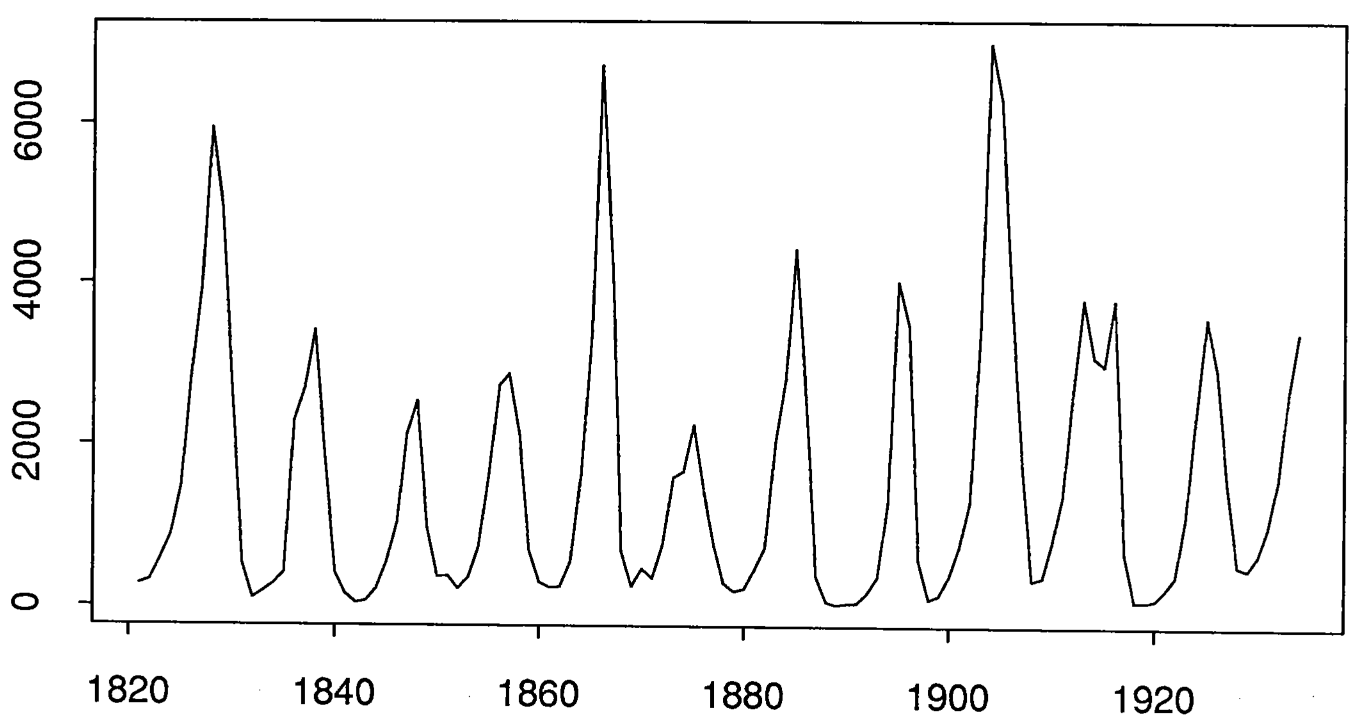

FIGURE 2. The time series of annual number of lynx trappings

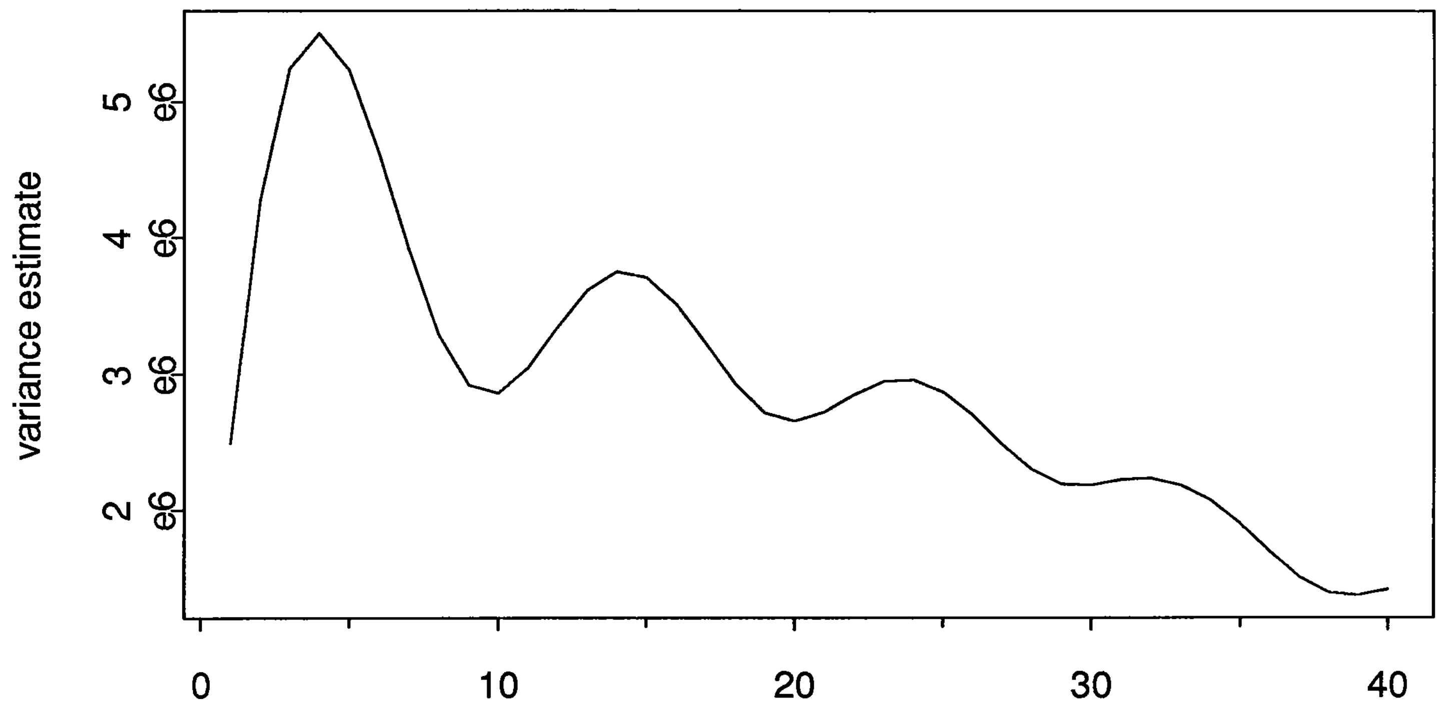

FIGURE 3. The moving blocks Lynx data estimate as a function of b

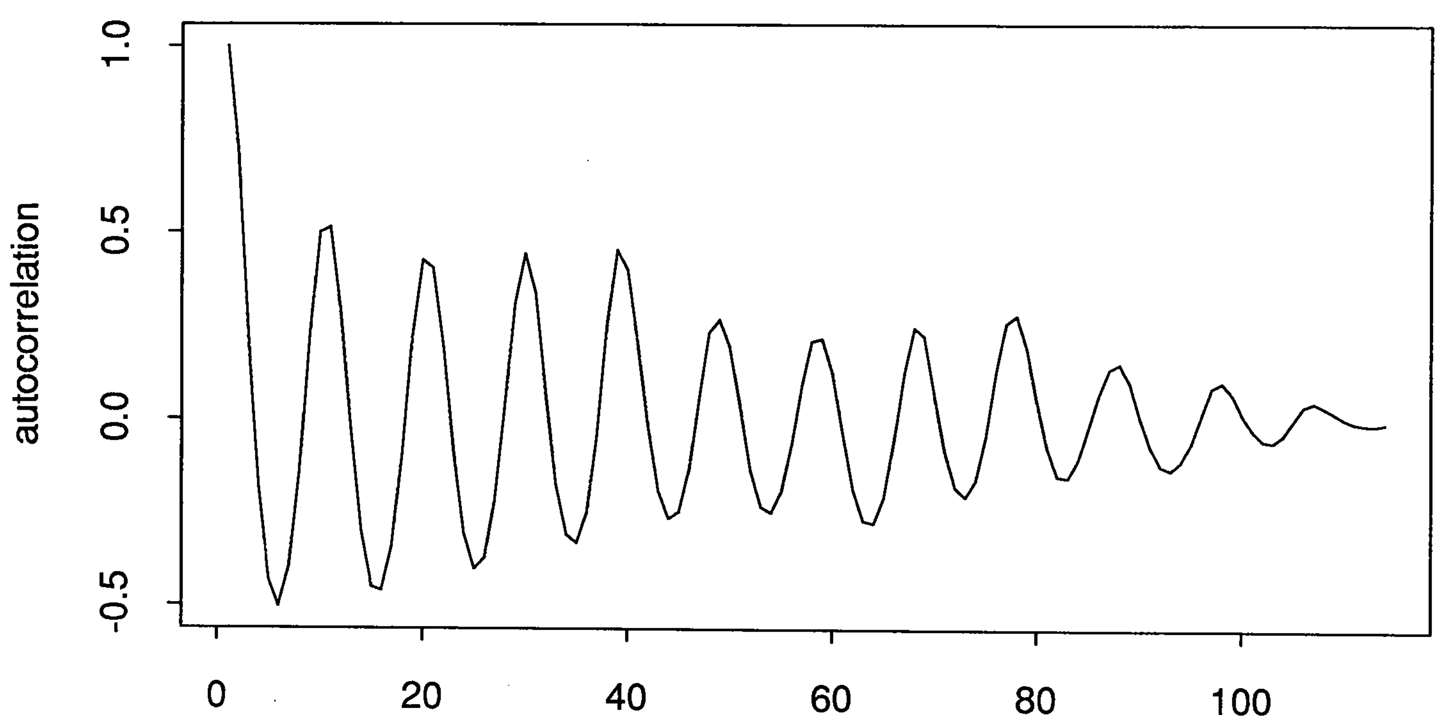

FIGURE 4. Sample autocorrelation sequence of Lynx data

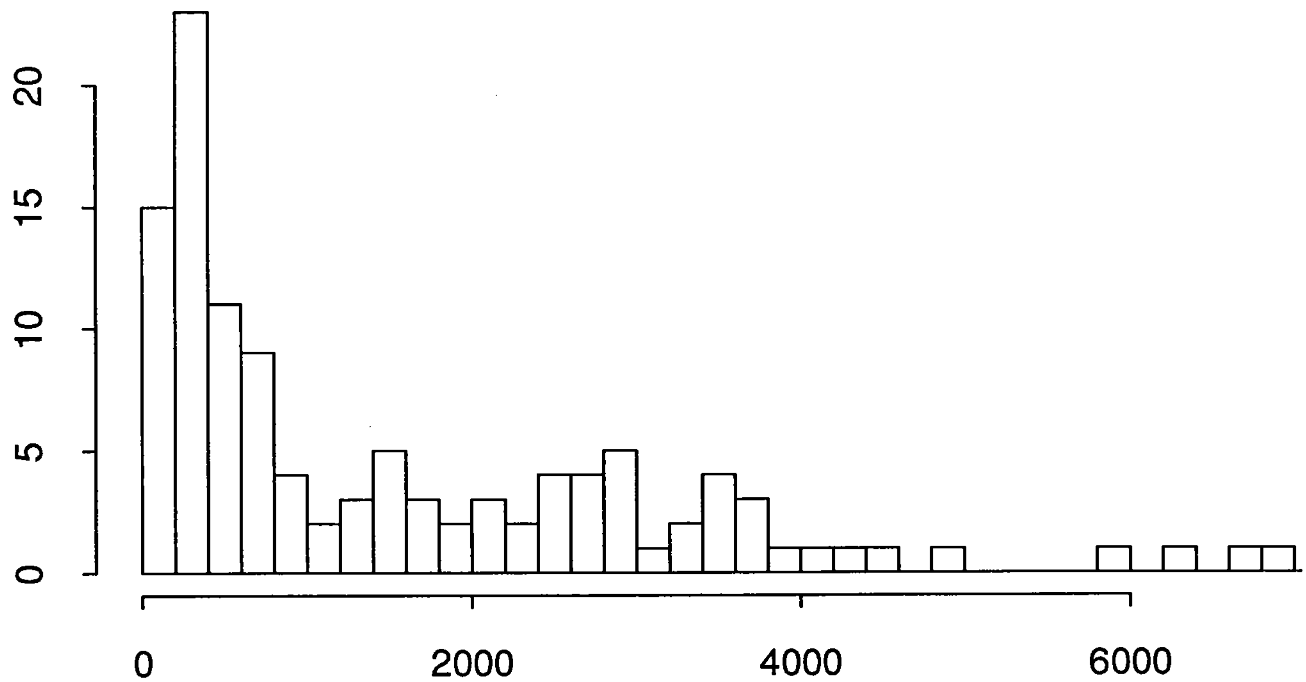

FIGURE 5. Histogram of Lynx data

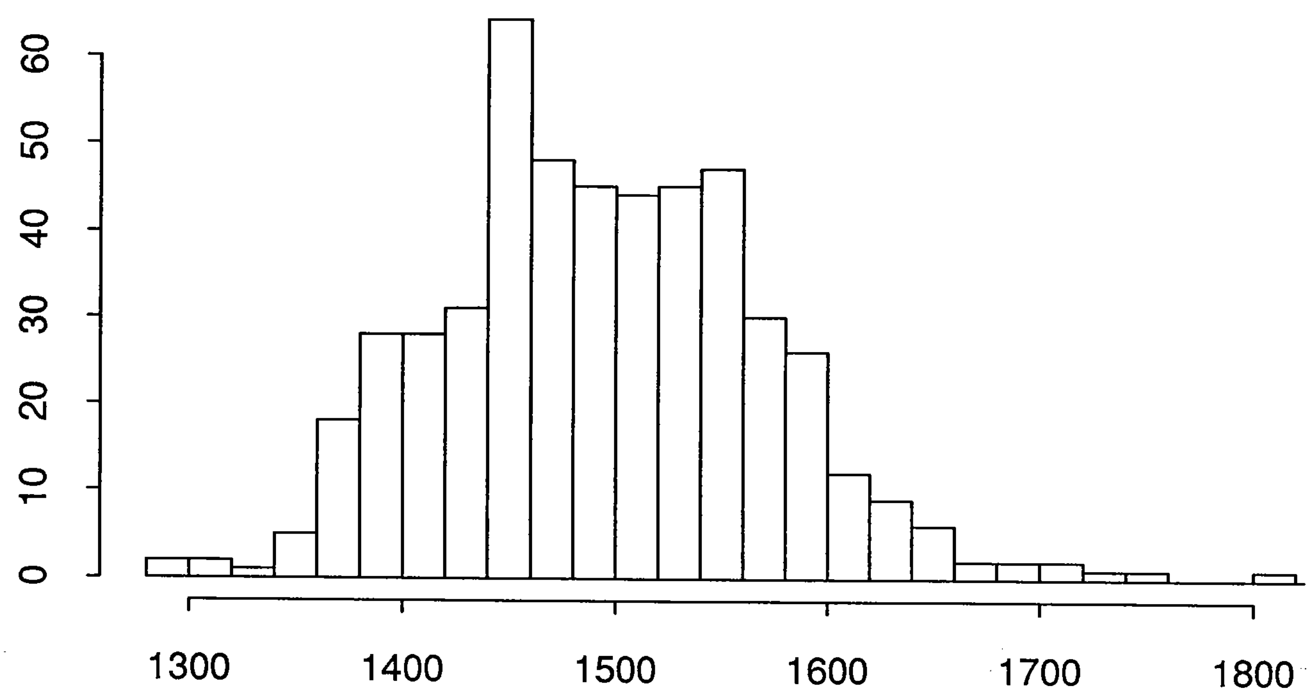

FIGURE 6. Bootstrap distribution of Lynx data sample mean (500 replications)

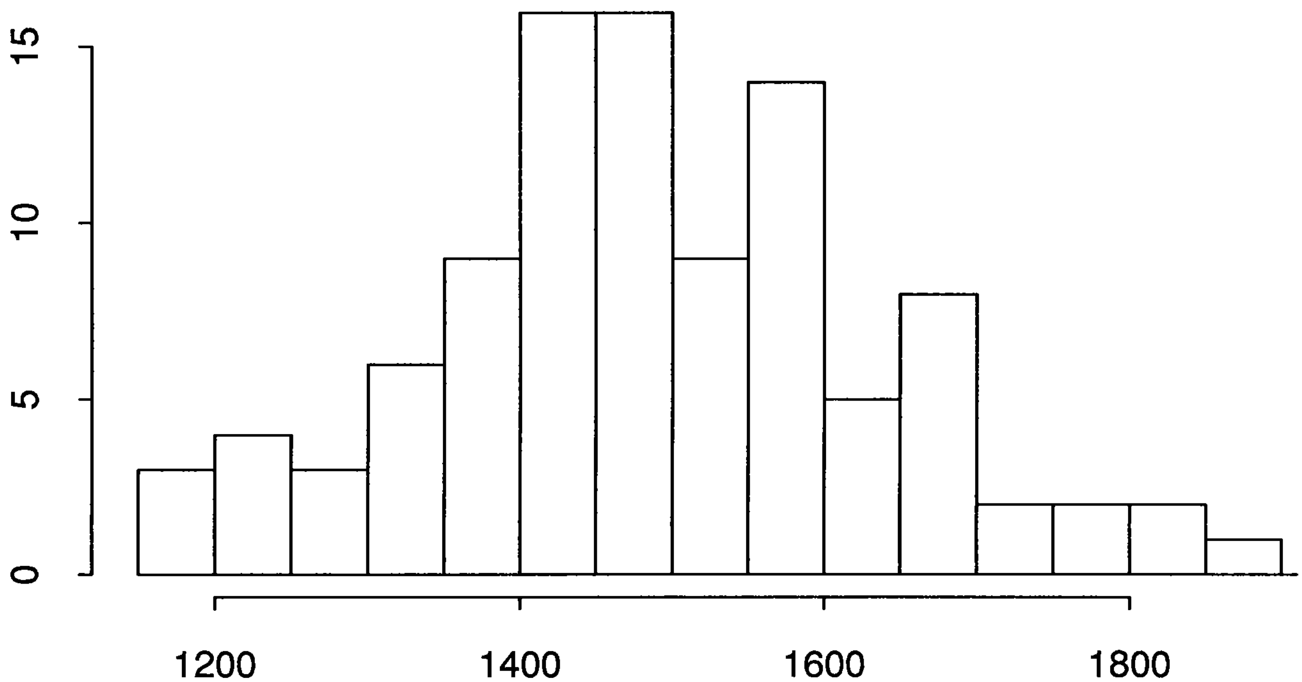

FIGURE 7. Bootstrap distribution of Lynx data sample mean (100 replications)

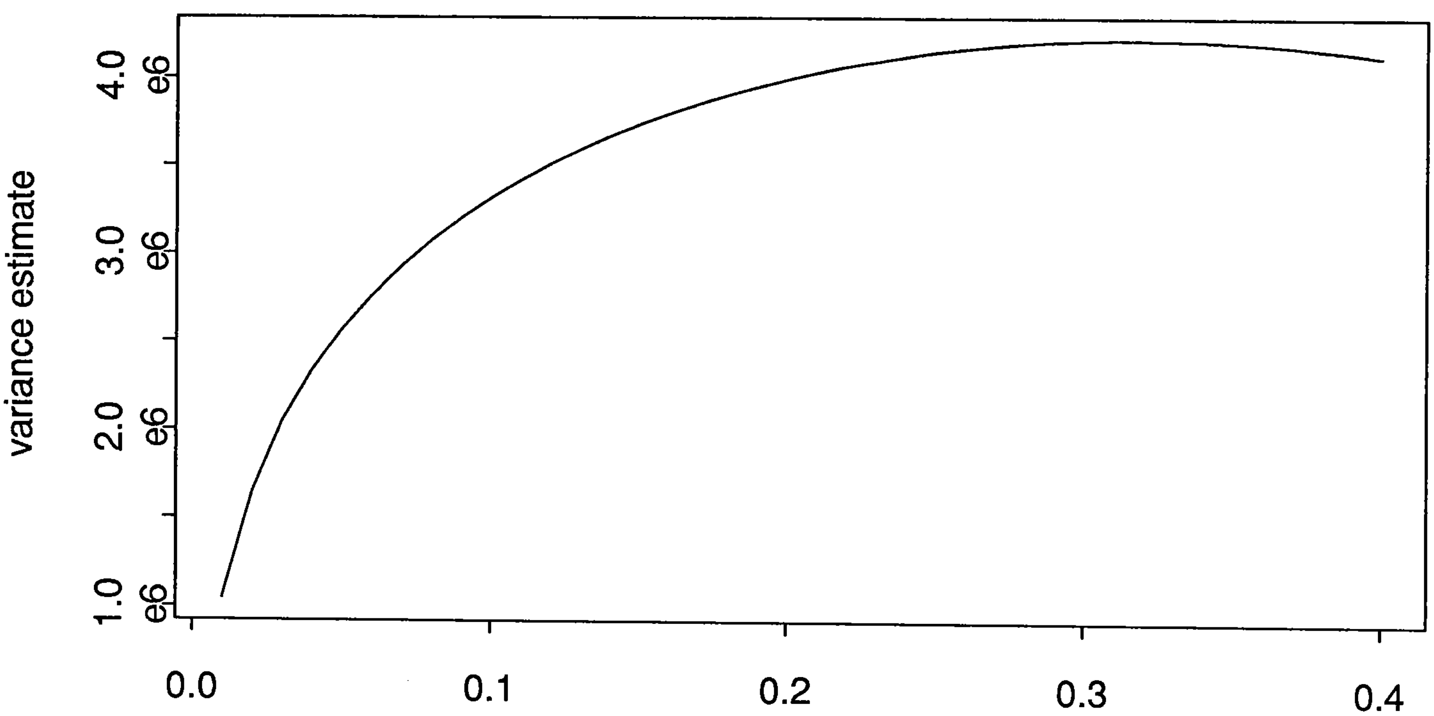

FIGURE 8. The stationary bootstrap Lynx estimate as a function of p

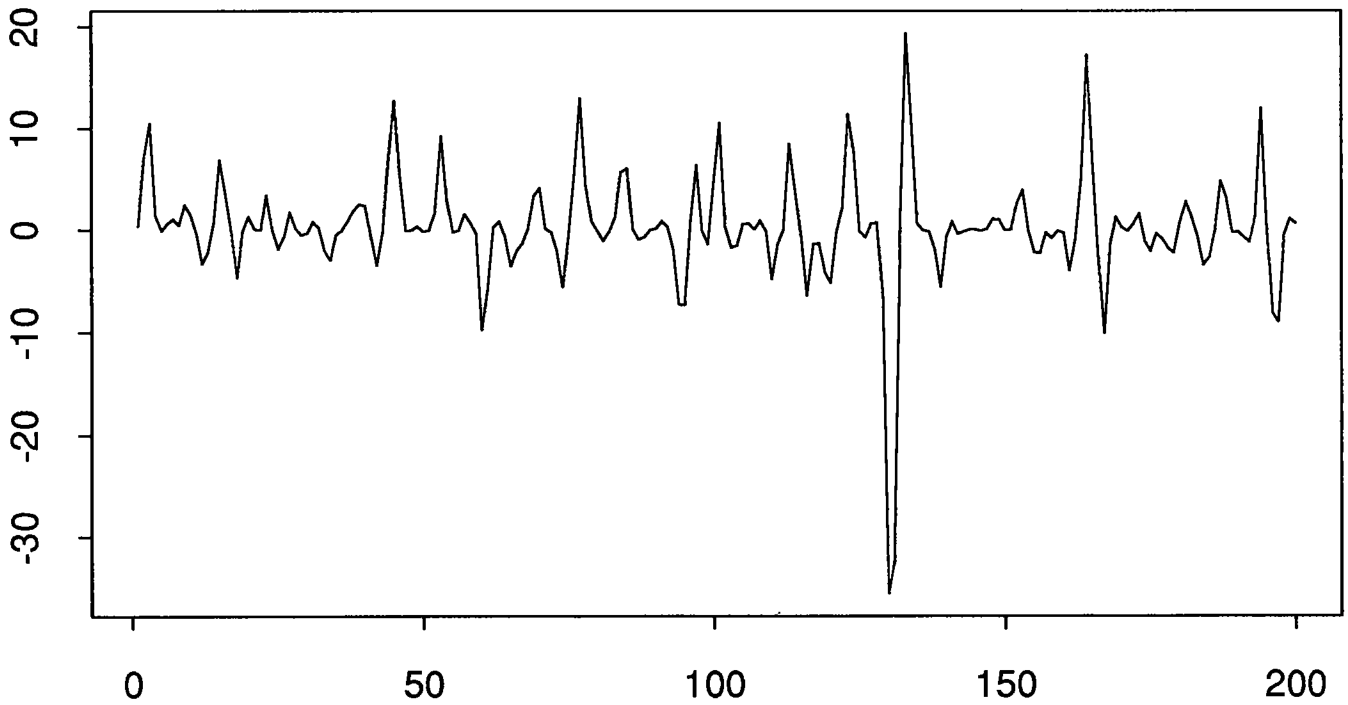

FIGURE 9. The time series Y

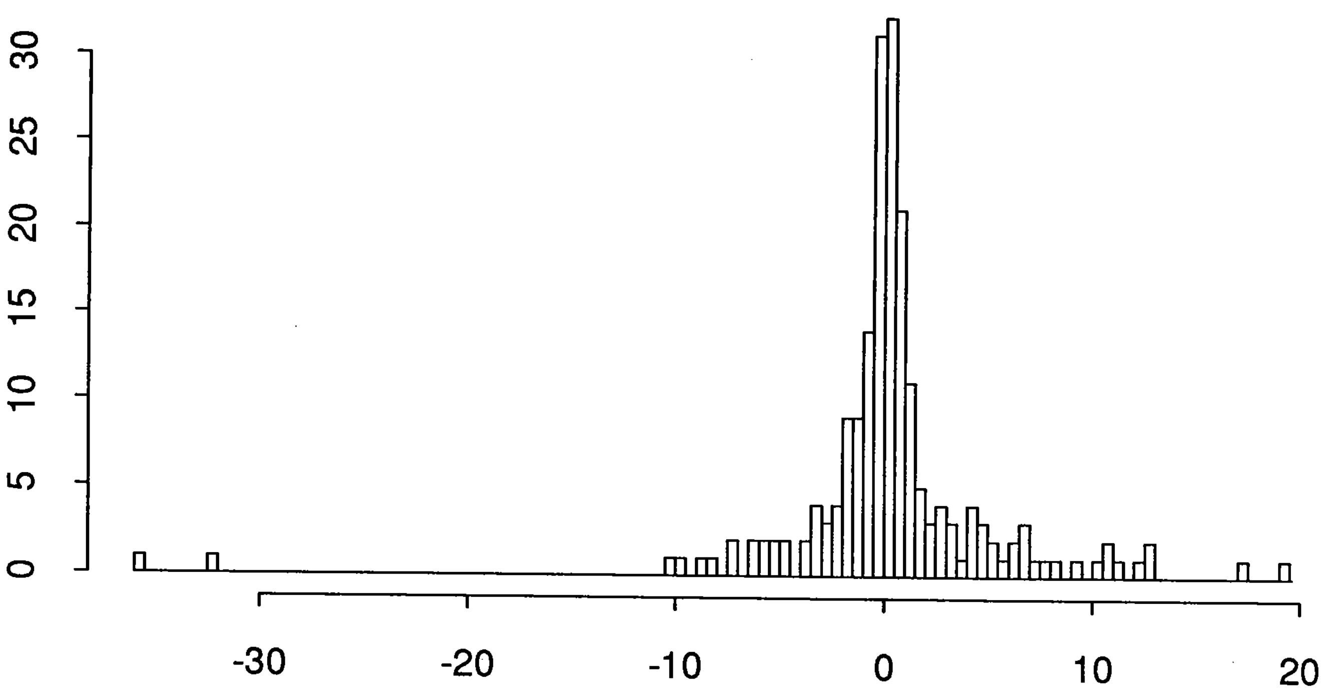

FIGURE 10. Histogram of the Y data

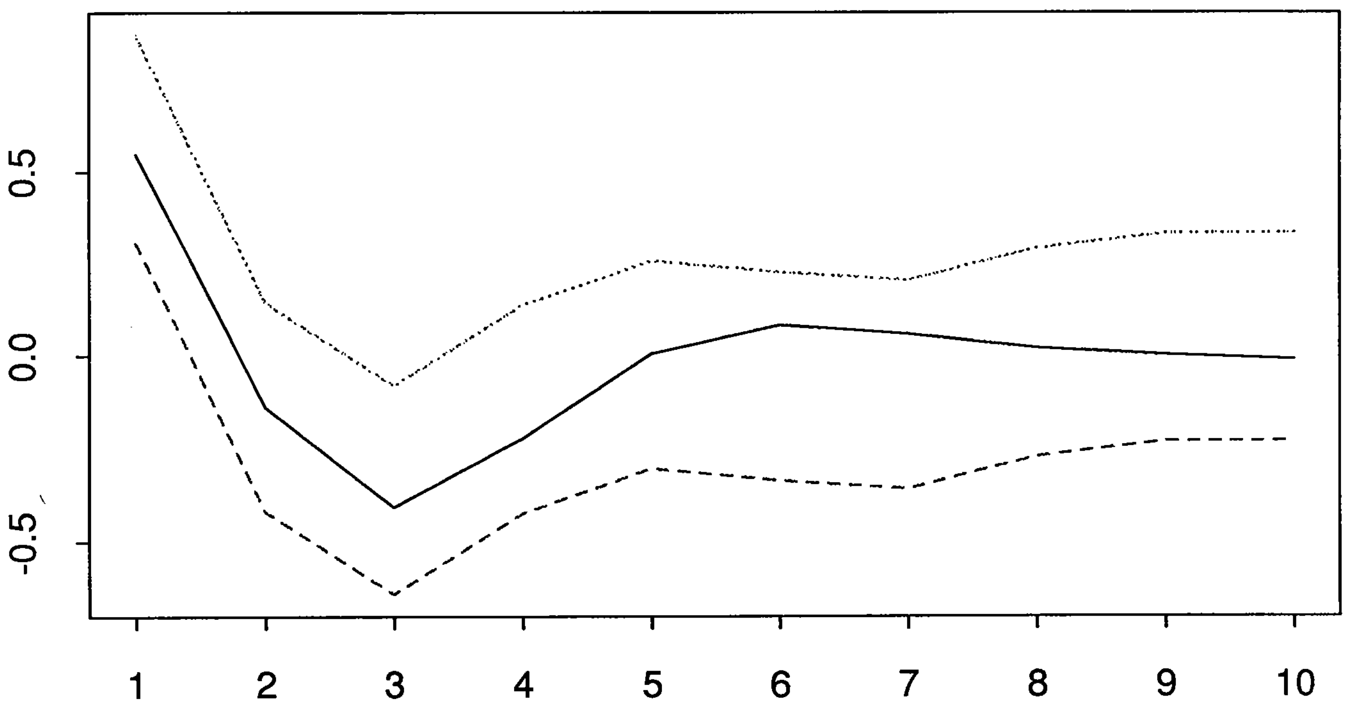

FIGURE 11. 95% confidence band for the trasformed autocorrelations

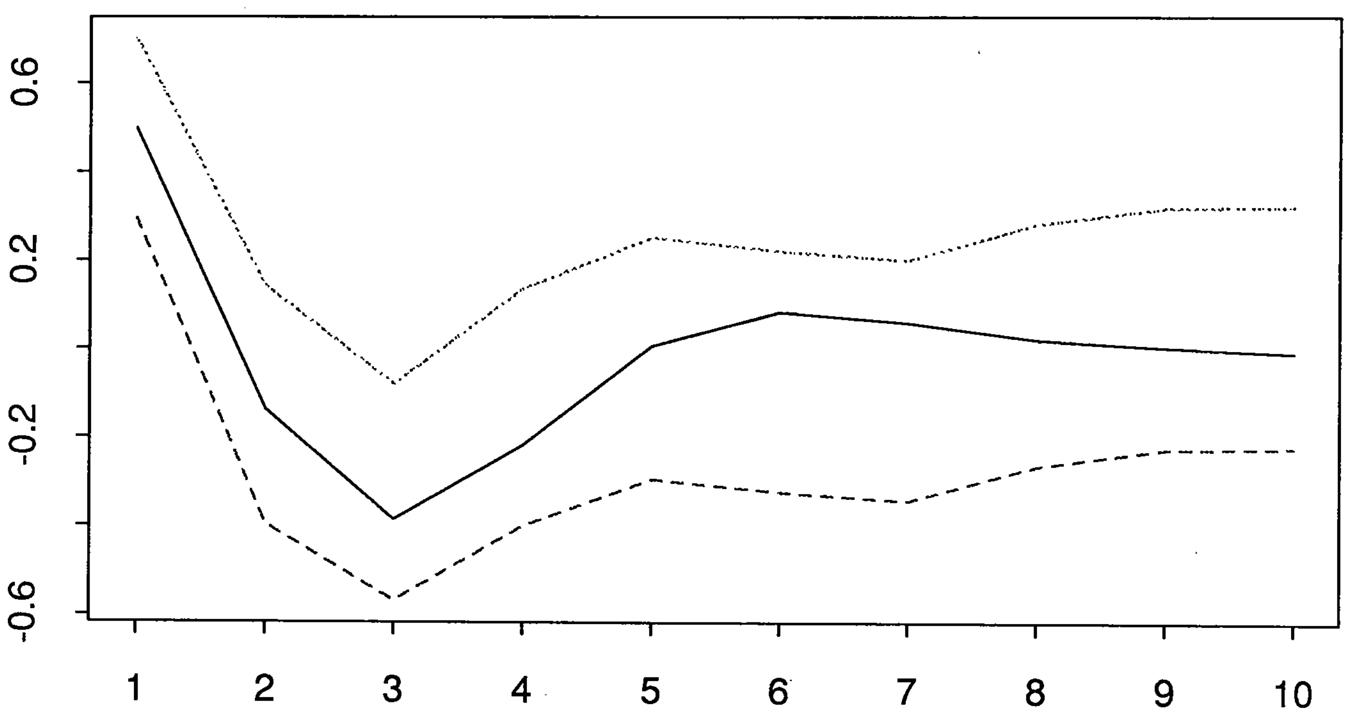

FIGURE 12. 95% confidence band for the autocorrelation coefficients

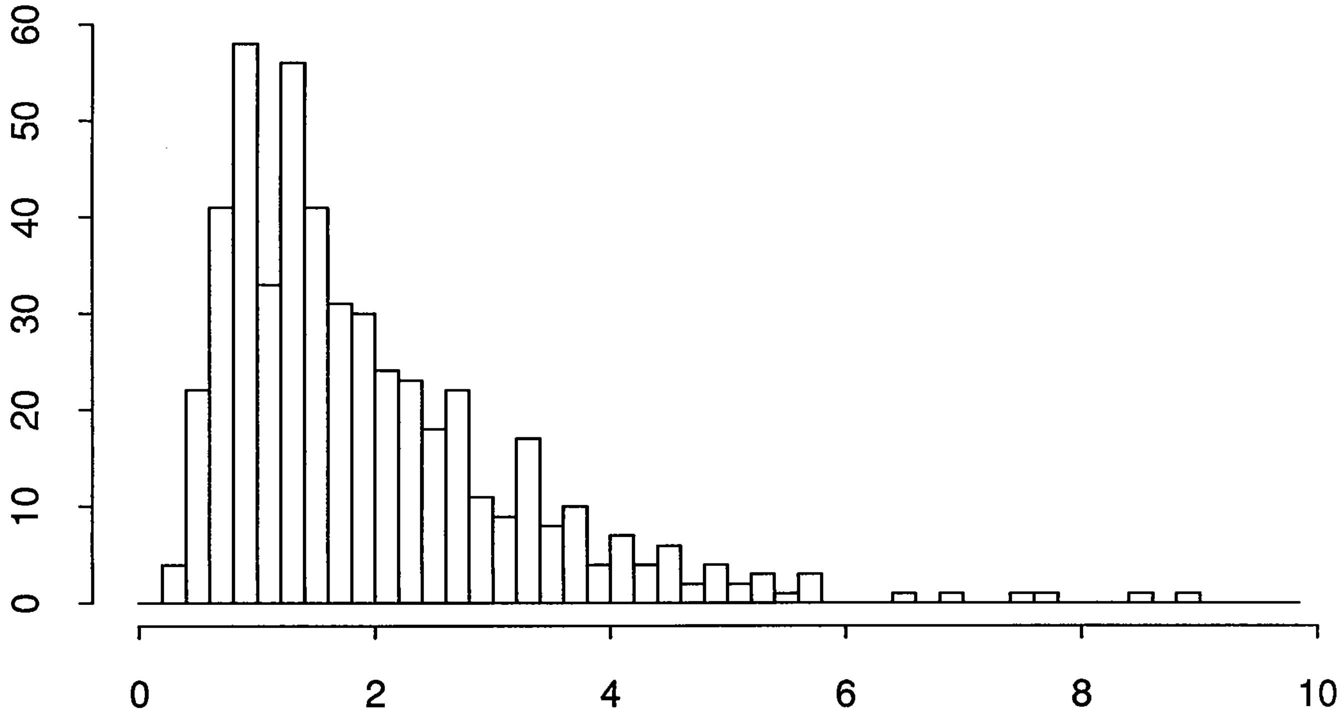

FIGURE 13. Bootstrap histogram of maximum modulus z value

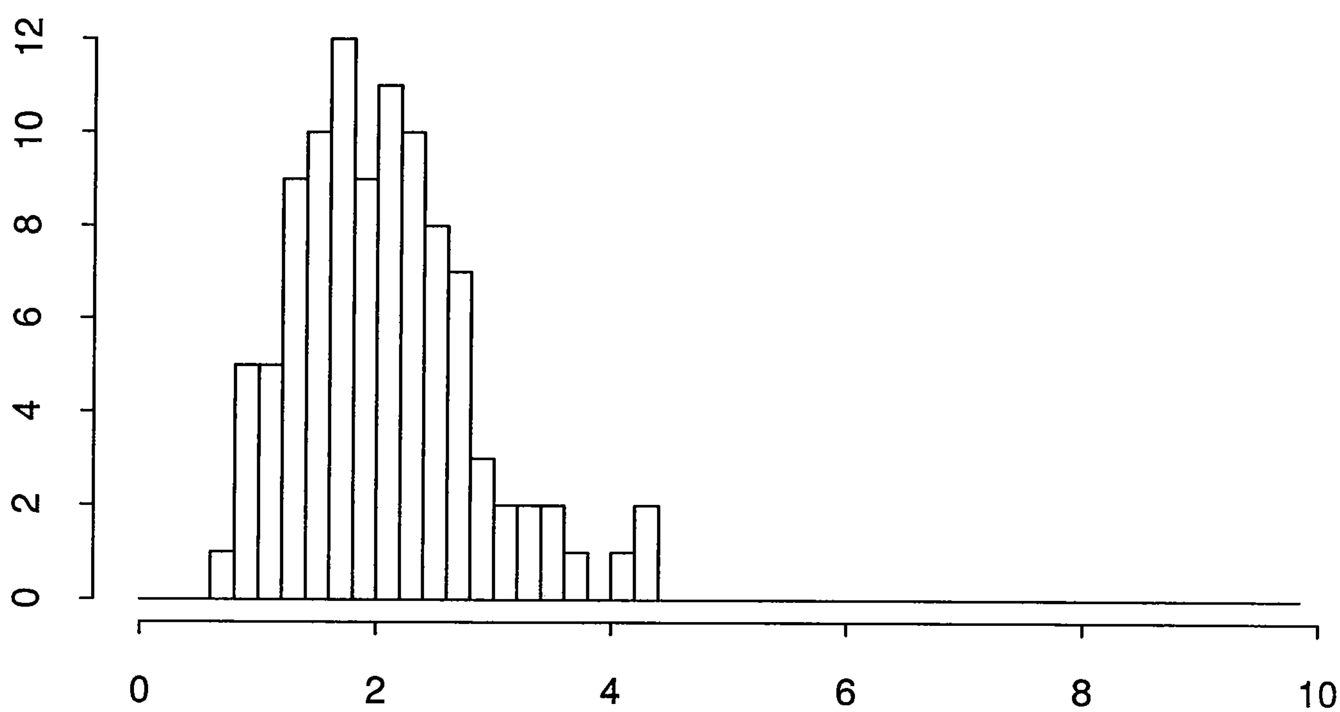

FIGURE 14. Empirical histogram of maximum modulus z value