Biometrika (1995), **82**, 3, *pp.* 561–74
Printed in Great Britain

On blocking rules for the bootstrap with dependent data

BY PETER HALL
Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia

JOEL L. HOROWITZ
Department of Economics, University of Iowa, Iowa City, Iowa 52242–1000, U.S.A.

AND BING-YI JING
Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

SUMMARY

We address the issue of optimal block choice in applications of the block bootstrap to dependent data. It is shown that optimal block size depends significantly on context, being equal to $n^{1/3}$, $n^{1/4}$ and $n^{1/5}$ in the cases of variance or bias estimation, estimation of a one-sided distribution function, and estimation of a two-sided distribution function, respectively. A clear intuitive explanation of this phenomenon is given, together with outlines of theoretical arguments in specific cases. It is shown that these orders of magnitude of block sizes can be used to produce a simple, practical rule for selecting block size empirically. That technique is explored numerically.

Some key words: Autoregression; Bias; Blocking methods; Bootstrap; Mean squared error; Variance.

1. INTRODUCTION

There are, broadly speaking, two approaches to using bootstrap methods for strictly stationary, dependent data. One is to describe the dependence through a structural model involving 'independent disturbances'. The best known models are those from time series analysis, such as autoregression or moving average models. The procedure there is first to fit the model, typically using standard methods, and then estimate residuals. The bootstrap may be implemented by resampling from the collection of estimated residuals, perhaps after adjustment for location and scale. Monte Carlo methods may then be employed to estimate more sophisticated quantities, for example the distributions of the earlier estimators of structural parameters. See for example Davis (1977), Freedman (1984), Efron & Tibshirani (1986) and Bose (1988). Despite the very good performance of these methods, they are restricted to relatively simple contexts where structural models are both plausible and tractable. More generally, bootstrap methods for less restrictive, more nonparametric contexts have been suggested by Hall (1985), Carlstein (1986) and Künsch (1989). They are based on 'blocking' arguments, in which the data are divided into blocks and those, rather than individual data values or estimated residuals, are resampled. There are basically two different ways of proceeding, depending on whether the blocks are overlapping or nonoverlapping. Both approaches were suggested by Hall (1985) in the context of spatial data. Carlstein (1986) proposed nonoverlapping blocks for univariate time series data, and Künsch (1989) suggested overlapping blocks in the same setting. Subsequent workers have tended to focus on the contributions of Carlstein and Künsch, ascribing the two methods to these authors, and to avoid ambiguity we shall make the same attribution in all that follows. Thus, 'Carlstein's rule' is taken to be the one based on nonoverlapping blocks, and 'Künsch's rule' on overlapping blocks. The method actually employed by Carlstein was a little different from that generally ascribed to him, in that he did not treat the case where the blocks are laid end-to-end to form a new time series of the same length as the original, but nevertheless that approach had been anticipated by Hall (1985). Götze & Künsch (1990), Lahiri (1991, 1992) and Davison & Hall (1993) discussed block bootstrap methods in the context of distribution estimation.

Our purpose in the present paper is to address the problem of optimal block choice when the block bootstrap is used in a variety of different contexts. It turns out that optimal block size depends very much on context. We identify three different settings of practical importance: estimation of bias or variance, estimation of a one-sided distribution function, and estimation of a two-sided distribution function. This last function is used to construct symmetric confidence intervals for an unknown parameter, and these enjoy enhanced coverage accuracy among two-sided confidence regions (Hall, 1988). We shall show that optimal block lengths in these three problems are of different sizes, being $n^{1/3}$, $n^{1/4}$ and $n^{1/5}$ respectively, where n is the length of the time series. At first sight this disparity is baffling. One of our contributions is to provide a simple explanation for it, as follows. Optimal block length is achieved by balancing error-about-the-mean against bias, to minimise mean squared error or indeed the error in any other L_p metric, for finite p. Bias terms in all three problems are of similar sizes, but variances are quite different for the three problems: they are essentially the variances of standardised second, third and fourth cumulants, in the respective problems of bias or variance estimation, one-sided distribution estimation, and two-sided distribution estimation. Elementary calculations based on this observation lead to the claims made earlier about orders of block length in the different problems.

The optimal asymptotic formula for block length, l, is $l \sim Cn^{1/k}$, where $k = 3$, 4 or 5. As indicated above, the value of k is known, determined by context. We shall show that this result is of practical benefit as well as theoretical interest, since it may be used as the basis for a simple rule for choosing block size. The rule operates by using empirical methods, similar to cross-validation, to choose the block size for a subseries of the original data set, of length $m < n$ say. This quantity may be re-calibrated so that it applies to the original, larger sample size, by multiplying by the factor $(n/m)^{1/k}$.

The length of blocks, or equivalently the number of blocks, might be considered to be a smoothing parameter since its adjustment generally has the effect of changing variance and bias in opposite directions. However, this interpretation is not entirely correct. For particularly large values of block length l, both squared bias and variance can increase with l; we shall provide details of this property in the case of estimation of bias or variance, but it emerges in other settings too. Fortunately the size of block length for which this occurs is an order of magnitude larger than that which produces overall minimisation of mean squared error, and so the difficulty is not as serious as might be feared. In some situations it may be removed by making a simple multiplicative correction to the bootstrap estimator.

In the case of distribution estimation, and in order to make our discussion as clear as possible, our development of methodology will follow a traditional line as given, for example, by Hall (1992). Specifically, if θ, $\hat{\theta}$, $\hat{\theta}^*$, s and $\hat{s}$ denote respectively a parameter, its estimator, a bootstrap version of $\hat{\theta}$, the standard deviation of $\hat{\theta}$, and a bootstrap version of s, then we develop methodology by considering the conditional distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{s}$, given the data, as an estimator of the unconditional distribution of $(\hat{\theta} - \theta)/s$. Approaches such as this presuppose the existence of an appropriate variance estimator, which may not be realistic, particularly for relatively small samples. Sometimes this difficulty may be remedied by applying a variance-stabilising transformation prior to applying the bootstrap, and back-transforming afterwards, but in other instances a block version of the iterated bootstrap is more appropriate. We shall not consider the latter approach here.

Section 2 details our theoretical results on block length for the three contexts discussed above. Section 3 introduces our empirical rule for choosing block length, and describes the results of numerical simulation experiments. Technical details leading to some of the results in § 2 are sketched in the Appendix.

2. Properties of block bootstrap estimators

2·1. *Blocking rules and bootstrap estimators*

Let the observed stationary sequence be $\mathscr{X} = (X_1, \ldots, X_n)$, and let b, l denote integers such that $n = bl$. Carlstein's rule asks that $\mathscr{X}$ be divided among just b disjoint blocks, the kth being $\mathscr{B}_k = (X_{(k-1)l+1}, \ldots, X_{kl})$ for $1 \leqslant k \leqslant b =: N$. Künsch's rule produces $n - l + 1$ overlapping blocks, the kth being

$$\mathscr{B}_k = (X_k, \ldots, X_{k+l-1}) \quad (1 \leqslant k \leqslant n - l + 1 =: N).$$

In either case the 'block bootstrap' method asks that we choose blocks $\mathscr{B}_1^*, \ldots, \mathscr{B}_b^*$ by resampling randomly, with replacement, from among $\mathscr{B}_1, \ldots, \mathscr{B}_N$. If $\mathscr{B}_i^* = (X_{i1}^*, \ldots, X_{il}^*)$, the bootstrap version of $\mathscr{X}$ is

$$\mathscr{X}^* = (X_1^*, \ldots, X_n^*) = (X_{11}^*, \ldots, X_{1l}^*, X_{21}^*, \ldots, X_{2l}^*, \ldots, X_{b1}^*, \ldots, X_{bl}^*).$$

In their simplest form, block bootstrap estimators may be described as follows. Let $\hat{\theta}$, a function of the data $\mathscr{X}$, denote an estimator of a parameter θ, and write $\hat{\theta}^*$ for the same function of $\mathscr{X}^*$. For example if $\bar{X} = n^{-1} \sum X_i$ is the sample mean then $\bar{X}^* = n^{-1} \sum X_i^*$ is the resample mean. Let E', var$'$ and pr$'$ represent expectation, variance and probability measure in the bootstrap distribution, conditional on $\mathscr{X}$. The bias and variance of $\hat{\theta}$ are given by $\beta = E(\hat{\theta}) - \theta$ and $s^2 = \text{var}\,(\hat{\theta})$, and their bootstrap estimators by $\hat{\beta} = E'(\hat{\theta}^*) - \hat{\theta}$ and $\hat{s}^2 = \text{var}'\,(\hat{\theta}^*)$. Bootstrap estimators of the distribution functions

$$F_1(x) = \text{pr}\,\{(\hat{\theta} - \theta)/s \leqslant x\}, \quad F_2(x) = \text{pr}\,(|\hat{\theta} - \theta|/s \leqslant x)$$

are respectively $\hat{F}_1(x) = \text{pr}'\,\{(\hat{\theta}^* - \hat{\theta})/\hat{s} \leqslant x\}$ and $\hat{F}_2(x) = \text{pr}'\,(|\hat{\theta}^* - \hat{\theta}|/\hat{s} \leqslant x)$.

In the case of bias estimation under Künsch's rule the estimator $\hat{\beta}$, as defined just above, is not as efficacious as one might hope. For example, if $\theta = \mu = E(X_i)$ denotes the marginal mean, then its bias estimator is $\hat{\beta} = \bar{X}' - \bar{X}$, where under Künsch's rule

$$\bar{X}' := E'(\bar{X}^*) = \bar{X} + (n - l + 1)^{-1} l^{-1} \{l(l-1)\bar{X} - T_1 - T_2\},$$

$$T_1 = \sum_{k=1}^{l-1} (l - k) X_k, \quad T_2 = \sum_{k=n-l+2}^{n} \{k - (n - l + 1)\} X_k.$$

In particular, $E'(\bar{X}^*) \neq \bar{X}$, and so the bootstrap estimator of the bias of the sample mean, which of course equals zero, is nonzero. When θ is a smooth function of the marginal mean μ, say $\theta = f(\mu)$, a remedy for the problem described above is to replace $\hat{\theta} = f(\bar{X})$ by

$\hat{\theta}' = f(\bar{X}')$ in the definition of the bias estimator, obtaining $\hat{\beta}' = E'\{f(\bar{X}^*)\} - f(\bar{X}')$. If resampling is done under Carsltein's rule then $\hat{\beta}$ and $\hat{\beta}'$ are identical. Without such a modification the variance of $\hat{\beta}$, under Künsch's rule, is increased by a factor proportional to n, with consequent inferior performance; see the second last paragraph of § 2·3, and also Lahiri (1991, 1992).

Similar modifications are possible whenever $\hat{\theta}$ may be interpreted as a smooth function of a vector mean. This includes the case where $\theta = \gamma(\theta) = E(X^2) - (EX)^2$, the marginal population variance, of which $\hat{\theta} = n^{-1} \sum X_i^2 - \bar{X}^2$ is an estimator. The latter is a bivariate function of means of the (X, X^2) process, and the Künsch-type block bootstrap estimator of the bias of $\hat{\theta}$ may be modified as before to reduce variance. We should distinguish here between the marginal variance of X_i, $\gamma(0)$, and the variance of the sample mean,

$$s^2 = \text{var}(\bar{X}) = n^{-1} \left\{ \gamma(0) + 2 \sum_{1 \leqslant j \leqslant n} (1 - n^{-1} j) \gamma(j) \right\} \sim n^{-1} \sum_{-\infty < j < \infty} \gamma(j) =: n^{-1} \sigma^2,$$

where $\gamma(j) = \text{cov}(X_0, X_j)$.

It is not always possible to achieve the 'ideal' identity $n = bl$. For a given $l < n$ we might take b to be equal to the integer part of n/l, and then consider $n' = bl$ as a substitute for n. We may construct block bootstrap estimators of bias or variance for each of the data sequences $X_{i+1}, \ldots, X_{i+n'}$ $(0 \leqslant i \leqslant n - n')$, and take the estimators' average value as 'the' bootstrap estimator of the relevant quantity. Or we might use b blocks of length l and one block of length $n - n'$ at the end. In sheer asymptotic terms, even replacing n by $n' = n + O(l) = n\{1 + O(b^{-1})\}$ does not alter the conclusions drawn later, even though it amounts to ignoring the last $n - n'$ data values in the original time series.

2·2. *Main results*

We shall consider block bootstrap estimators of (a) bias or variance, (b) one-sided distribution functions, and (c) two-sided distribution functions. If the quantity in question may be written as the expected value of a function of a vector mean of random variables, representing the data, then its bootstrap estimator equals the expected value, conditional on the data, of the same function of resampled data derived using the block bootstrap.

Consider a block bootstrap rule where there are b blocks of length l, with $bl \sim n$. Section 2·1 prescribed resampling schemes of this type. We shall show that in the cases (a), (b) and (c) noted above, mean squared errors of estimators are asymptotic to

$$n^{-2}(C_1 l^{-2} + C_2 n^{-1} l), \quad n^{-1}(C_1 l^{-2} + C_2 n^{-1} l^2), \quad n^{-2}(C_1 l^{-2} + C_2 n^{-1} l^3), \qquad (2·1)$$

respectively, where the positive numbers C_1 and C_2 do not depend on n, b or l. In the case of distribution function estimation, C_1 and C_2 depend on the argument of that function. The case of bias or variance estimation will be addressed in § 2·3, while that of distribution estimation will be described in § 2·4 and the Appendix. Minimising over l the mean squared error formulae at (2·1), we see that the optimal block sizes are of order $n^{1/3}$, $n^{1/4}$ and $n^{1/5}$ in the respective cases. The formulae at (2·1) do not depend on whether Carlstein's or Künsch's rule is used, and in fact C_1 is the same for both rules. The constant C_2 assumes a smaller value under Künsch's rule, with a consequent reduction in mean squared error and increase in optimal block length. However, this asymptotic improvement is hardly apparent in numerical studies, and in practice it would seem that Carlstein's rule is competitive with Künsch's.

Thus, the optimal block sizes are quite different in different contexts. To appreciate why they have the form that they do, note that the dominant contributions to the errors of block bootstrap estimators of bias or variance, one-sided distribution functions and two-sided distribution functions derive from second, third and fourth cumulants, respectively. There can be other contributions that provide terms of similar orders, but they do not dominate those just mentioned. Therefore they affect only the constant multiplier in the formula for optimal block size, not the actual order of block size.

For example, in the case of Carlstein's rule the dominant contributors to the errors of bias or variance estimators, of one-sided distribution estimators, and of two-sided distribution estimators are proportional to

$$K_2 = lb^{-1} \sum_{j=1}^{b} (Y_j - \bar{Y})^2, \quad K_3 = l^2 b^{-1} \sum_{j=1}^{b} (Y_j - \bar{Y})^3,$$
$$K_4 = l\left\{ l^2 b^{-1} \sum_{j=1}^{b} (Y_j - \bar{Y})^4 - 3\sigma^4 \right\}, \tag{2·2}$$

respectively, where

$$Y_j = l^{-1} \sum_{1 \leqslant i \leqslant l} X_{(j-1)l+i}, \quad \bar{Y} = b^{-1} \sum_{1 \leqslant j \leqslant b} Y_j = \bar{X},$$

σ^2 denotes the limit of $n \operatorname{var}(\bar{X})$ as $n \to \infty$, and the factors l, l^2 and l^3 in (2·2) ensure that the expected value of each quantity there equals $O(1)$. In more detail, let k_i denote the ith cumulant of $n^{\frac{1}{2}}\bar{X}$. Then the biases of estimators of bias or variance, one-sided distribution functions and two-sided distribution functions are of the same order as $n^{-1}|EK_2 - k_2|$, $n^{-\frac{1}{2}}|EK_3 - k_3|$ and $n^{-1}|EK_4 - k_4|$, respectively, and the variances are the same as those of $\operatorname{var}(n^{-1}K_2)$, $\operatorname{var}(n^{-\frac{1}{2}}K_3)$ and $\operatorname{var}(n^{-1}K_4)$, respectively. Thus, the respective mean squared errors are of the same sizes as $n^{-2}E(K_2 - k_2)^2$, $n^{-1}E(K_3 - k_3)^2$ and $n^{-2}E(K_4 - k_4)^2$.

Asymptotic formulae for the latter three quantities are identical to those given in (2·1). Of course, this does not amount to a proof of (2·1), but nevertheless it does give insight into why the quantities there have the form they do. It is clear after a little Taylor expansion that second cumulants should control the convergence rates of bias and variance estimators; the contribution of first cumulants vanishes from both quantities. And it is understandable that third and fourth cumulants should drive convergence rates in the case of estimators of one-sided and two-sided distributions, respectively, given that they appear in Edgeworth expansions of those quantities. However, it is not clear without some technical work that this argument should produce exactly the convergence rates given in (2·1). The Appendix outlines technicalities in the case of distribution estimation. The arguments are similar, but simpler, in the context of bias or variance estimation, which will therefore not be treated in any detail. In the Appendix we concentrate on the percentile bootstrap and Carlstein's blocking rule. It is straightforward to extend our argument to Künsch's rule, and it will be found that this rule reduces variance without asymptotically affecting the bias term; this is intuitively clear without doing the calculations, given the way that Künsch's overlapping blocks use the data more efficiently and consequently reduce variability. The percentile-t bootstrap is similar, provided the asymptotic variance is used as the standard for scaling. Extra bias terms can be introduced if Studentising is by an estimate of the finite-sample variance, and then the ubiquitous l^{-2} term in the formula for bias changes, with subsequent alterations to optimal block size and deterioration in the convergence rate of the distribution estimator. These matters are discussed in an unpublished research report by F. Götze and H. R. Künsch.

2·3. *Bias and variance estimation*

Let $(\psi, \hat{\psi})$ denote either $(\beta, \hat{\beta}')$ or $(s^2, \hat{s}^2)$, computed under either Carlstein's or Künsch's rule. The mean squared error of $\hat{\psi}$ is given by

$$E(\hat{\psi} - \hat{\psi})^2 \sim n^{-2}(C_1 l^{-2} + C_2 n^{-1} l), \qquad (2\cdot3)$$

where C_1, C_2 are positive constants. The first component on the right-hand side, with factor C_1, derives from squared bias, while the other is contributed by variance. The asymptotically optimal block size is obtained by minimising the right-hand side over l, and equals $l_0 = C_0 n^{1/3}$, where $C_0 = (2C_1/C_2)^{1/3}$. The fact that the exponent of n in the formula for l_0 is a known, absolute constant, and only C_0 need be estimated empirically, is critical to the empirical methods suggested in § 3.

It may be shown by Taylor expansion that when $\theta = f(\mu)$ and the mean is univariate,

$$(C_1, C_2) = C_3^2(\tau^2, 2v\sigma^4), \qquad (2\cdot4)$$

where

$$C_3 = \begin{cases} -\frac{1}{2}f''(\mu) & \text{if } \psi = \beta, \\ f'(\mu) & \text{if } \psi = s^2, \end{cases}$$

$$\tau = 2\sum_{j=1}^{\infty} j\gamma(j), \quad \sigma^2 = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j),$$

and $v = 1$ under Carlstein's rule, $v = \frac{2}{3}$ under Künsch's. Thus, the value of C_0 is a function only of the covariance structure of the underlying process $\{X_i\}$, and is not influenced in any other way by the strength of dependence of the process, or even by the nature of the problem being addressed. In asymptotic terms the number of blocks should be $1\cdot5^{1/3} - 1 =$ 14% larger in the case of Carlstein's rule than under Künsch's rule; the minimum asymptotic mean squared error is $1\cdot5^{2/3} - 1 = 31\%$ larger in the case of Carlstein's rule than it is for Künsch's rule; and blocks should be longer for processes with larger values of

$$\rho = (vC_0^3)^{\frac{1}{2}} = \left|\sum_{j=1}^{\infty} j\gamma(j)\right| \Big/ \left\{\frac{1}{2}\gamma(0) + \sum_{j=1}^{\infty} \gamma(j)\right\}.$$

The first term on the right-hand side of (2·3) derives from the fact that

$$|E(\hat{\psi}) - \psi| = C_1^{-\frac{1}{2}}(nl)^{-1} + o\{(nl)^{-1}n^{-1}b^{-\frac{1}{2}}\}.$$

Longer and more informative expansions may be developed in specific contexts. For example, in the setting described in the previous paragraph it may be shown that

$$E(\hat{\psi}) - \psi = n^{-1}(l^{-1}\tau + b^{-1}\sigma^2)C_3 + o\{n^{-1}(l^{-1} + b^{-1})\}.$$

Therefore, provided $\tau > 0$,

$$|E(\hat{\psi}) - \psi| \sim \text{const}\, n^{-1}(l^{-1}\tau + n^{-1}l\sigma^2)$$

is approximately convex in l, minimised by taking $l \sim (\tau\sigma^{-2}n)^{\frac{1}{2}}$. Most importantly, the bias contribution to mean squared error of $\hat{\psi}$ is not a decreasing function of l, even to a first-order approximation. In the vicinity of values l that minimise mean squared error the bias contribution is decreasing, but for larger l's it is actually increasing. This contrasts with the roles played by other smoothing parameters in Statistics, e.g. bandwidth in curve estimation problems, where bias is a monotone function of the parameter. However, the estimator $\tilde{\psi} = (1 - b^{-1})^{-1}\hat{\psi}$ has relatively monotone bias, satisfying

$$E(\tilde{\psi}) - \psi = C_3(nl)^{-1}\tau + o\{(nl)^{-1}\} + O(n^{-1}b^{-2}).$$

Furthermore the factor $(1 - b^{-1})^{-1}$ does not asymptotically affect the validity of (2·3), which continues to hold with (C_1, C_2) given by (2·4) if $\hat{\psi}$ is replaced by $\tilde{\psi}$.

In the case of Künsch's rule, if the estimator is not modified by recentring as suggested in § 2·1, the variance is increased from order $n^{-3}l$ to $n^{-2}l$, with consequent deterioration in performance. While the formula for bias changes, and the multiplicative correction discussed just above is no longer efficacious, the order of magnitude of bias does not alter. Thus, the effect of not centring is to increase error about the mean, not bias.

Künsch (1989, p. 1226) discussed a version of (2·3), and also the multiplicative correction, in the context of the jackknife estimator of the variance of the arithmetic mean. Sufficient regularity conditions for (2·3) are that θ and $\hat{\theta}$ be representable as smooth functions of a multivariate population and sample mean, respectively, where 'smooth' entails one derivative in the case $\psi = s^2$ and two derivatives when $\psi = \beta$; that sufficiently many moments of X_i are finite; and that the process $\{X_i\}$ is Rosenblatt mixing (Götze & Hipp, 1983).

2·4. *Distribution estimation*

Let $\hat{F}_j$ denote the bootstrap estimate of F_j, defined in § 2·1. The mean squared error of $\hat{F}_j$ satisfies

$$E(\hat{F}_j - F_j)^2 \sim n^{-j}(C_{1j}l^{-2} + C_{2j}n^{-1}l^{j+1})\phi^2, \tag{2·5}$$

where C_{1j} and C_{2j} denote squares of polynomials, and ϕ is the standard Normal density function. The two components on the right-hand side, with coefficients C_{1j} and C_{2j}, derive from squared bias and variance respectively. Asymptotically optimal block length for estimation of F_j at a particular point is given by $l_{0j} = C_{0j}n^{1/(j+3)}$, where $C_{0j} = [2C_{1j}/\{(j+1)C_{2j}\}]^{1/(j+3)}$. Optimality in a global sense may be achieved by replacing the function C_{1j} by an integral average. A proof of (2·5) under Carlstein's rule and in the case where $\theta = \mu$ is a mean is outlined in the Appendix. In fact we show there that $(C_{11}(x), C_{21}(x)) = (D_1, D_2)\pi_1(x)^2$, where

$$\pi_1(x) = \tfrac{1}{6}(x^2 - 1), \quad D_1 = (\tfrac{3}{2}\tau\delta\sigma^{-5} - \tau_3\sigma^{-3})^2, \quad D_2 = 9 - 6\mu_4\sigma^{-4} + \mu_6\sigma^{-6},$$

provided that the latter two constants are nonzero; and if the stochastic process $\{X_i\}$ is regarded as doubly infinite and centred at its mean,

$$\delta = \sum\sum_{-\infty < i_1, i_2 < \infty} E(X_0 X_{i_1} X_{i_2}),$$

$$\tau_3 = 3\sum_{j=1}^{\infty} jE(X_0X_j^2 + X_0^2X_j) + 6\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(i+j)E(X_0X_iX_{i+j})$$

and $\mu_{2k} = \lim n^k E(\bar{X}^{2k})$. If Künsch's rule is employed instead of Carlstein's, the effect is to replace D_2 by $\frac{2}{3}D_2$ in the formula $(C_{11}, C_{21}) = (D_1, D_2)\pi_1^2$. If $\{X_i\}$ is a Gaussian process with zero mean then $\delta = \tau_3 = 0$, and so $D_1 = 0$. Thus, in this special case the contribution of squared bias, but not variance, to (2·5) is of smaller order than $n^{-j}l^{-2}$.

The version of (2·5) in the case where θ is a mean and $j = 1$ is of particular interest, since then, global optimisation over l is equivalent to local optimisation. This follows from the fact that in the formula $(C_{11}, C_{21}) = (D_1, D_2)\pi_1^2$, D_1 and D_2 are real numbers rather than polynomials.

Appropriate regularity conditions for (2·5) are those in § 2·3, together with a smoothness condition on the sampling distributions; for the latter, see Götze & Hipp (1983) or Götze & Künsch (1993).

3. EMPIRICAL METHOD FOR BLOCK CHOICE

3·1. *General*

Section 2 showed that in quite general contexts the optimal block size is $n^{1/k}$, with the value of k being 3, 4 or 5, depending on circumstance. In the present section we suggest an empirical rule for estimating optimal block size for a time series of smaller length than the original, say $m<n$. Once this has been determined, at $\hat{l}_m$ say, the optimal block size $\hat{l}_n$ for the original time series of length n may be estimated from the formula $\hat{l}_n=(n/m)^{1/k}\hat{l}_m$.

Our technique is as follows. Let $\mathscr{S}$ denote the set of all $n-m+1$ runs of length m obtainable from the original time series, and apply the block bootstrap method to each subseries from $\mathscr{S}$. Let l' denote block size here. Each application of the block bootstrap produces a point estimate of the quantity ψ of interest; the latter may be a bias, a variance or a distribution function. Let $\hat{\psi}_i$ $(1\leqslant i\leqslant n-m)$ denote the bootstrap estimates of ψ computed from the $n-m$ runs of length m in $\mathscr{S}$, and write $\hat{\psi}$ for the estimate of ψ computed from the entire data set of length n, using a plausible pilot block size l. An estimate of mean squared error in a sample of size m, using block length l', is the average of the squares of the differences $\hat{\psi}_i-\hat{\psi}$. We select that value of l', say $\hat{l}_m$, which minimises this quantity, and then revise our choice of l to $(n/m)^{1/k}\hat{l}_m$. This procedure may be iterated if desired, replacing the original pilot choice of l by an updated version.

To assess the performance of this method we applied it to data simulated from moving average model and autoregressive models.

3·2. *Moving average model*

We consider the model

$$X_i=\theta+(Y_{i+1}+Y_{i+2})2^{-\frac{1}{2}}\quad(-\infty<i<\infty),\tag{3·1}$$

where the Y_j's are independent and identically distributed random variables with a centred χ_1^2 distribution. We took $\hat{\theta}=\bar{X}$, the sample mean, as our estimator of θ, the population mean, and studied two block bootstrap estimation problems: first, estimating the variance, $s^2=\text{var}(\hat{\theta})$, and secondly, estimating the one-sided distribution function, $F_1(x)=\text{pr}\{(\hat{\theta}-\theta)/s\leqslant x\}$. In both cases we employed Künsch's form of the block bootstrap.

Table 1 presents mean squared errors of estimates of s^2 and $F_1(x)$ for various block lengths l, when $n=100$. Those data make it clear that optimal block sizes for variance estimation and distribution estimation are, respectively, $l=3$ and $l=8$.

We generated samples of size 100 from the moving average model at (3·1), and applied our empirical method for determining the optimal block length l. For a particular sample, Table 2 describes the main elements of our procedure in the cases of variance and distribution estimation. For the particular sample that is the subject of Table 2(a), we chose $\hat{l}_n=5$ as the initial block length. Mean squared error, given in Table 2(a), was then calculated as described in § 3·1, giving the values listed in the table. It was minimised by $\hat{l}_m=2$, which in turn gave $\hat{l}_n=(n/m)^{1/3}\hat{l}_m\simeq 3·2$ as an estimate of optimal block size for the full sample. In the second iteration we chose $\hat{l}_n=3$ as the initial block length; again, $\hat{l}_m=2$ minimised mean squared error. Thus, the process converged after only one iteration, and

Table 1. *Determination of optimal block length for variance estimation and distribution estimation, for the case $n = 100$; each entry is the average 500 simulations*

(*a*) *Variance estimation*

l	$E(\hat{s}^2)$	Bias	SD	MSE
1	1·94	−2·04	0·78	4·78
2	2·85	−1·13	1·13	2·56
3	3·28	−0·70	1·39	2·42*
4	3·30	−0·68	1·49	2·68
5	3·40	−0·58	1·62	2·96
6	3·48	−0·50	1·60	2·81
7	3·57	−0·41	2·04	4·33
8	3·61	−0·37	1·83	3·48
9	3·50	−0·48	1·88	3·75
10	3·39	−0·60	1·93	4·06
True	3·98			

(*b*) *Distribution estimation*

l	$E\{\hat{F}_1(0)\}$	Bias	SD	$10^3 \times$ MSE
1	0·491	−0·0438	0·0355	3·17
2	0·505	−0·0297	0·0337	2·02
3	0·511	−0·0239	0·0331	1·67
4	0·507	−0·0276	0·0259	1·43
5	0·512	−0·0229	0·0252	1·16
6	0·511	−0·0235	0·0243	1·15
7	0·509	−0·0254	0·0241	1·22
8	0·512	−0·0222	0·0243	1·08*
9	0·510	−0·0245	0·0237	1·16
10	0·509	−0·0257	0·0242	1·24
11	0·509	−0·0260	0·0236	1·23
12	0·508	−0·0264	0·0243	1·28
13	0·510	−0·0249	0·0228	1·14
14	0·510	−0·0248	0·0246	1·22
15	0·507	−0·0273	0·0247	1·36
True	0·5346			

In (a) for the purpose of calculating bias, standard deviation (SD) and mean squared error (MSE), the true value of s^2 was computed exactly.
In (b), the true value of $F_1(0)$ was computed by averaging over 20 000 simulations.
* Indicates the minimum MSE value.

the optimal block size was estimated as $\hat{l}_n = 3$, to be compared with the theoretical optimum of 3, determined from Table 1(a). Similarly, starting with $\hat{l}_n = 8$ in Table 2(b) gave convergence to $\hat{l}_n = 7$ on the next iteration, and later iterations did not alter this value. The theoretical optimum is 8.

Tables 3 and 4 describe the average performance over 1000 and 200 samples, respectively. Table 3 illustrates the distribution of empirically chosen block length in the cases of vari-

Table 2. *Estimates of mean squared error* (MSE) *for various block lengths, in the cases of variance estimation and distribution estimation, and a particular sample*

(*a*) *Variance estimation*

$\hat{l}_m$	MSE after first iteration with $\hat{l}_n = 5$	MSE after second iteration with $\hat{l}_n = 3$
1	1·85	1·54
2	0·53*	0·53*
3	0·84	0·92
4	1·25	1·33
5	1·47	1·49
6	1·94	1·95
7	2·80	2·88
8	2·11	2·05
9	3·35	3·44
10	2·47	2·39

(*b*) *Distribution estimation*

$\hat{l}_m$	$10^3 \times$ MSE after first iteration with $\hat{l}_n = 10$	$10^3 \times$ MSE after second iteration with $\hat{l}_n = 8$
1	5·45	3·90
2	2·88	2·51
3	2·06	1·37
4	1·57	1·55
5	1·14	1·31
6	1·20	1·27
7	0·78*	1·06*
8	0·96	1·18
9	0·84	1·12
10	1·02	1·18

In (a) we took $n = 100$ and $m = 25$, and $\hat{l}_n = 8$ as initial block size. With this choice, the second column of (a) depicts estimates of MSE in a sample of size $m = 25$, for various block sizes $\hat{l}_m$. Those values are minimised by $\hat{l}_m = 2$, which suggests taking $\hat{l}_n = (n/m)^{1/3}\hat{l}_m = 3{\cdot}2 \simeq 3$ as the optimal block size. With this choice of $\hat{l}_n$, the third column of (a) depicts revised estimates of MSE for a sample of size $m = 25$. Again, MSE is a minimum at $\hat{l}_m = 2$, and (a) suggests $\hat{l}_n = 3$.
Details of (b) are as for (a) except that the estimated value of block size was 7 after both one and two iterations, and the exponent of n/m is now $\frac{1}{4}$ instead of $\frac{1}{3}$.
* Indicates minimum MSE value.

ance and distribution estimation, respectively. In the former case the distribution has a pronounced mode in the vicinity of the optimal block size, $l = 3$, and little mass elsewhere. In the latter, the distribution has most of its mass between $l = 6$ and $l = 10$. Note that we cannot actually obtain the theoretical optimum, at $l = 8$, because of the nature of our Richardson extrapolation. Larger sample sizes give better performance.

The number of iterations required in order for our procedure to converge was 1 for 93% of our variance estimation simulations, and 1 for 92% of our distribution estimation

Table 3. *Distribution of empirically chosen block length for variance and distribution estimation*

(*a*) *Variance estimation*

$\hat{l}_n$	2	3	5	6	8	10	11	13
Frequency	0·27	0·52	0·06	0·07	0·02	0·02	0·02	0·02

(*b*) *Distribution estimation*

$\hat{l}_n$	5	6	7	9	10	11	12	13	15
Frequency	0·06	0·22	0·16	0·21	0·26	0·01	0·01	0·05	0·02

$n = 100$, $m = 25$, and number of simulations is 1000 and 200, respectively. In each case, the total frequency outside the tabulated range is less than 0·005.

Table 4. *Determination of optimal block length for distribution estimation in the autoregressive model with* $\alpha = 0{\cdot}3$ *and* $\alpha = -0{\cdot}1$

	$\alpha = 0{\cdot}3$				$\alpha = -0{\cdot}1$			
l	$E\{\hat{F}_1(0)\}$	Bias	SD	$10^3 \times$ MSE	$E\{\hat{F}_1(0)\}$	Bias	SD	$10^3 \times$ MSE
1	0·500	0·0164	0·0468	2·197	0·501	0·00119	0·0445	1·983
2	0·500	0·0139	0·0371	1·377	0·501	0·00065	0·0351	1·234
3	0·500	0·0190	0·0293	0·861	0·500	−0·00019	0·0301	0·903
4	0·500	0·0117	0·0266	0·708	0·501	0·00055	0·0268	0·721
5	0·499	0·0081	0·0253	0·641	0·499	−0·00060	0·0254	0·645
6	0·500	0·0163	0·0246	0·607	0·501	0·00099	0·0243	0·592
7	0·501	0·0232	0·0242	0·592	0·500	0·00039	0·0237	0·563
8	0·500	0·0121	0·0228	0·522	0·501	0·00046	0·0210	0·441*
9	0·500	0·0153	0·0224	0·503	0·500	−0·00024	0·0232	0·538
10	0·500	0·0159	0·0223	0·499*	0·501	0·00064	0·0229	0·524
11	0·500	0·0125	0·0224	0·504	0·501	0·00083	0·0233	0·545
12	0·500	0·0153	0·0231	0·534	0·500	0·00040	0·0221	0·487
13	0·500	0·0121	0·0235	0·554	0·501	0·00058	0·0230	0·527
14	0·500	0·0191	0·0236	0·563	0·501	0·00084	0·0236	0·558
15	0·499	0·0095	0·0239	0·574	0·500	−0·00034	0·0225	0·507
True	0·4985				0·50			

Details are the same as for the case of Table 1(b).

simulations. In the remaining 7 or 8% of cases, convergence was always achieved after 2 iterations. Details of parameter settings are as for Table 3.

3·3. *Autoregressive model*

Consider the model

$$X_i = \alpha X_{i-1} + \varepsilon_i \quad (-\infty < i < \infty), \tag{3·2}$$

where the ε_i's are independent and identically distributed random variables with a centred χ_1^2 distribution. Again we took $\hat{\theta} = \bar{X}$, the sample mean, as our estimator of θ, the population mean, and studied block bootstrap estimation of the one-sided distribution function, $F_1(x) = \text{pr}\,\{(\hat{\theta} - \theta)/s \leqslant x\}$.

Two different values of α were chosen: $\alpha = 0{\cdot}3$ and $\alpha = -0{\cdot}1$. Table 4 presents mean squared errors of estimates of $F_1(x)$ for various block lengths l, when $n = 100$. The optimal block sizes are respectively $l = 10$ and $l = 8$.

Table 5. *Frequencies of empirically chosen block length for distribution estimation in the autoregressive model with $\alpha = 0{\cdot}3$ and $\alpha = -0{\cdot}1$*

α	$l_n = 4$	$l_n = 6$	$l_n = 7$	$l_n = 8$	$l_n = 10$	$l_n = 11$	$l_n = 13$	$l_n = 14$
0·3	0·05	0·22	0·09	0·10	0·44	0·01	0·08	0·01
−0·1	0·01	0·24	0·12	0·17	0·36	0·01	0·008	0·01

Here $n = 100$, $m = 25$, and number of simulations is 1000 and 500, respectively. In each case, the total frequency outside the tabulated range is less than 0·005.

Table 5 describes the average performance over 200 samples. It illustrates the distribution of empirically chosen block length for $\alpha = 0{\cdot}3$ and $\alpha = -0{\cdot}1$. In both cases, the distribution has most of its mass between $l = 6$ and $l = 10$. However, the medians of the two distributions are 10 and 8, which are the optimal block sizes for $\alpha = 0{\cdot}3$ and $\alpha = -0{\cdot}1$, respectively.

ACKNOWLEDGEMENT

We are grateful to two referees, the Associate Editor and the Editor for their very helpful comments, which have led to this improved version of the paper. The research of J. L. Horowitz was supported in part by a National Science Foundation grant.

APPENDIX

Derivation of (2·5) *in the case of sample mean*

When $\theta = \mu = 0$ we have, by formal Edgeworth expansion,

$$F_1(x) = \text{pr}\,\{(\hat{\theta} - \theta)/s \leqslant x\} = \Phi(x) + n^{-\frac{1}{2}} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + O(n^{-3/2}),$$

where Φ and ϕ denote the standard normal distribution and density functions respectively,

$$p_1 = \alpha_1 \pi_1, \quad p_2 = \alpha_1^2 \pi_2 + \alpha_2 \pi_3, \quad \alpha_1 = n^{\frac{1}{2}} E(\bar{X}^3) s^{-3}, \quad \alpha_2 = n\{E(\bar{X}^4) s^{-4} - 3\},$$

$$\pi_2(x) = \tfrac{1}{72}(x^5 - 10x^3 + 15x), \quad \pi_3(x) = \tfrac{1}{24}(x^3 - 3x),$$

and π_1 is as in § 2·4. Similarly,

$$\hat{F}_1(x) = \text{pr}'\,\{(\hat{\theta}^* - \hat{\theta})/\hat{s} \leqslant x\} = \Phi(x) + n^{-\frac{1}{2}} \hat{p}_1(x)\phi(x) + n^{-1} \hat{p}_2(x)\phi(x) + O_p(n^{-3/2}),$$

where

$$\hat{p}_1 = \hat{\alpha}_1 \pi_1, \quad \hat{p}_2 = \hat{\alpha}_1^2 \pi_2 + \hat{\alpha}_2 \pi_3, \quad \hat{\alpha}_1 = l^{\frac{1}{2}} E'(\bar{X}^* - \bar{X})^3 \hat{s}^{-3}, \quad \hat{\alpha}_2 = l\{E'(\bar{X}^* - \bar{X})^4 \hat{s}^{-4} - 3\}.$$

Arguing thus we may show that

$$E(\hat{F}_1 - F_1)^2 \sim n^{-1} E(\hat{p}_1 - p_1)^2 \phi^2 = n^{-1} E(\hat{\alpha}_1 - \alpha_1)^2 \pi_1^2 \phi^2, \qquad (\text{A}{\cdot}1)$$

$$E(\hat{F}_2 - F_2)^2 \sim 4n^{-2} E(\hat{p}_2 - p_2)^2 \phi^2. \qquad (\text{A}{\cdot}2)$$

The latter result uses the fact that p_1 and $\hat{p}_1$ are even polynomials, whence

$$\hat{F}_2 - F_2 = 2n^{-1}(\hat{p}_2 - p_2)\phi + O_p(n^{-3/2}),$$

and in fact the remainder is $O_p(n^{-2})$. In the rest of this appendix we shall sketch proofs that, for constants $A_1, A_2, \ldots$,

$$E(\hat{\alpha}_1) - \alpha_1 = A_1 l^{-1}\{1 + o(1)\} + O(b^{-1} l^{\frac{1}{2}}), \qquad \text{var}\,(\hat{\alpha}_1) \sim A_2 n^{-1} l^2, \quad \text{var}\,(\hat{\alpha}_2) \sim A_3 n^{-1} l^3. \tag{A·3}$$

Similarly, it may be shown that

$$E(\hat{\alpha}_1^2) - \alpha_1^2 \sim A_4 l^{-1}, \quad E(\hat{\alpha}_2) - \alpha_2 \sim A_5 l^{-1}, \quad \text{var}\,(\hat{\alpha}_1^2) = O(n^{-1} l), \quad \text{cov}\,(\hat{\alpha}_1^2, \hat{\alpha}_2) = O(n^{-1} l^2).$$

Therefore,

$$E(\hat{\alpha}_1 - \alpha_1)^2 \sim A_1^2 l^{-2} + A_2 n^{-1} l^2, \quad E(\hat{p}_1 - p_1)^2 \sim (A_4 \pi_2 + A_5 \pi_3)^2 l^{-2} + A_3 \pi_3^2 n^{-1} l^3,$$

which proves (2·5).

We begin by deriving the first two results in (A·3). Write α_{1n} for α_1, to stress dependence on n, and put

$$\bar{X}_l = l^{-1} \sum_{1 \leqslant i \leqslant l} X_i, \quad Y_j = l^{-1} \sum_{1 \leqslant i \leqslant l} X_{(j-1)l+i}, \quad \bar{Y}^{(k)} = b^{-1} \sum_{1 \leqslant j \leqslant b} Y_j^k, \quad S_k = \bar{Y}^{(k)} - E\bar{Y}^{(k)},$$

$$\hat{\mu}_{kl} = b^{-1} \sum_{1 \leqslant j \leqslant b} (Y_j - \bar{Y})^k, \quad \rho_l^2 = lE(\bar{X}_l^2), \quad \alpha_{1l} = l^2 E(\bar{X}_l^3) \rho_l^{-3},$$

$$B_1 = -3\rho_l^{-1}, \quad B_2 = -\tfrac{3}{2}\alpha_{1l}\rho_l^{-2}, \quad B_3 = \rho_l^{-3}, \quad B_{11} = 3\alpha_{1l}\rho_l^{-2}, \quad B_{22} = \tfrac{15}{4}\alpha_{1l}\rho_l^{-4},$$

$$B_{12} = \tfrac{3}{2}\rho_l^{-3}, \quad B_{23} = \tfrac{3}{2}\rho_l^{-5}.$$

By Taylor expansion,

$$\begin{aligned} \hat{\alpha}_1 &= l^{1/2} \hat{\mu}_{3l} \hat{\mu}_{2l}^{-3/2} \\ &= \alpha_{1l} + (B_1 l S_1 + B_2 l S_2 + B_3 l^2 S_3) + \tfrac{1}{2}(B_{11} l S_1^2 + B_{22} l^2 S_2^2) + (B_{12} l^2 S_1 S_2 + B_{23} l^3 S_2 S_3) + \ldots, \end{aligned} \tag{A·4}$$

where here and below '...' indicates terms of such high order that they do not contribute to the final result. It follows that

$$\begin{aligned} E(\hat{\alpha}_1) - \alpha_1 &= \alpha_{1l} - \alpha_{1n} + \tfrac{1}{2}(B_{11} l E S_1^2 + B_{22} l^2 E S_2^2) + (B_{12} l^2 E S_1 S_2 + B_{23} l^3 E S_2 S_3) + \ldots, \\ \text{var}\,(\hat{\alpha}_1) &= \text{var}\,(B_1 l S_1 + B_2 l S_2 + B_3 l^2 S_3) + \ldots . \end{aligned} \tag{A·5}$$

Put $\mu_{il} = l^{\lfloor (i+1)/2 \rfloor} E(\bar{X}_l^i)$, where $\lfloor . \rfloor$ denotes the integer part function. Now,

$$E(S_1^2) \sim n^{-1}\sigma^2, \quad E(S_2^2) = O(n^{-1} l^{-1}), \quad E(S_3^2) \sim b^{-1} E(Y_1^6) \sim n^{-1} l^{-2} \mu_{6l},$$

$$E(S_1, S_3) \sim b^{-1} E(Y_1^4) \sim n^{-1} l^{-1} \mu_{4l}.$$

Therefore,

$$\begin{aligned} E(\hat{\alpha}_1) - \alpha_1 &= \alpha_{1l} - \alpha_{1n} + O(lES_1^2 + l^2 ES_2^2 + l^2 |ES_1 S_2| + l^3 |ES_2 S_3|) \\ &= \alpha_{1l} - \alpha_{1n} + O(b^{-1} l^{\frac{1}{2}}), \\ \text{var}\,(\hat{\alpha}_1) &\sim \text{var}\,(B_1 l S_1 + B_3 l^2 S_3) \\ &\sim n^{-1} l^2 (B_1^2 \sigma^2 + 2 B_1 B_3 \mu_4 + B_3^2 \mu_6) \sim n^{-1} l^2 (9 - 6\mu_4 \sigma^{-4} + \mu_6 \sigma^{-6}). \end{aligned}$$

The second part of (A·3) follows from the latter formula, while the first is a consequence of the former formula and the fact that

$$\alpha_{1l} - \alpha_{1n} = (\tfrac{3}{2}\tau\delta\sigma^{-5} - \tau_3\sigma^{-3}) l^{-1} + o(l^{-1}). \tag{A·6}$$

To derive (A·6), observe that since

$$\mu_{2n} = \gamma(0) + 2 \sum_{j=1}^{n} (1 - n^{-1} j)\gamma(j),$$

$$\mu_{3n} = E(X_1^3) + 3 \sum_{j=1}^{n} (1 - n^{-1} j) E(X_0 X_j^2 + X_0^2 X_j) + 6 \sum\sum_{i,j \geqslant 1; i+j \leqslant n} \{1 - n^{-1}(i+j)\} E(X_0 X_i X_{i+j})$$

then, writing $\mu_{i\infty} = \lim \mu_{in}$, we have $\mu_{2\infty} - \mu_{2n} \sim n^{-1}\tau$ and $\mu_{3\infty} - \mu_{3n} \sim n^{-1}\tau_3$. Therefore, since $\mu_{3\infty} = \delta$,

$$\alpha_{1n} = \mu_{3n}\mu_{2n}^{-3/2} = \mu_{3\infty}\mu_{2\infty}^{-3/2} + n^{-1}(\tfrac{3}{2}\tau\delta\sigma^{-5} - \tau_3\sigma^{-3}) + o(n^{-1}),$$

with an analogous formula holding for α_{1l}.

Derivation of the third part of (A·3) is similar to that of the second. First develop an expansion of $\hat{\alpha}_2$ similar to that of $\hat{\alpha}_1$ at (A·4), in which the linear term is $(B_4 lS_1 + B_5 l^2 S_2 + B_6 l^3 S_4)$ instead of $(B_1 lS_1 + B_2 lS_2 + B_3 l^2 S_3)$, where the constants B_j are functions of l that have finite limits as $l \to \infty$. An argument similar to that producing (A·5) gives

$$\text{var}\,(\hat{\alpha}_2) = \text{var}\,(B_4 lS_1 + B_5 l^2 S_2 + B_6 l^3 S_4) + \dots .$$

Now,

$$E(S_1^2) = O(n^{-1}), \quad E(S_2^2) \sim \text{const}\, n^{-1} l^{-1}, \quad E(S_4^2) \sim \text{const}\, n^{-1} l^{-3}, \quad E(S_2 S_4) \sim \text{const}\, n^{-1} l^{-2}.$$

Therefore,

$$\text{var}\,(\hat{\alpha}_2) \sim \text{var}\,(B_5 l^2 S_2 + B_6 l^3 S_4) \sim \text{const}\, n^{-1} l^3,$$

as had to be shown.

References

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16**, 1709–22.
Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.* **14**, 1171–9.
Davis, W. W. (1977). Robust interval estimation of the innovation variance of an ARMA model. *Ann. Statist.* **5**, 700–8.
Davison, A. C. & Hall, P. (1993). On Studentizing and blocking methods for implementing the bootstrap with dependent data. *Aust. J. Statist.* **35**, 215–24.
Efron, B. & Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1**, 54–77.
Freedman, D. A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12**, 827–42.
Götze, F. & Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahr. verw. Geb.* **64**, 211–39.
Götze, F. & Künsch, H. R. (1990). Blockwise bootstrap for dependent observations: higher order approximations for Studentized statistics (Abstract). *Bull. Inst. Math. Statist.* **19**, 443.
Hall, P. (1985). Resampling a coverage process. *Stoch. Proces. Applic.* **20**, 231–46.
Hall, P. (1988). On symmetric bootstrap confidence intervals. *J. R. Statist. Soc.* B **50**, 35–45.
Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York: Springer.
Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217–41.
Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statist. Prob. Let.* **11**, 335–41.
Lahiri, S. N. (1992). Edgeworth correction by "moving block" bootstrap for stationary and nonstationary data. In *Exploring the Limits of the Bootstrap*, Ed. R. Lepage and L. Billard, pp. 183–214. New York: Wiley.

[*Received May* 1993. *Revised January* 1995]